Supporting Data

Structure-Based Virtual Screening, Docking, ADMET, Molecular Dynamics, and MM-PBSA Calculations for the discovery of Potential Natural SARS-CoV-2 Helicase Inhibitors From The Traditional Chinese Medicine

Ahmed M. Metwaly[a,b]*, Alaa Elwan [c], Abdul-Aziz M.M. El-Attar[d], Sara T. Al-Rashood [e], Ibrahim. H. Eissa[c]*

[a] Pharmacognosy and Medicinal Plants Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt

[b]. Biopharmaceutical Products Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications (SRTA-City), Alexandria 21934, Egypt

[c] Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo11884, Egypt.

[d] Pharmaceutical Analytical Chemistry Department, Faculty of Pharmacy, Al-Azhar University, 11884, Nasr City, Cairo, Egypt

[e] Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, P.O. Box 2457, Riyadh 11451, Saudi Arabia, salrashood@ksu.edu.sa

***Corresponding authors:**

Ahmed M. Metwaly

Pharmacognosy and Medicinal Plants Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt**.** **Email:** ametwaly@azhar.edu.eg

Ibrahim H. Eissa

Medicinal Chemistry Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt**.** **Email:** Ibrahimeissa@azhar.edu.eg

Content

Method

Fingerprint study

Fingerprint study of the selected compounds against the co-crystallized ligand was carried out calculated using Discovery studio 4.0. At first, the CHARMM force field was applied then the compounds were prepared using prepare ligand protocol. Then, the compounds were used as a test set while the co-crystallized ligand was used as a reference compounds. The protocol was adjusted to give the most related compounds to the co-crystallized ligand. The default molecular properties were applied. The used fingerprints were based on some parameters related to type of atoms which may be one of the following: charge, hybridization, H-bond acceptor, H-bond donor, Positive ionizable, Negative ionizable, Halogen, Aromatic, or None of the above. In addition, it includes the ALogP category of atoms.

Molecular Similarity

Molecular Similarity of the examined natural compounds against the co-crystallized ligand was carried out calculated using Discovery studio 4.0. At first, the CHARMM force field was applied then the compounds were prepared using prepare ligand protocol. Then, the compounds were used as a test set while the co-crystallized ligand was used as a reference compounds. The default molecular properties were applied. The molecular properties include number of rotatable bonds, number of rings, number of aromatic rings, number of hydrogen bond donors (HBA), number of hydrogen bond acceptors (HBD), partition coefficient (ALog p), molecular weight (M. Wt), and molecular fractional polar surface area (MFPSA).

Docking studies

Crystal structure target enzyme was obtained from Protein Data Bank. The docking investigation was accomplished using MOE2014 software. At first, the crystal structure of the protein was prepared by removing water molecules. Only one chain was retained beside the co-crystallized ligand. Then, the selected chain was protonated and subjected to minimization of energy process. Next, the active site of the target protein was defined.

Structures of the tested compounds and the co-crystallized ligand were drawn using ChemBioDraw Ultra 14.0 and saved as MDL-SD format. Such file was opened using MOE to display the 3D structures which were protonated and subjected to energy minimization. Formerly, validation of the docking process was performed by docking the co-crystallized ligand against the isolated pocket of active site. The produced RMSD value indicated the validity of process. Finally,

docking of the tested compounds was done through the dock option inserted in compute window. For each docked molecule, 30 docked poses were produced using ASE for scoring function and force field for refinement. The results of the docking process were then visualized using Discovery Studio 4.0 software.

ADMET studies

ADMET descriptors (absorption, distribution, metabolism, excretion and toxicity) of the compounds were determined using Discovery studio 4.0. At first, the CHARMM force field was applied then the tested compounds were prepared and minimized according to the preparation of small molecule protocol. Then ADMET descriptors protocol was applied to carry out these studies.

Toxicity studies

The toxicity parameters of the tested compounds were calculated using Discovery studio 4.0. At first, the CHARMM force field was applied then the compounds were prepared and minimized according to the preparation of small molecule protocol. Then different parameters were calculated from the toxicity prediction (extensible) protocol.

Molecular Dynamics (MD) Simulations

The system was prepared using the web-based CHARMM-GUI[1-3] interface with the CHARMM36 force field[4]. All the simulations were done using the NAMD 2.13[5] package. The TIP3P explicit solvation model was used[6], and the periodic boundary conditions were set with a dimension of the dimensions ---- Å, ----- Å, and ----- Å in x, y, and z, respectively. The parameters for the top docking results were generated using the CHARMM general force field[7] Afterward, the system was neutralized using ---- (Cl^-/Na^+) ions. The MD protocols involved minimization, equilibration, and production. a 2 *fs* time step of integration was chosen for all MD simulations, the equilibration was carried in the canonical (*NVT*) ensemble, while the isothermal–isobaric (*NPT*) ensemble was for the production. Through the 100 *ns* of MD production, the pressure was set at 1 atm using the Nose´–Hoover Langevin piston barostat[8, 9] with a Langevin piston decay of 0.05 *ps* and a period of 0.1 *ps*. The temperature was set at 298.15 K using the Langevin thermostat[10]. A distance cutoff of 12.0 Å was applied to short-range nonbonded interactions with a pair list distance of 16 Å, and Lennard Jones interactions were smoothly truncated at 8.0 Å. Long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method[11, 12], where a grid spacing of 1.0 Å was used for all simulation cells. All covalent bonds involving

hydrogen atoms were constrained using the SHAKE algorithm[13]. For consistency, we have applied the same protocol for all MD simulations.

Binding Energy Calculations

The one-average molecular mechanics generalized Born surface area (MM/GBSA)[14, 15] approach implemented in the MOLAICAL code[16] was used for the relative binding energy calculations, in which the ligand (L) binds to the protein receptor (R) to form the complex (RL),

$$\Delta G_{bind} = \Delta G_{RL} - \Delta G_R - \Delta G_L$$

which can be represented by contributions of different interactions,

$$\Delta G_{bind} = \Delta H - T\Delta S = \Delta E_{MM} + \Delta G_{Sol} - T\Delta S$$

where the changes in the gas phase molecular mechanics (ΔE_{MM}), solvation Gibbs energy (ΔG_{Sol}), and conformational entropy $(-T\Delta S)$ are determined as follows: ΔE_{MM} is the sum of the changes in the electrostatic energies ΔE_{ele}, the van der Waals energies ΔE_{vdW}, and the internal energies ΔE_{int} (bonded interactions); ΔG_{Sol} is the total of both the polar solvation (calculated using the generalized Born model) and the nonpolar solvation (the solvent-accessible surface area) and $-T\Delta S$ is calculated by the normal mode analysis. The solvent dielectric constant of 78.5 and the surface tension constant of 0.03012 kJ mol^{-1} Å^2 were used for MM/GBSA calculations.

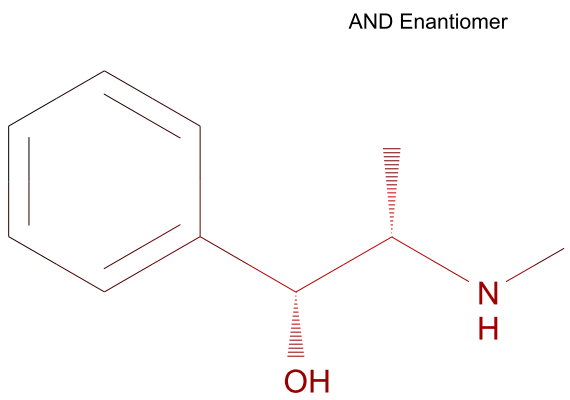
MMPBSA

The g_mmpbsa package of GROMACS was utilized to calculate the MM/PBSA (Molecular Mechanics/Poisson Boltzmann Surface Area) binding free energies followed by final MD production run to get a detailed overview of the molecular interactions between the protein and ligand. The free solvation energy (polar and nonpolar solvation energies) and potential energy (electrostatic and Van der Waals interactions) of each protein-ligand complex were analyzed to determine the total ΔG_{bind} of the complex. The binding energies were calculated using the following equation in this method:

$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand})$

Here, the $\Delta G_{binding}$ = the total binding energy of the protein-ligand complex, $G_{protein}$ = the energy of free protein, and G_{ligand} = the of unbounded ligand.

Toxicity Report



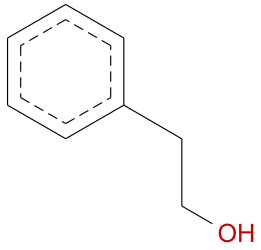
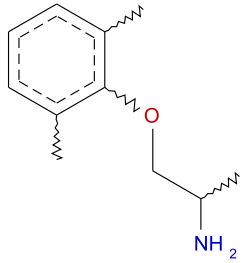
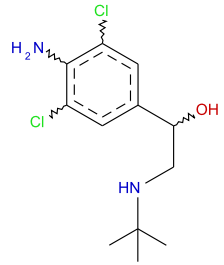
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.694
Enrichment: 1.32
Bayesian Score: 3.75
Mahalanobis Distance: 10.1
Mahalanobis Distance p-value: 0.0165

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenethyl Alcohol	Mexiletine .HCl (Free base form)	Clenbuterol
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.514	0.540	0.621
Reference	J Toxicol Environ Health 12(2-3):235-44; 1983	Iyakuhin Kenkyu 14(4):550-570; 1983	Iyakuhin Kenkyu 15(4):597-613; 1984

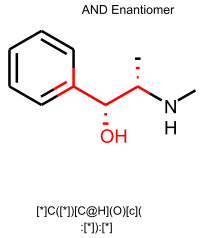
Model Applicability

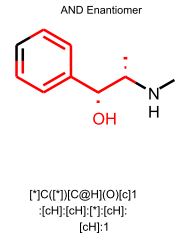
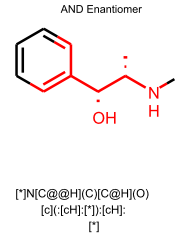
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

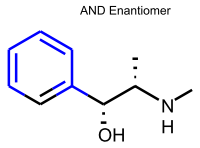
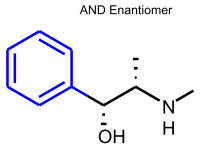
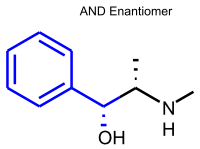
Feature Contribution

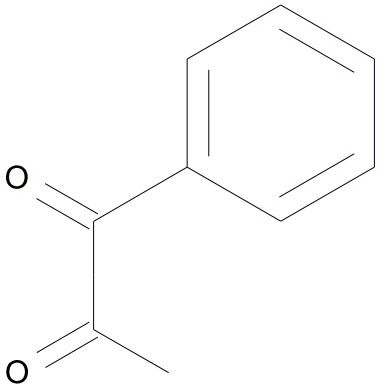
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1702724181		0.558	9 out of 9

SCFP_6	-561151481	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	0.523	6 out of 6
SCFP_6	1318513260	 [*]N[C@@H](C)[C@H](O) [c](:[cH]:[*]):[cH]: [*]	0.478	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1379591900	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.282	33 out of 84
SCFP_6	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.188	12 out of 28
SCFP_6	-1631132401	 [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.0973	8 out of 17



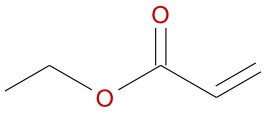
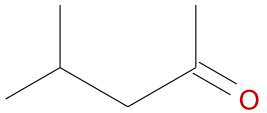
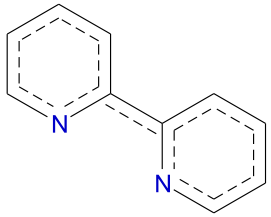
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Non-Toxic
Probability: 0.514
Enrichment: 0.977
Bayesian Score: -0.907
Mahalanobis Distance: 7.79
Mahalanobis Distance p-value: 0.705

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ethyl Acrylate	Methylisobutylketone	2;2'-Dipyridyl
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.530	0.558	0.558
Reference	Toxicol Appl Pharmacol 60:106-111; 1981	Fundam Appl Toxicol 8:310-327; 1987	Teratology 18(1):63-70; 1978

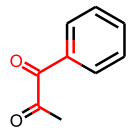
Model Applicability

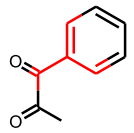
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

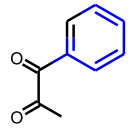
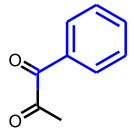
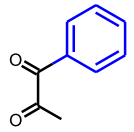
1. All properties and OPS components are within expected ranges.

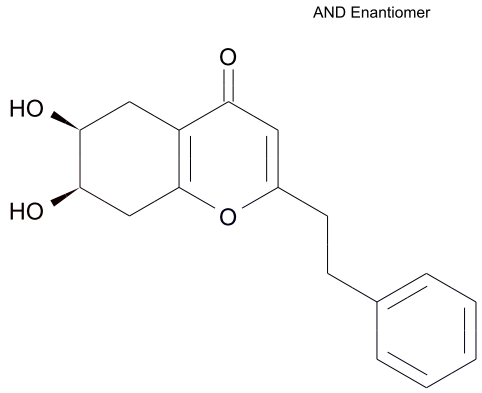
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1980302127	 [*]C(=[*])C(=O)[c](:[*]):[*]	0.25	5 out of 7

SCFP_6	-2056718782	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.0786	33 out of 58

Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Toxic in training set**
SCFP_6	-1379591900	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84
SCFP_6	1649310982	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.212	10 out of 24
SCFP_6	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.188	12 out of 28



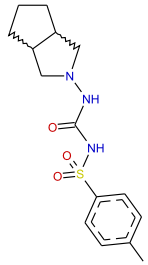
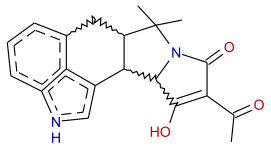
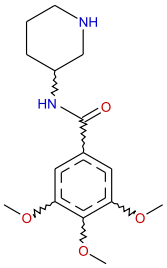
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.546
Enrichment: 1.04
Bayesian Score: -0.0291
Mahalanobis Distance: 11
Mahalanobis Distance p-value: 0.00128

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Sulfonylurea Gliclazide	Cyclopiazonic Acid	N-(3-Piperidyl)3;4;5-trimethoxybenzamide
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.536	0.591	0.603
Reference	Yakuri to Chiryo 9:3551-3571; 1981	J Toxicol Environ Health 14:585-594; 1984	Kiso to Rinsho 18:91-101; 1984

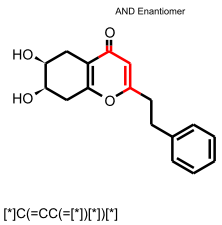
Model Applicability

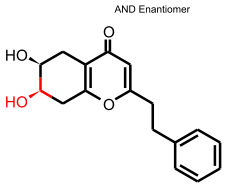
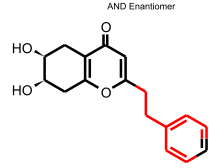
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

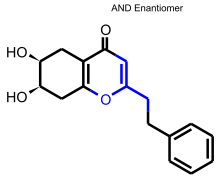
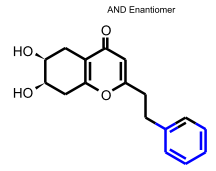
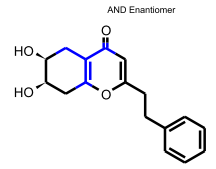
Feature Contribution

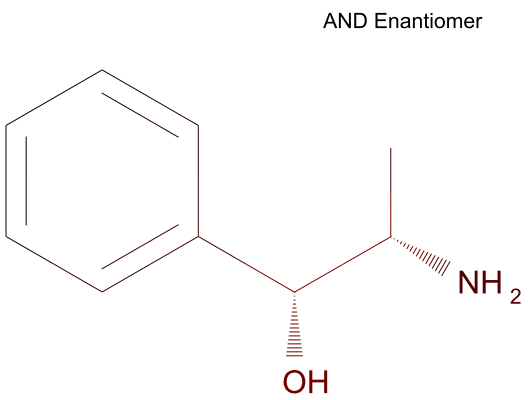
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1971196727	 [*]C(=CC(=[*])[*])[*]	0.293	13 out of 18

SCFP_6	-424515134	 [*]C([*])O	0.275	39 out of 56
SCFP_6	-1211866396	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.21	8 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	616547045	 [*]C\C(=C\[*])\O[*]	-0.438	1 out of 4
SCFP_6	-1379591900	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84
SCFP_6	55464376	 [*]CC(=C([*])[*])C(=[*])[*]	-0.252	1 out of 3



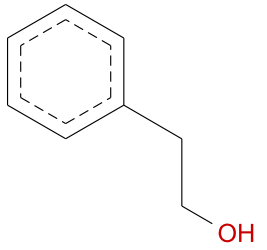
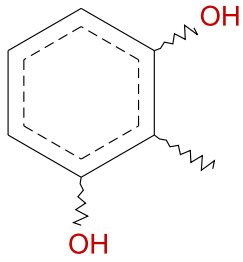
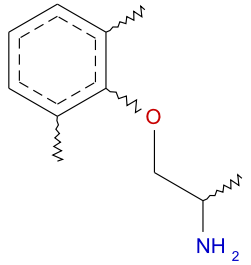
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.638
Enrichment: 1.21
Bayesian Score: 2.33
Mahalanobis Distance: 8.88
Mahalanobis Distance p-value: 0.214

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenethyl Alcohol	2-Methylresorcinol	Mexiletine .HCl (Free base form)
Structure			
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic
Distance	0.557	0.560	0.592
Reference	J Toxicol Environ Health 12(2-3):235-44; 1983	Fundam Appl Toxicol 7:293-298; 1986	Iyakuhin Kenkyu 14(4):550-570; 1983

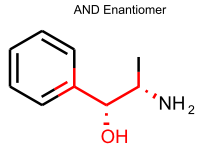
Model Applicability

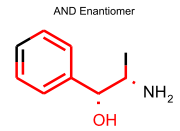
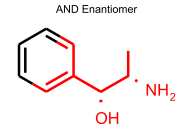
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

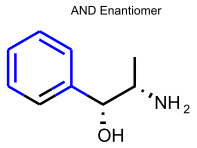
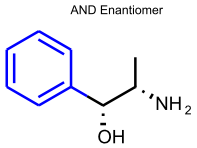
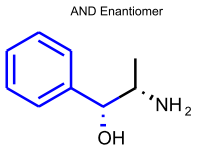
Feature Contribution

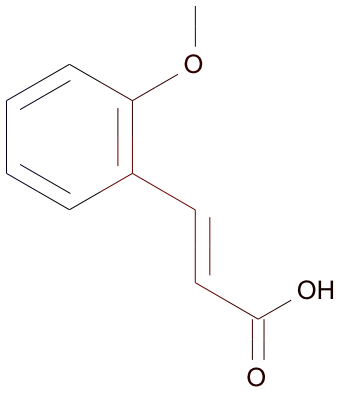
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1702724181	 [*]C([*])[C@H](O)[c](:[*]):[*]	0.558	9 out of 9

SCFP_6	-561151481	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	0.523	6 out of 6
SCFP_6	1318513260	 [*]N[C@@H](C)[C@H](O) [c](:[cH]:[*]):[cH]: [*]	0.478	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1379591900	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.282	33 out of 84
SCFP_6	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.188	12 out of 28
SCFP_6	-1631132401	 [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.0973	8 out of 17



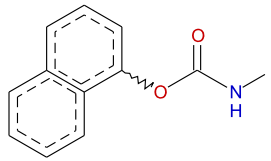
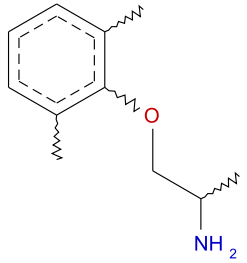
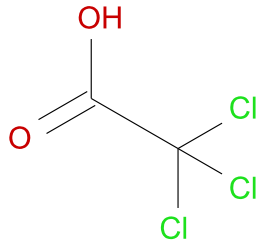
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.625
Enrichment: 1.19
Bayesian Score: 1.99
Mahalanobis Distance: 9.21
Mahalanobis Distance p-value: 0.122

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Carbaryl	Mexiletine .HCl (Free base form)	Trichloroacetic Acid
Structure			
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Non-Toxic
Distance	0.540	0.562	0.570
Reference	Toxicol Appl Pharmacol 26(4):621-38; 1973	Iyakuhin Kenkyu 14(4):550-570; 1983	Teratology 40:445-451; 1989

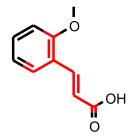
Model Applicability

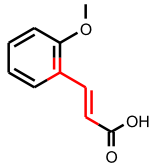
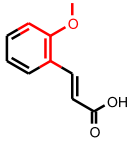
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

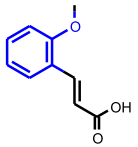
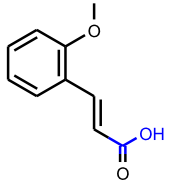
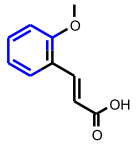
Feature Contribution

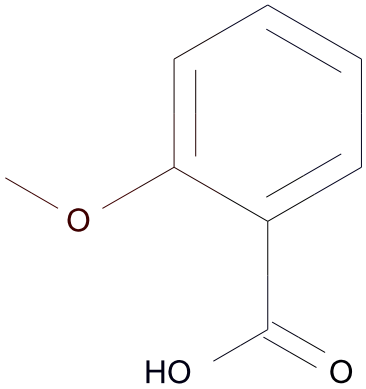
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1977229858	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	0.478	4 out of 4

SCFP_6	-1971137145	 [*]\C=C\[c](:[*]):[*]	0.431	7 out of 8
SCFP_6	591469355	 [*][c](:[*]):[c](OC):[cH]:[*]	0.411	10 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1652428703	 [*]O[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1[*]	-0.438	1 out of 4
SCFP_6	-424485343	 [*]C(=[*])O	-0.328	19 out of 51
SCFP_6	-1379591900	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84



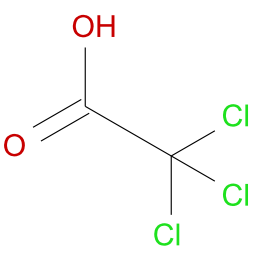
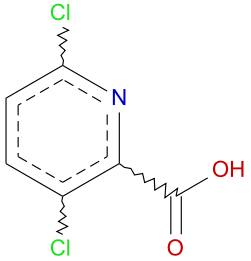
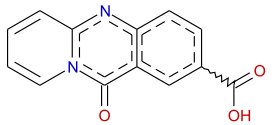
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.558
Enrichment: 1.06
Bayesian Score: 0.278
Mahalanobis Distance: 8.01
Mahalanobis Distance p-value: 0.602

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Trichloroacetic Acid	3;6-Dichloropicolinic Acid	11-oxo-11H-Pyrido(2;1-b)quinazoline-2-carboxylic Acid
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Toxic
Distance	0.490	0.519	0.565
Reference	Teratology 40:445-451; 1989	Fundam Appl Toxicol 4:91-97; 1984	Teratology 38(4):351-67; 1988

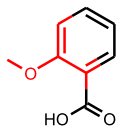
Model Applicability

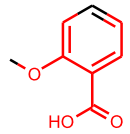
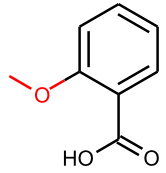
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

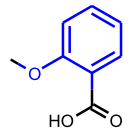
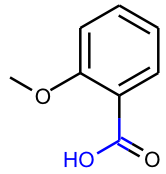
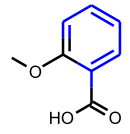
Feature Contribution

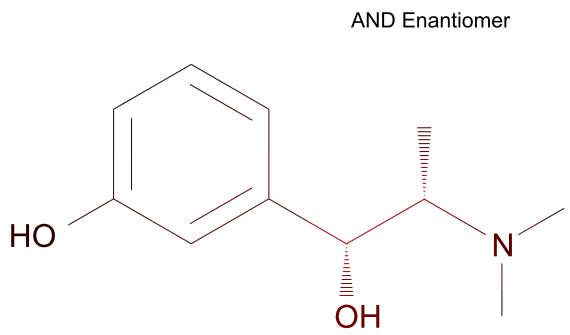
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	591469355	 [*][c](:[*]):[c](OC):[cH]:[*]	0.411	10 out of 12

SCFP_6	1627842774	 [*]O[c]1:[cH]:[*]:[cH]:[cH]:[c]:1C(=O)O	0.271	1 out of 1
SCFP_6	136239834	 [*]OC	0.242	23 out of 34

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1652428703	 [*]O[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1[*]	-0.438	1 out of 4
SCFP_6	-424485343	 [*]C(=[*])O	-0.328	19 out of 51
SCFP_6	-1379591900	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84



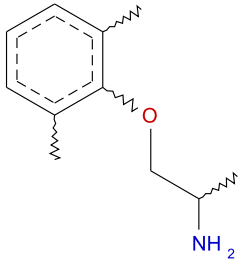
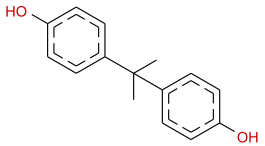
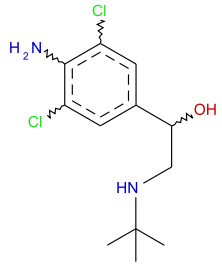
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.686
Enrichment: 1.3
Bayesian Score: 3.55
Mahalanobis Distance: 10.6
Mahalanobis Distance p-value: 0.00397

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Mexiletine .HCl (Free base form)	Bisphenol a	Clenbuterol
Structure			
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Toxic
Distance	0.546	0.586	0.599
Reference	Iyakuhin Kenkyu 14(4):550-570; 1983	Fundam Appl Toxicol 8:571-582; 1987	Iyakuhin Kenkyu 15(4):597-613; 1984

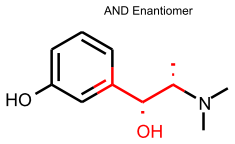
Model Applicability

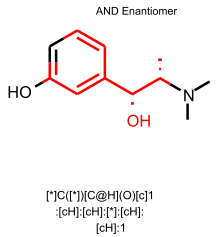
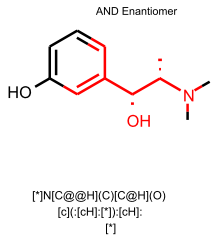
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

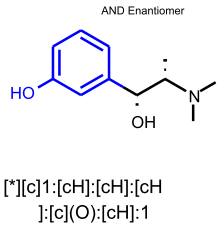
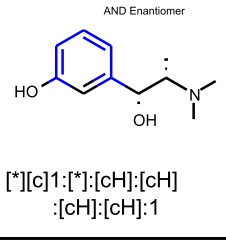
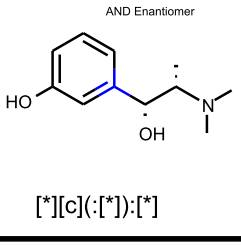
Feature Contribution

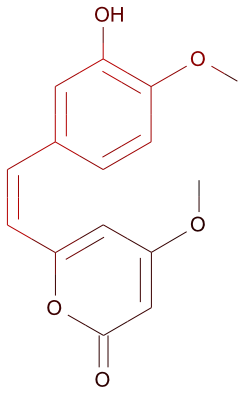
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1702724181	 [*]C([*])[C@H](O)[c](:[*]):[*]	0.558	9 out of 9

SCFP_6	-561151481		0.523	6 out of 6
SCFP_6	1318513260		0.478	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1434483901		-0.422	0 out of 1
SCFP_6	-1379591900		-0.282	33 out of 84
SCFP_6	3		0	92 out of 181



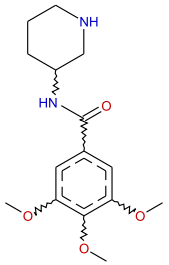
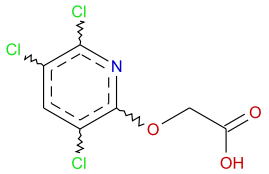
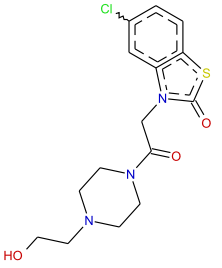
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.8
Enrichment: 1.52
Bayesian Score: 6.7
Mahalanobis Distance: 9.67
Mahalanobis Distance p-value: 0.0466

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	N-(3-Piperidyl)3;4;5-trimethoxybenzamide	Triclopyr	Tiaramide .HCl (Free base form)
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.573	0.580	0.587
Reference	Kiso to Rinsho 18:91-101; 1984	Fundam Appl Toxicol 4:872-882; 1984	Arzneimittelforschung 23(4):504-8; 1973

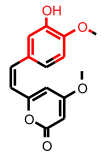
Model Applicability

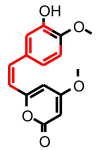
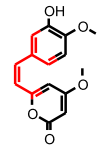
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

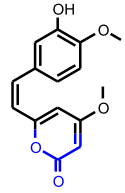
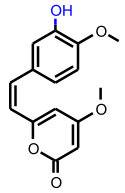
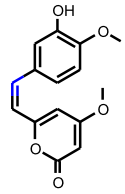
Feature Contribution

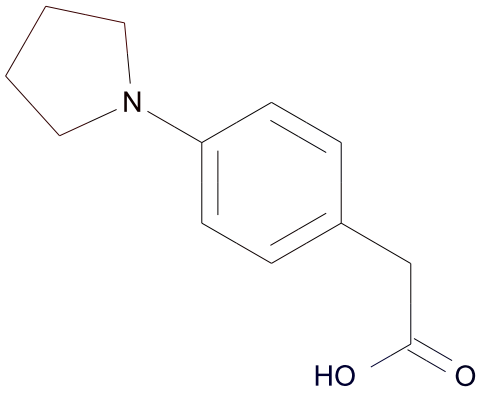
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	2116304939	 [*]O[c]1:[cH]:[*]:[c]([*]):[cH]:[c]:1O	0.504	5 out of 5

SCFP_6	-538866216	 [*]\C=C/[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.478	4 out of 4
SCFP_6	-1977229858	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	0.478	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1132907712	 [*]=CC(=O)O	-0.184	9 out of 21
SCFP_6	12	 [*]O	0	97 out of 178
SCFP_6	1	 [*]C(=[*])[*]	0	90 out of 173



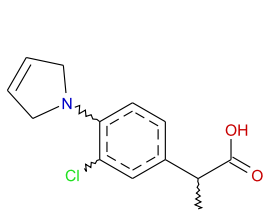
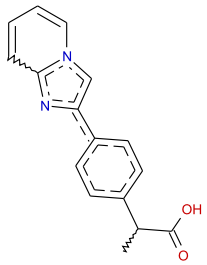
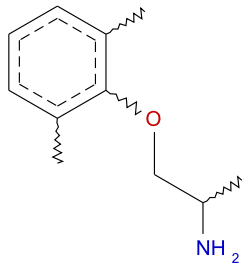
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.546
Enrichment: 1.04
Bayesian Score: -0.0369
Mahalanobis Distance: 7.92
Mahalanobis Distance p-value: 0.647

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Pirprofen	Miroprofen	Mexiletine .HCl (Free base form)
Structure			
Actual Endpoint	Non-Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Non-Toxic
Distance	0.482	0.515	0.528
Reference	Kiso to Rinsho 18:5651-5673; 1984	Iyakuhin Kenkyu 12:808-826; 1981	Iyakuhin Kenkyu 14(4):550-570; 1983

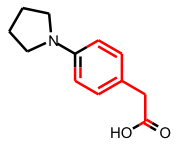
Model Applicability

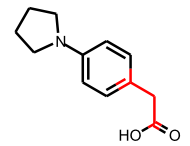
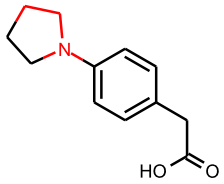
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC8 out of range. Value: 4.9801. Training min, max, SD, explained variance: -3.1563, 4.7297, 1.569, 0.0361.

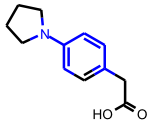
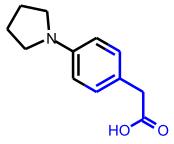
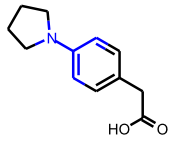
Feature Contribution

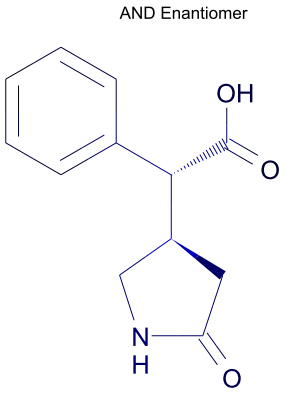
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1100167450	 [*]C(=[*])C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.441	3 out of 3

SCFP_6	470101049	 [*]C(=[*])C[c](:[*]):[*]	0.322	4 out of 5
SCFP_6	1155241219	 [*]N1[*][*]CC1	0.302	11 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.449	6 out of 19
SCFP_6	-1476721585	 [*]:[cH]:[c](CC(=O)O):[cH]:[*]	-0.422	0 out of 1
SCFP_6	1334669481	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.355	10 out of 28



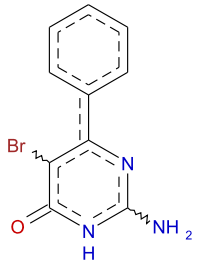
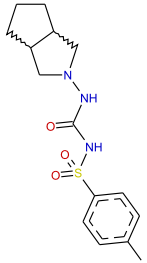
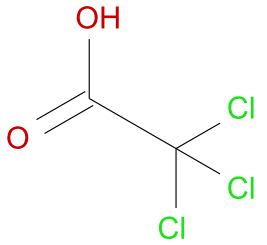
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.38
Enrichment: 0.723
Bayesian Score: -5.13
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.000431

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Bropirimine	Sulfonylurea Gliclazide	Trichloroacetic Acid
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Non-Toxic
Distance	0.581	0.624	0.627
Reference	Teratology 38(1):7-14; 1988	Yakuri to Chiryo 9:3551-3571; 1981	Teratology 40:445-451; 1989

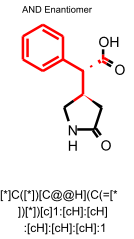
Model Applicability

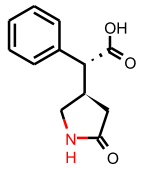
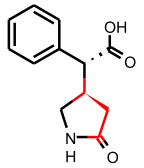
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

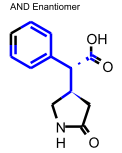
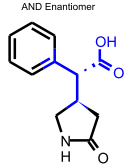
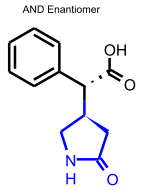
Feature Contribution

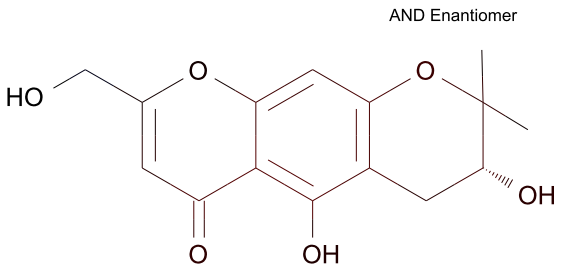
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1983581611	 [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH] :[cH]:[cH]:[cH]:1	0.271	1 out of 1

SCFP_6	9	AND Enantiomer  [*]N[*]	0.0928	45 out of 78
SCFP_6	-1272768868	AND Enantiomer  [*]C([*])CC(=[*])[*]	0.0607	14 out of 25

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1214451192	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.849	1 out of 7
SCFP_6	413428563	AND Enantiomer  [*]C([*])[C@@H](C(=O)O)[c](:[*]):[*]	-0.729	1 out of 6
SCFP_6	-1047009138	AND Enantiomer  [*][C@@H]1CNC(=O)C1	-0.718	0 out of 2



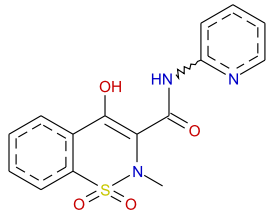
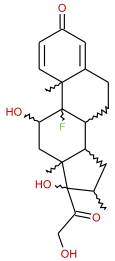
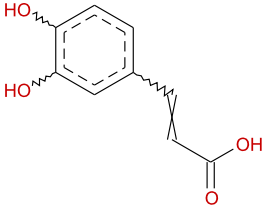
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Toxic
Probability: 0.656
Enrichment: 1.25
Bayesian Score: 2.79
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 3.28e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Piroxicam	Dexamethasone	Caffeic Acid
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.628	0.667	0.672
Reference	Yakuri to Chiryo 8:4655-4671; 1980	Preclin Rep Cent Inst Exp Anim 8:95-115; 1982	Toxicol Appl Pharmacol 36(2):227-37; 1976

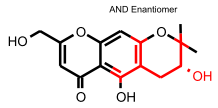
Model Applicability

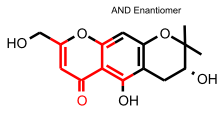
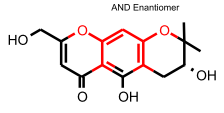
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

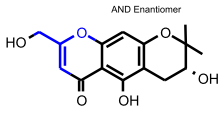
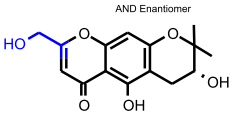
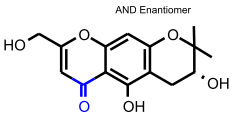
Feature Contribution

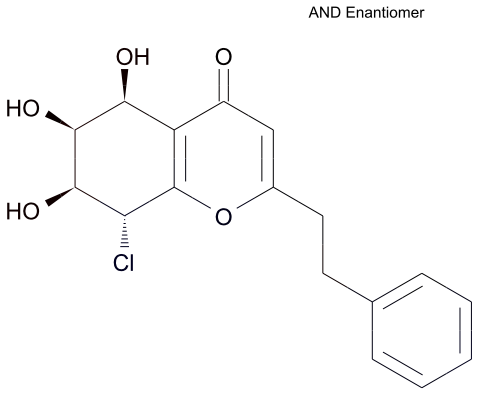
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1849894309	 [*][c](:[*]):[c]1C[C@@H](O)C([*])([*])[*][c]:1:[*]	0.381	2 out of 2

SCFP_6	-617610981	 [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*]]	0.381	2 out of 2
SCFP_6	130348166	 [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]:1[*]	0.369	5 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	616547045	 [*]C\C(=C\[*])\O[*]	-0.438	1 out of 4
SCFP_6	-711656408	 [*]C(=[*])CO	0	7 out of 13
SCFP_6	1311071855	 [*]C(=O)[*]	0	72 out of 141



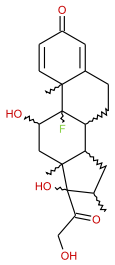
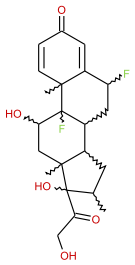
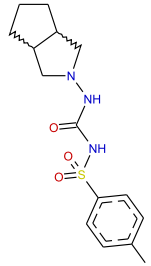
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Non-Toxic
Probability: 0.534
Enrichment: 1.02
Bayesian Score: -0.361
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 3.71e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Dexamethasone	Diflorasone Diacetate	Sulfonylurea Gliclazide
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.564	0.586	0.610
Reference	Preclin Rep Cent Inst Exp Anim 8:95-115; 1982	Oyo Yakuri 28(2):207-224; 1984	Yakuri to Chiryo 9:3551-3571; 1981

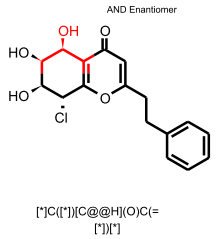
Model Applicability

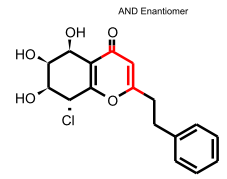
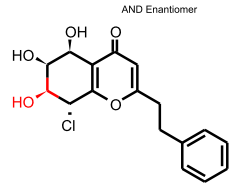
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

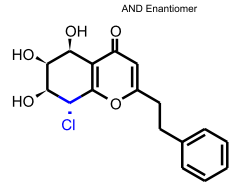
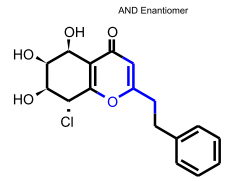
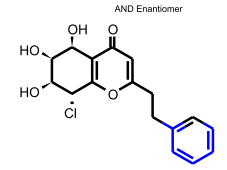
Feature Contribution

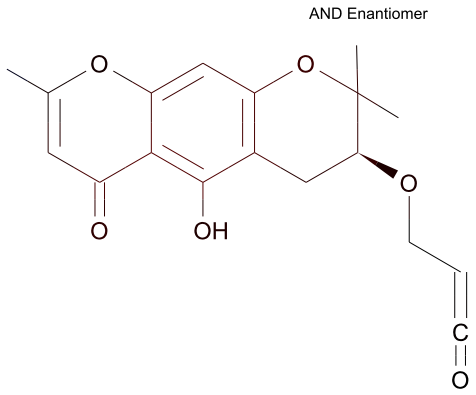
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1702664599	 [*]C([*])[C@@H](O)C(=[*])[*]	0.322	4 out of 5

SCFP_6	-1971196727	 [*]C(=CC(=[*])[*])[*]	0.293	13 out of 18
SCFP_6	-424515134	 [*]C([*])O	0.275	39 out of 56

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-52163885	 [*]C([*])Cl	-0.496	6 out of 20
SCFP_6	616547045	 [*]C\C(=C\[*])\O[*]	-0.438	1 out of 4
SCFP_6	-1379591900	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84



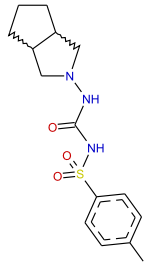
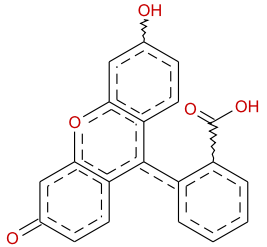
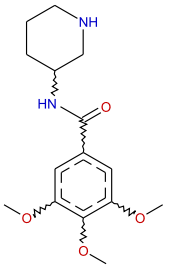
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.658
Enrichment: 1.25
Bayesian Score: 2.84
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 1.3e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Sulfonylurea Gliclazide	D&C Yellow 8	N-(3-Piperidyl)3;4;5-trimethoxybenzamide
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.539	0.551	0.613
Reference	Yakuri to Chiryo 9:3551-3571; 1981	Food Chem Toxicol 24:819-823; 1986	Kiso to Rinsho 18:91-101; 1984

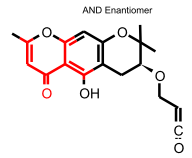
Model Applicability

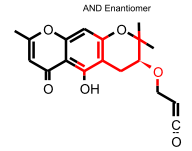
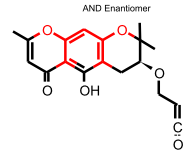
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

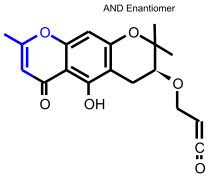
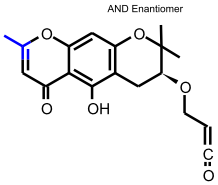
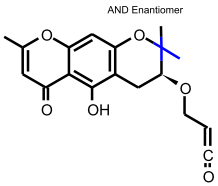
Feature Contribution

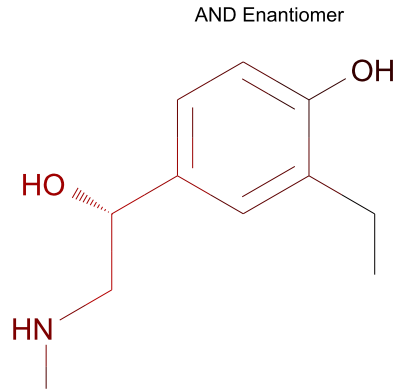
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-617610981	 [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*]	0.381	2 out of 2

SCFP_6	-1849894309	 [*][c](:[*]):[c]1C[C@ @H](O)C([*])([*])[*] [c]:1:[*]	0.381	2 out of 2
SCFP_6	130348166	 [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]: 1[*]	0.369	5 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	616547045	 [*]C\C(=C\[*])\O[*]	-0.438	1 out of 4
SCFP_6	136627117	 [*]C(=[*])C	-0.0885	10 out of 21
SCFP_6	136597326	 [*]C([*])C	0	50 out of 97



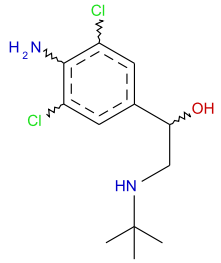
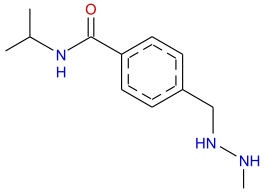
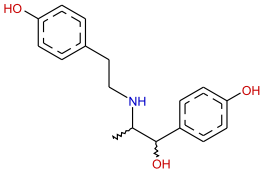
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Toxic
Probability: 0.744
Enrichment: 1.41
Bayesian Score: 5.08
Mahalanobis Distance: 9.04
Mahalanobis Distance p-value: 0.164

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Clenbuterol	Procarbazine .HCl (Free base form)	Ritodrine .HCl (Free base form)
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.491	0.539	0.567
Reference	Iyakuhin Kenkyu 15(4):597-613; 1984	Teratology 2(1):23-31; 1969	Kiso to Rinsho 19:2002-2018; 1985

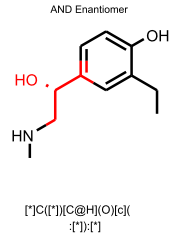
Model Applicability

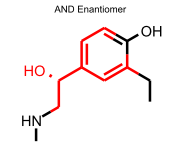
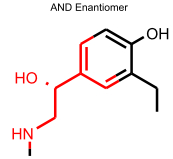
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

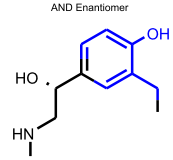
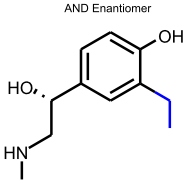
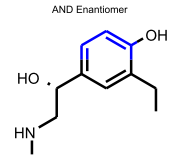
Feature Contribution

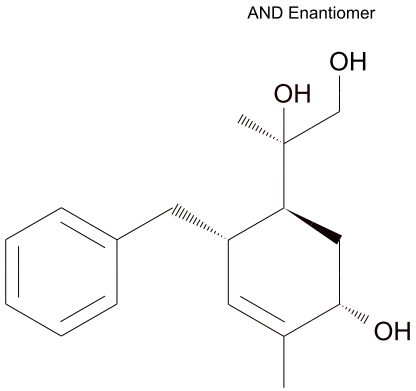
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1702724181	 [*]C([*])[C@H](O)[c](:[*]):[*]	0.558	9 out of 9

SCFP_6	-561151481	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	0.523	6 out of 6
SCFP_6	-1059390504	 [*]NC[C@H](O)[c](:[cH]:[*]):[cH]:[*]	0.441	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1931277081	 [*]C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1O	0	1 out of 2
SCFP_6	136597326	 [*]C([*])C	0	50 out of 97
SCFP_6	-496409612	 [*][c](:[*]):[cH]:[cH]:[*]	0	82 out of 163



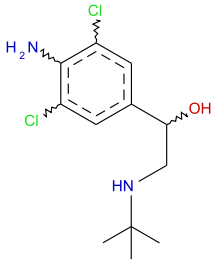
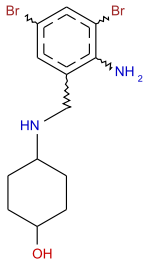
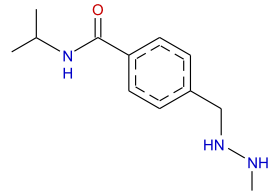
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Toxic
Probability: 0.573
Enrichment: 1.09
Bayesian Score: 0.671
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.00035

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Clenbuterol	Ambroxol	Procarbazine .HCl (Free base form)
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.468	0.608	0.610
Reference	Iyakuhin Kenkyu 15(4):597-613; 1984	Iyakuhin Kenkyu 12(1):371-387; 1981	Teratology 2(1):23-31; 1969

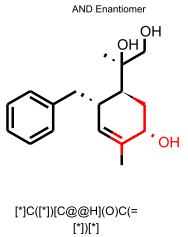
Model Applicability

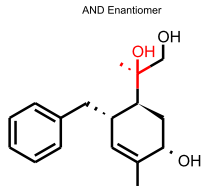
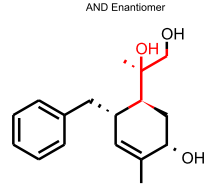
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC9 out of range. Value: -4.1153. Training min, max, SD, explained variance: -3.9182, 4.9981, 1.45, 0.0308.

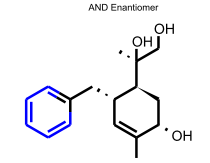
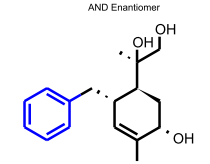
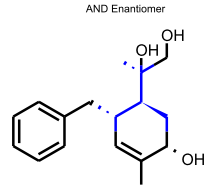
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1702664599	 [*]C([*])[C@@H](O)C(=[*])[*]	0.322	4 out of 5

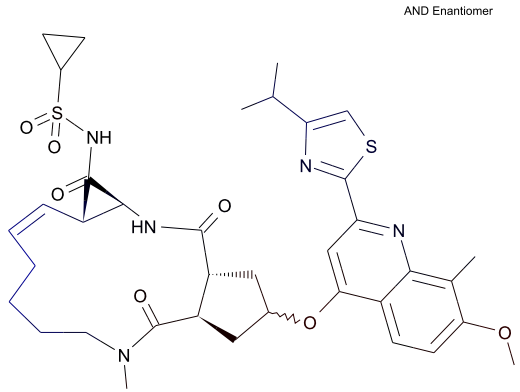
SCFP_6	-424515134	 [*]C([*])O	0.275	39 out of 56
SCFP_6	1416196903	 [*]OC(C)(C)C([*])[*]	0.255	3 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1379591900	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84
SCFP_6	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.188	12 out of 28
SCFP_6	-1043339860	 [*]C([*])C1C[*][*]C1	-0.114	13 out of 28

Simeprevir

TOPKAT_Developmental_Toxicity_Potential



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: Non-Toxic

Probability: 0.438

Enrichment: 0.833

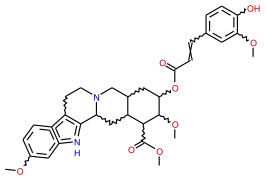
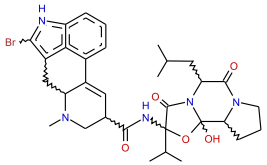
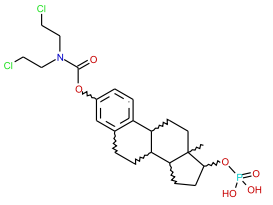
Bayesian Score: -3.14

Mahalanobis Distance: 13.4

Mahalanobis Distance p-value: 9.2e-008

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Reserpate	Bromocriptine	Estramustine Phosphate Disodium (Free acid form)
Structure			
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic
Distance	0.675	0.778	0.819
Reference	Oyo Yakuri 18:105-124; 1979	Toxicol Lett 50:189-194; 1990	Oyo Yakuri 20(6):1219-1236; 1980

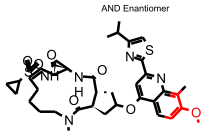
Model Applicability

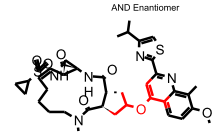
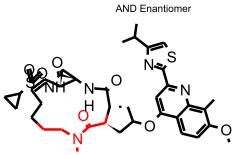
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: 3.5244. Training min, max, SD, explained variance: -3.7514, 3.3159, 1.318, 0.0255.

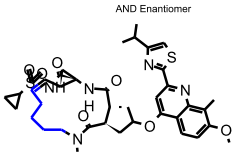
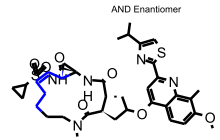
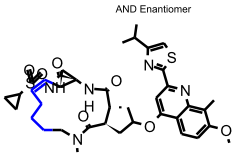
Feature Contribution

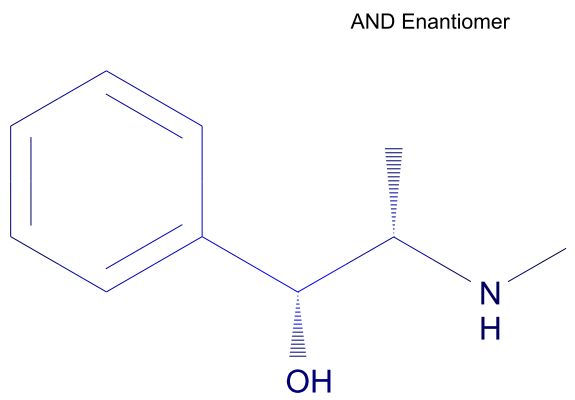
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	591469355	 [*][c](:[*]):[c](OC):[cH]:[*]	0.411	10 out of 12

SCFP_6	-395254381	 [*]:[cH]:[c](OC1C[*][*]C1):[c](:[*]):[*]	0.271	1 out of 1
SCFP_6	-109092631	 [*]CCN(C)C(=O)C([*])[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-98332825	 [*]CCCCC=[*]	-0.718	0 out of 2
SCFP_6	-1476112164	 [*]CC\C=C/C1[*][*]1	-0.718	0 out of 2
SCFP_6	1260369147	 [*]CCC\C=C/[*]	-0.718	0 out of 2



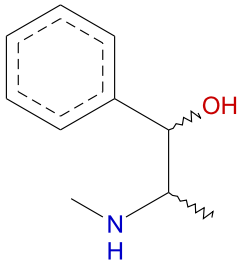
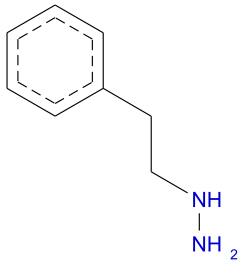
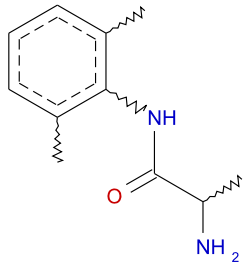
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.207
Enrichment: 0.647
Bayesian Score: -4.88
Mahalanobis Distance: 9.32
Mahalanobis Distance p-value: 0.749

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Phenelzine	Tocainide
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.000	0.523	0.550
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

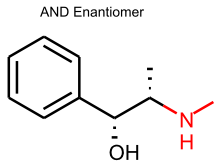
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

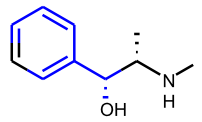
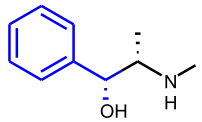
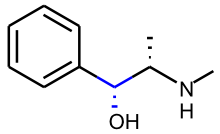
Feature Contribution

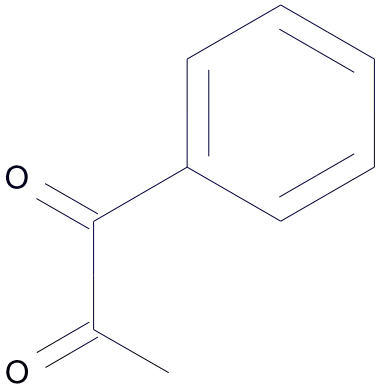
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	864287155	 [*]NC	0.2	4 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-388186450	AND Enantiomer  [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.857	1 out of 12
ECFP_6	2014710090	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.716	1 out of 10
ECFP_6	-1910270391	AND Enantiomer  [*]C([*])[*]	-0.307	20 out of 89



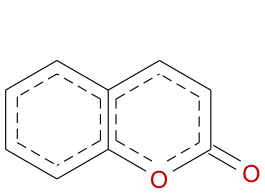
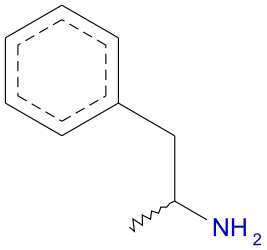
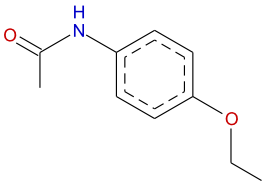
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.237
Enrichment: 0.739
Bayesian Score: -1.53
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.0935

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Coumarin	Amphetamine	Phenacetin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.570	0.576	0.580
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

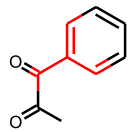
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

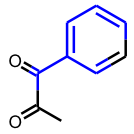
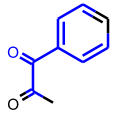
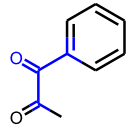
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -472613004: [*]C(=[*])C(=O)C

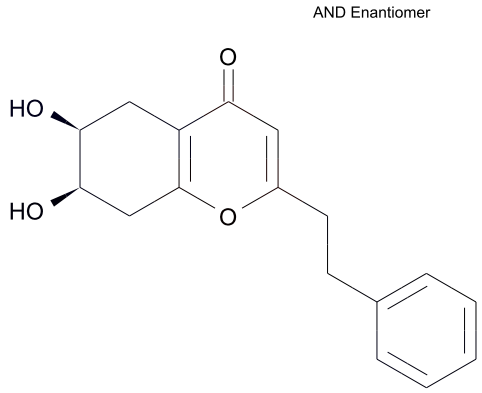
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.171	9 out of 24

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1451403962	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.459	1 out of 7
ECFP_6	169261700	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.27	0 out of 1
ECFP_6	1432101658	 [*]C(=[*])C(=O)[c](:[*]):[*]	-0.27	0 out of 1



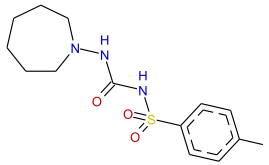
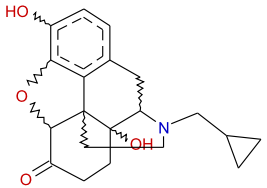
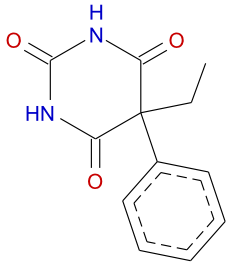
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.26
Enrichment: 0.811
Bayesian Score: -0.148
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.282

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Tolazamide	Naltrexone	Phenobarbital
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.557	0.593	0.614
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

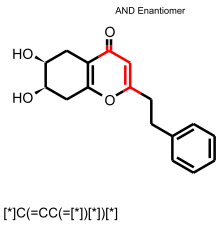
Model Applicability

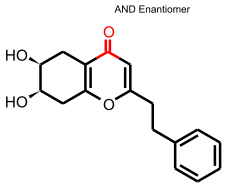
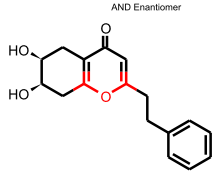
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1794461805: [*]CC(=C([*])[*])C(=[*])[*]
3. Unknown ECFP_2 feature: 1650944136: [*]CC(=C([*])[*])O[*]

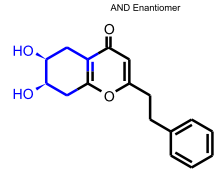
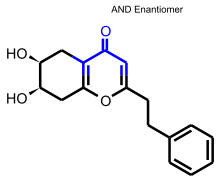
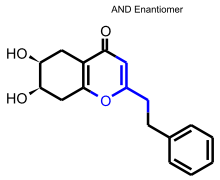
Feature Contribution

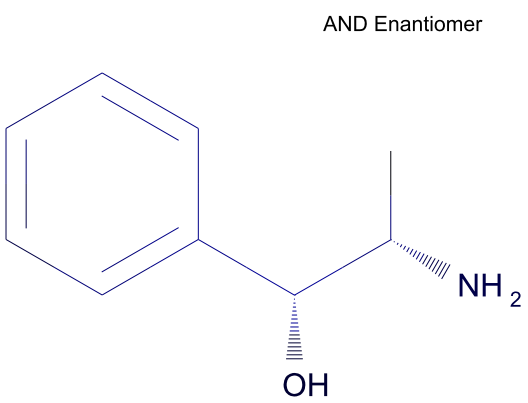
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	464808839	 [*]C(=CC(=[*])[*])[*]	0.524	8 out of 14

ECFP_6	2106656448	 [*]C(=O)[*]	0.254	31 out of 77
ECFP_6	-560785749	 [*]C(=[*])OC(=[*])[*]	0.212	1 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	571598839	 [*]C1=[*]C[C@@H](O)[C@@H](O)C1	-0.27	0 out of 1
ECFP_6	1299558496	 [*]C(=[*])C(=O)C=[*]	-0.27	0 out of 1
ECFP_6	1875238785	 [*]C\C(=C\[*])\O[*]	-0.27	0 out of 1



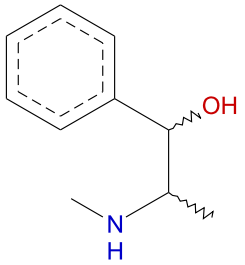
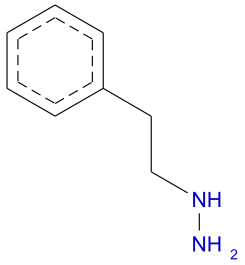
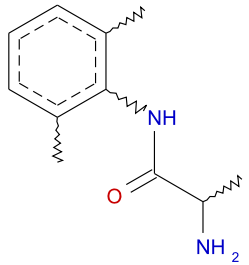
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.213
Enrichment: 0.665
Bayesian Score: -3.84
Mahalanobis Distance: 9.78
Mahalanobis Distance p-value: 0.547

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Phenelzine	Tocainide
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.391	0.479	0.503
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

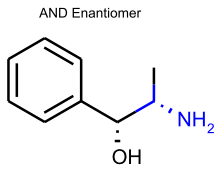
Model Applicability

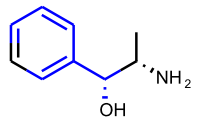
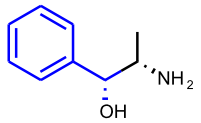
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

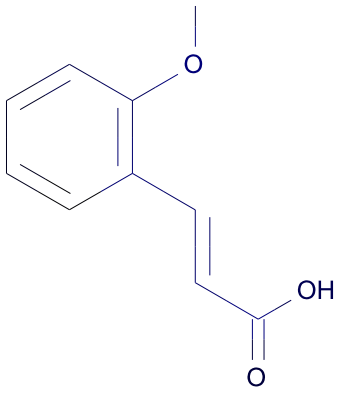
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -261056708: [*]C([*])[C@H](C)N

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-933808133	 [*]C([*])N	-1.33	0 out of 9

ECFP_6	-388186450	AND Enantiomer  [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.857	1 out of 12
ECFP_6	2014710090	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.716	1 out of 10



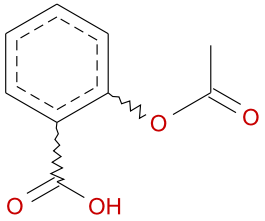
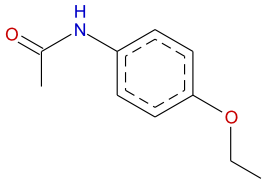
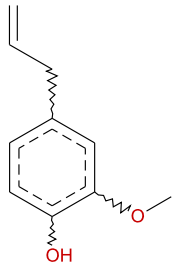
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.205
Enrichment: 0.641
Bayesian Score: -5.59
Mahalanobis Distance: 8.3
Mahalanobis Distance p-value: 0.976

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aspirin	Phenacetin	Eugenol
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.527	0.555	0.560
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

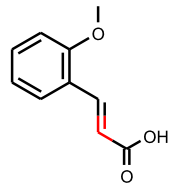
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

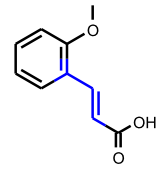
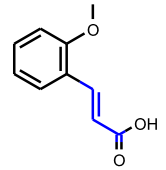
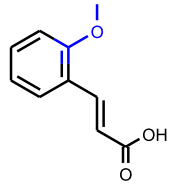
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1335702447: [*][c](:[*]):[c](C=[*]):c:[*]

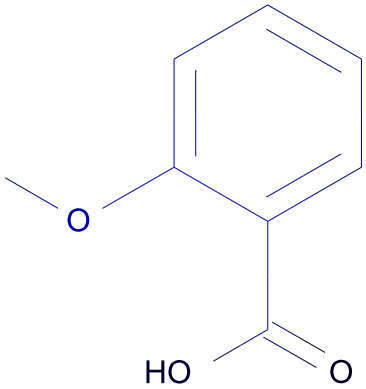
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1925046727	 [*]C=[*]	0.391	11 out of 23

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1831055759	 [*]\C=C\[c](:[*]):[*]	-0.805	0 out of 4
ECFP_6	-470416293	 [*]\C=C\C(=[*])[*]	-0.657	0 out of 3
ECFP_6	1307307440	 [*]:[c](:[*])OC	-0.558	4 out of 25



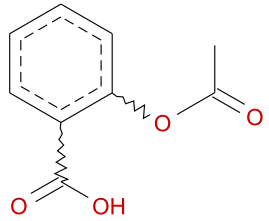
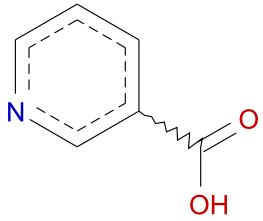
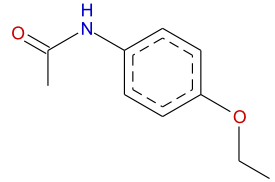
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.207
Enrichment: 0.647
Bayesian Score: -4.95
Mahalanobis Distance: 7.82
Mahalanobis Distance p-value: 0.996

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aspirin	Nicotinic acid	Phenacetin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.348	0.569	0.592
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

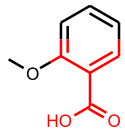
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

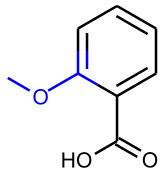
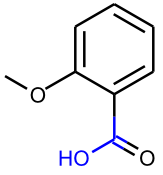
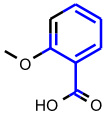
Feature Contribution

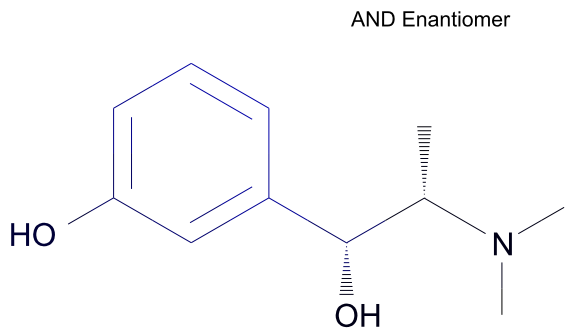
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	989674687	 [*][c](:[*]):[c](:[cH]:[*])C(=O)O	0.164	2 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	1307307440	 [*]:[c](:[*])OC	-0.558	4 out of 25
ECFP_6	2025485523	 [*]C(=[*])O	-0.506	8 out of 45
ECFP_6	1635415905	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-0.482	0 out of 2



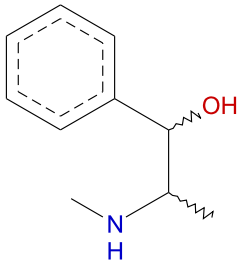
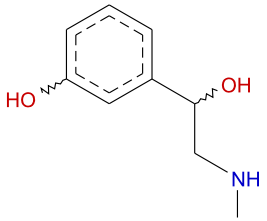
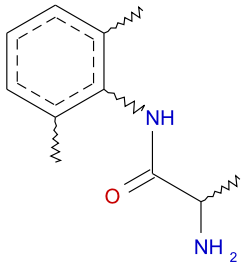
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.208
Enrichment: 0.65
Bayesian Score: -4.66
Mahalanobis Distance: 12
Mahalanobis Distance p-value: 0.0119

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Phenylephrine	Tocainide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.486	0.513	0.557
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

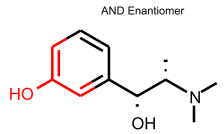
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

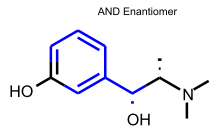
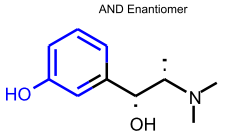
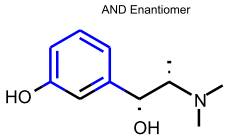
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1280034494: [*]C([*])[C@H](C)N([*])[*]

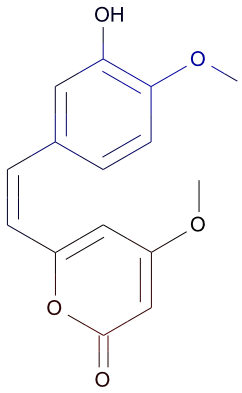
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-177786161	 [*]:[cH]:[c](O):[cH]:[*]	0.406	7 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-388186450	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.857	1 out of 12
ECFP_6	-783815036	 O[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.657	0 out of 3
ECFP_6	2007300961	 [*][c]1:[*]:[c]([*]):[cH]:[cH]:[cH]:1	-0.652	5 out of 34



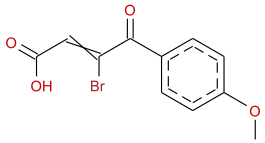
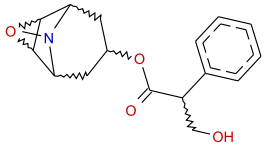
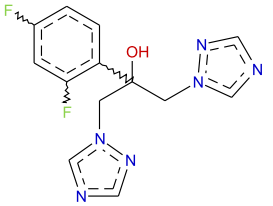
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.219
Enrichment: 0.683
Bayesian Score: -3.08
Mahalanobis Distance: 10
Mahalanobis Distance p-value: 0.435

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Cytembena	Scopolamine	Fluconazole
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.535	0.593	0.619
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

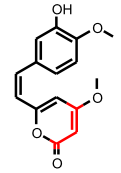
Model Applicability

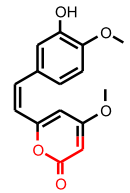
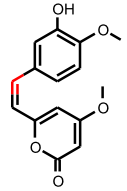
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC13 out of range. Value: 5.2896. Training min, max, SD, explained variance: -3.9176, 5.0348, 1.522, 0.0236.
2. Unknown ECFP_2 feature: -176483725: [*]=C[c](:c:[*]):c:[*]
3. Unknown ECFP_2 feature: -1053980253: [*]O\C(=C/[*])\C=[*]
4. Unknown ECFP_2 feature: -444332269: [*]O\C(=C\[*])\C=[*]

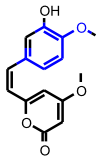
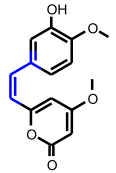
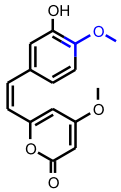
Feature Contribution

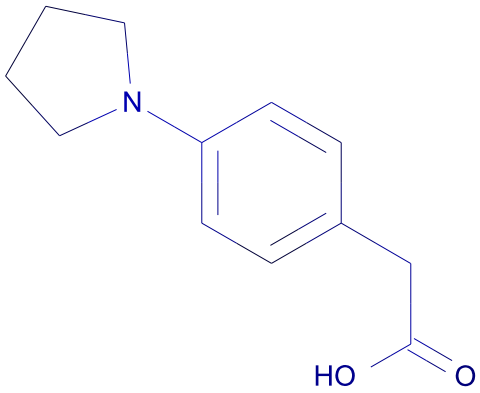
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	464808839	 [*]C(=CC(=[*])[*])[*]	0.524	8 out of 14

ECFP_6	-1885846789	 [*]OC(=O)C=[*]	0.424	1 out of 1
ECFP_6	-1925046727	 [*]C=[*]	0.391	11 out of 23

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	2077607946	 [*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-1.15	0 out of 7
ECFP_6	-1831055759	 [*]\C=C\[c](:[*]):[*]	-0.805	0 out of 4
ECFP_6	1307307440	 [*]:[c](:[*])OC	-0.558	4 out of 25



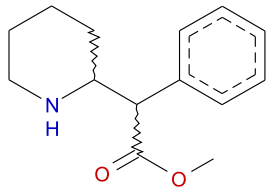
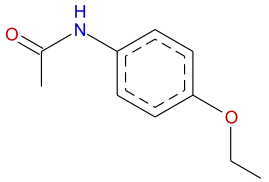
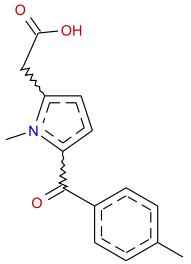
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.207
Enrichment: 0.645
Bayesian Score: -7.56
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 0.304

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Methylphenidate	Phenacetin	Tolmetin
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.535	0.564	0.566
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

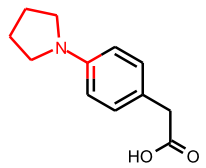
Model Applicability

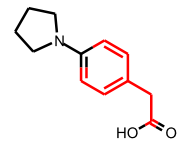
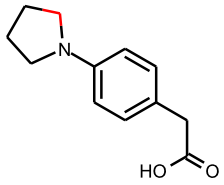
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

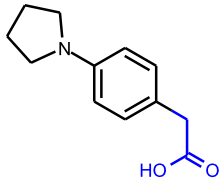
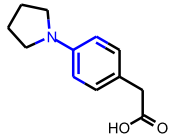
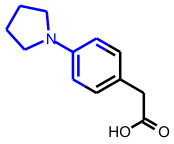
Feature Contribution

Top features for positive contribution

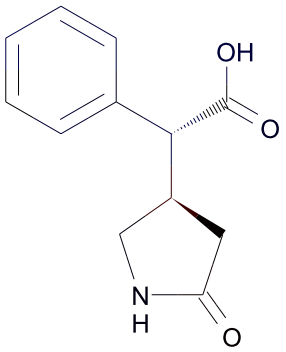
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	341335091	 [*]:[c](:[*])N1CCCC1	0.212	1 out of 2

ECFP_6	1737023319	 [*]C(=[*])C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.212	1 out of 2
ECFP_6	-992506539	 [*]C[*]	0.073	44 out of 132

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1731135544	 [*]CC(=O)O	-1.55	0 out of 12
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.805	0 out of 4
ECFP_6	-98352723	 [*]:[cH]:[c](:[cH]:[*])N1CCCC1	-0.657	0 out of 3

AND Enantiomer



$C_{12}H_{13}NO_3$

Molecular Weight: 219.237

ALogP: 0.621

Rotatable Bonds: 3

Acceptors: 3

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.228

Enrichment: 0.71

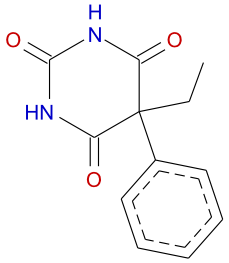
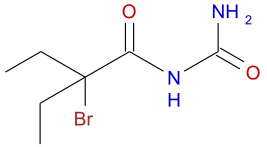
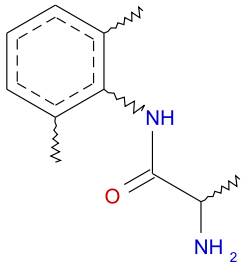
Bayesian Score: -2.25

Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.255

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenobarbital	Carbromal	Tocainide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.517	0.597	0.598
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

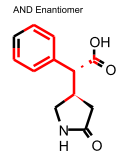
Model Applicability

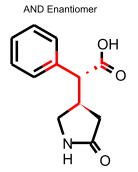
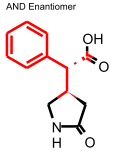
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -858846751: [*]C([*])C1C[*][*]C1

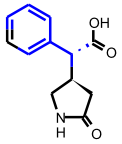
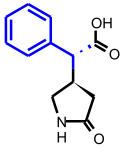
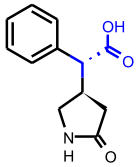
Feature Contribution

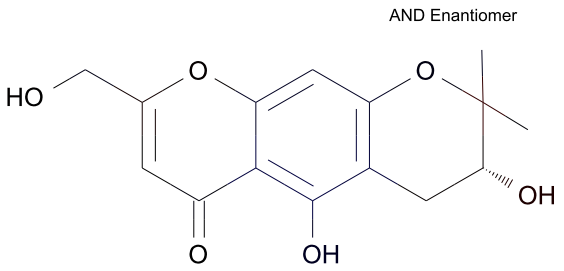
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-622605643	 [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH] :[*]:[cH]:[cH]:1	0.424	1 out of 1

ECFP_6	1603312431	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]	0.424	1 out of 1
ECFP_6	-1135409258	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH] :[cH]:[cH]:[cH]:1	0.424	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-388186450	AND Enantiomer  [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.857	1 out of 12
ECFP_6	2014710090	AND Enantiomer  [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.716	1 out of 10
ECFP_6	-1905455774	AND Enantiomer  [*]C([*])C(=O)O	-0.552	1 out of 8



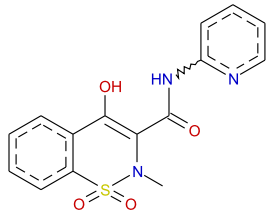
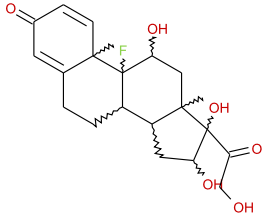
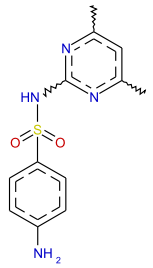
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Carcinogen
Probability: 0.27
Enrichment: 0.843
Bayesian Score: 0.352
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 0.0209

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Piroxicam	Triamcinolone	Sulfamethazine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.664	0.677	0.685
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

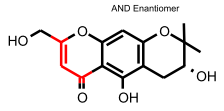
Model Applicability

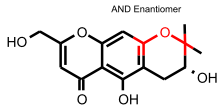
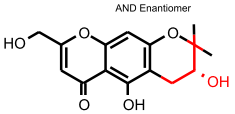
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

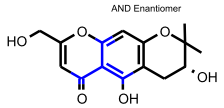
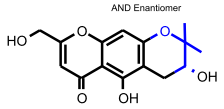
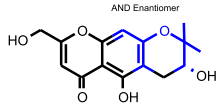
Feature Contribution

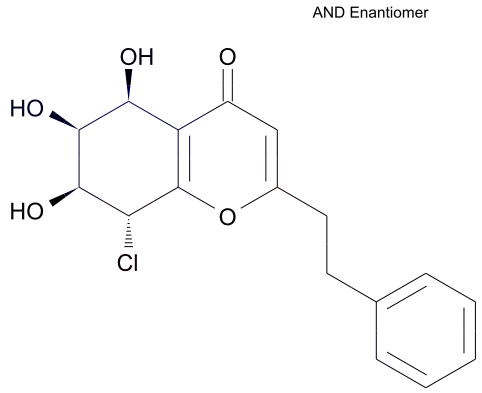
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	464808839	 [*]C(=CC(=[*])[*])[*]	0.524	8 out of 14

ECFP_6	-200406221	 [*]C([*])([*])O[c](:[*]):[*]	0.451	3 out of 5
ECFP_6	-1051556861	 [*]C[C@@H](O)C([*])([*])[*]	0.442	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-813997308	 [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.452	2 out of 12
ECFP_6	1778376725	 [*]OC(C)(C)C([*])[*]	-0.27	0 out of 1
ECFP_6	-1679469954	 [*][C@@H]1[*][c](:[*]):[c](OC1(C)C):[cH]:[*]	-0.27	0 out of 1



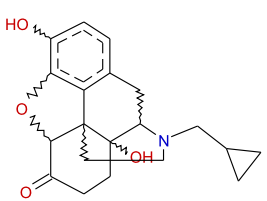
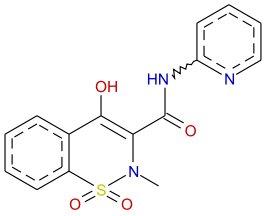
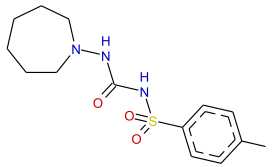
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.236
Enrichment: 0.735
Bayesian Score: -1.62
Mahalanobis Distance: 13.8
Mahalanobis Distance p-value: 3.11e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Naltrexone	Piroxicam	Tolazamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.632	0.660	0.661
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

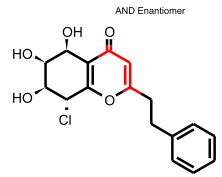
Model Applicability

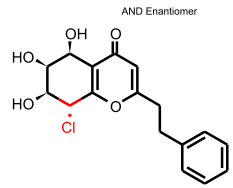
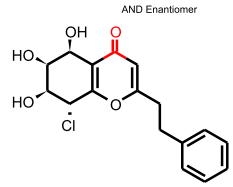
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -7106223: [*]C([*])[C@@H](Cl)C(=[*])[*]
3. Unknown ECFP_2 feature: 1652274794: [*]OC(=C([*])[*])C([*])[*]

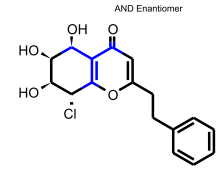
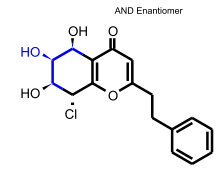
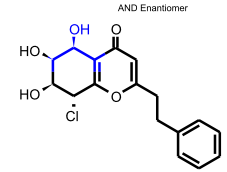
Feature Contribution

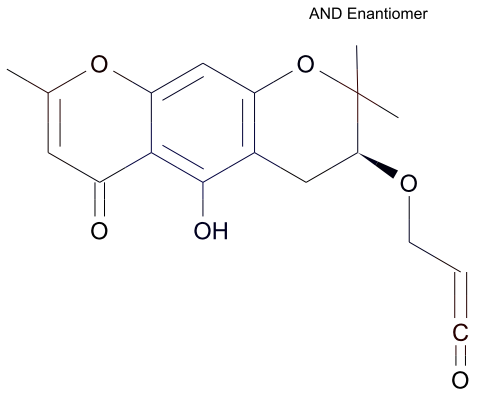
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	464808839	 [*]C(=CC(=[*])[*])[*]	0.524	8 out of 14

ECFP_6	105634199	 [*]C([*])Cl	0.424	1 out of 1
ECFP_6	2106656448	 [*]C(=O)[*]	0.254	31 out of 77

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1795792463	 [*]C([*])C(=C([*])[*])C(=[*])[*]	-0.657	0 out of 3
ECFP_6	305695353	 [*]C([*])C(O)C([*])[*]	-0.638	1 out of 9
ECFP_6	-1567907747	 [*]C([*])[C@@H](O)C(=[*])[*]	-0.27	0 out of 1



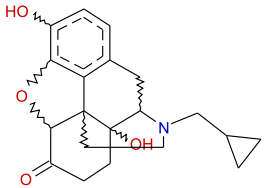
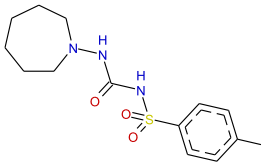
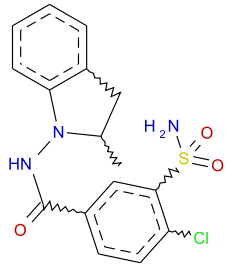
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.273
Enrichment: 0.853
Bayesian Score: 0.488
Mahalanobis Distance: 13.7
Mahalanobis Distance p-value: 4.68e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Naltrexone	Tolazamide	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.538	0.567	0.596
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

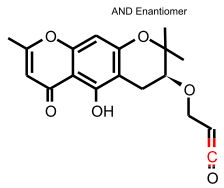
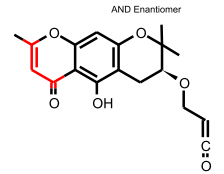
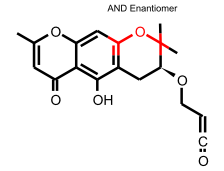
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -331149802: [*]O\C(=C/[*])\C
3. Unknown ECFP_2 feature: -1591590376: [*]C=C=O
4. Unknown ECFP_2 feature: 2106995136: [*]=C=O
5. Unknown ECFP_2 feature: -91536905: [*]CC=C=[*]
6. Unknown ECFP_2 feature: -1688150664: [*]OCC=[*]

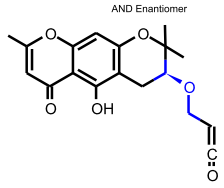
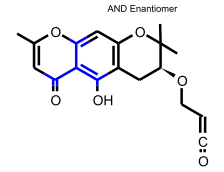
Feature Contribution

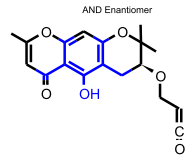
Top features for positive contribution

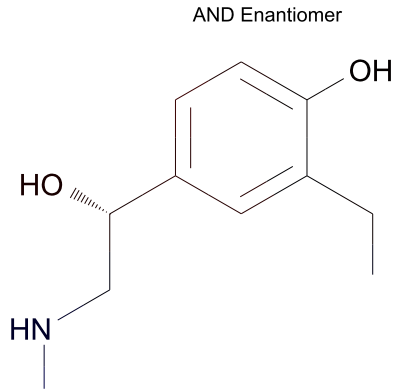
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-1114776580	 [*]=C=[*]	0.755	11 out of 15
ECFP_6	464808839	 [*]C(=CC(=[*])[*])[*]	0.524	8 out of 14
ECFP_6	-200406221	 [*]C([*])([*])O[c](:[*]):[*]	0.451	3 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1250019913	 [*]COC([*])[*]	-0.482	0 out of 2
ECFP_6	-813997308	 [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.452	2 out of 12

ECFP_6	-247712724	 [*]C[c]1:[c]([*]):[*] :[c]([*]):[c](C(=[*])[*]):[c]:1O	-0.27	0 out of 1



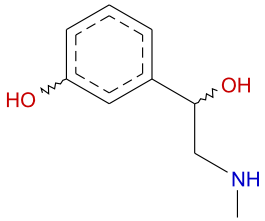
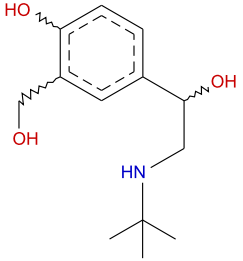
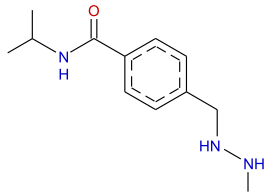
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.229
Enrichment: 0.713
Bayesian Score: -2.17
Mahalanobis Distance: 8.94
Mahalanobis Distance p-value: 0.875

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenylephrine	Albuterol	Procarbazine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.412	0.474	0.553
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

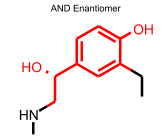
Model Applicability

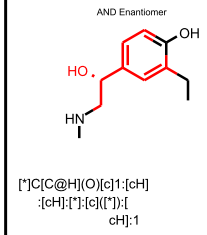
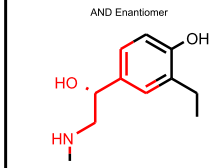
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

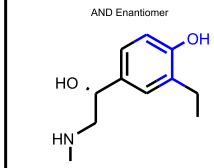
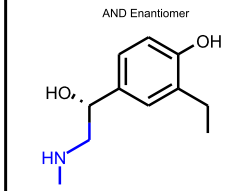
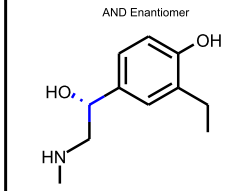
Feature Contribution

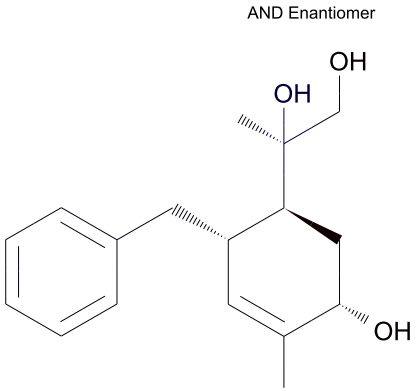
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-566124999	 [*]C[C@H](O)[c]1:[cH]:[cH]:[c](O):[c]([*]):[cH]:1	0.442	2 out of 3

ECFP_6	-1438409867	 [*]C[C@H](O)[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.337	3 out of 6
ECFP_6	628100036	 [*]NC[C@H](O)[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1334400011	 [*][c](:[*]):[c](O):[cH]:[*]	-0.496	3 out of 18
ECFP_6	493154328	 [*]CNC	-0.482	0 out of 2
ECFP_6	-1910270391	 [*]C([*])[*]	-0.307	20 out of 89



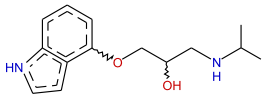
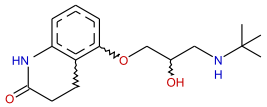
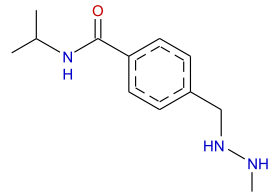
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.236
Enrichment: 0.735
Bayesian Score: -1.62
Mahalanobis Distance: 12.2
Mahalanobis Distance p-value: 0.00696

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Pindolol	Carteolol	Procarbazine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.598	0.623	0.637
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

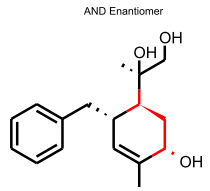
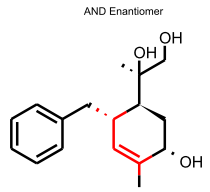
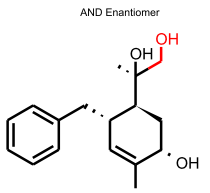
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC18 out of range. Value: 4.5259. Training min, max, SD, explained variance: -3.8038, 4.1085, 1.337, 0.0182.
2. Unknown ECFP_2 feature: -2097867909: [*]C[C@H](O)C(=[*])[*]
3. Unknown ECFP_2 feature: -327548242: [*]C[C@@H](C([*])[*])C([*])([*])[*]
4. Unknown ECFP_2 feature: -1263967621: [*]C[C@@H](C=[*])C([*])[*]
5. Unknown ECFP_2 feature: -1042330089: [*]\C=C(\C)/C([*])[*]
6. Unknown ECFP_2 feature: 1280892564: [*]C[C@](C)(O)C([*])[*]
7. Unknown ECFP_2 feature: -1907755304: [*]C([*])([*])CO
8. Unknown ECFP_2 feature: 771121623: [*]C([*])C[c](:[*]):[*]

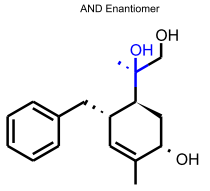
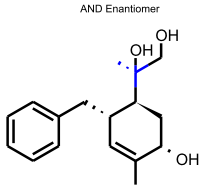
Feature Contribution

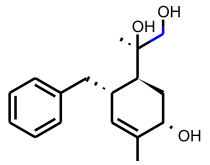
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-801490360	 [*]C([*])CC([*])[*]	0.297	12 out of 28
ECFP_6	470495651	 [*]C([*])C=C([*])[*]	0.2	4 out of 10
ECFP_6	2022454958	 [*]CO	0.135	9 out of 25

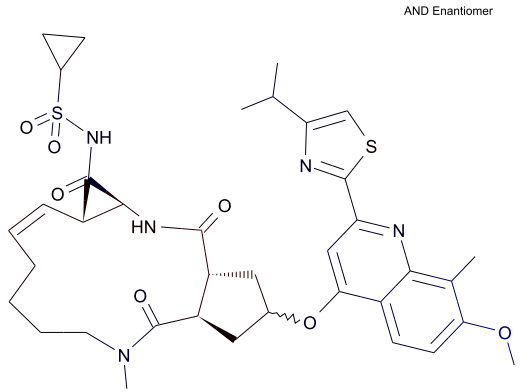
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	2025123907	 [*]C([*])([*])O	-1.05	0 out of 6
ECFP_6	-1085223908	 [*]C([*])([*])[*]	-0.168	9 out of 35

ECFP_6	1559650422	AND Enantiomer  [*]C[*]	-0.164	50 out of 191

Simeprevir

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.207

Enrichment: 0.646

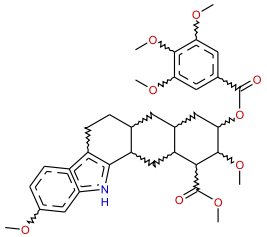
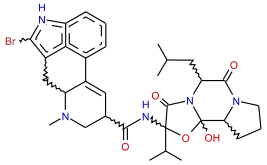
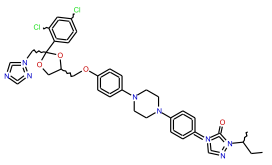
Bayesian Score: -4.97

Mahalanobis Distance: 15.4

Mahalanobis Distance p-value: 4.64e-008

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Reserpine	Bromocriptine	Itraconazole
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.783	0.847	0.870
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

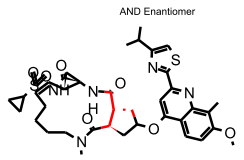
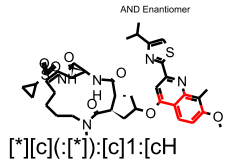
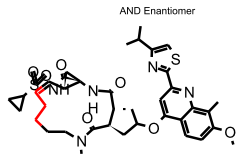
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

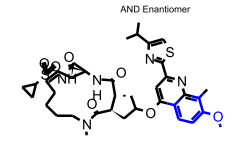
1. OPS PC4 out of range. Value: 5.1697. Training min, max, SD, explained variance: -4.7116, 4.7287, 2.103, 0.0450.
2. OPS PC14 out of range. Value: -3.7168. Training min, max, SD, explained variance: -3.5274, 4.3994, 1.457, 0.0216.
3. OPS PC17 out of range. Value: 3.3515. Training min, max, SD, explained variance: -4.7306, 3.3103, 1.364, 0.0189.
4. Unknown ECFP_2 feature: -309676563: [*]N[C@]1(C[C@H]1[*])C(=[*])[*]
5. Unknown ECFP_2 feature: -1049290660: [*]C1([*])C[C@H]1C=[*]
6. Unknown ECFP_2 feature: 360408239: [*]C\C=C/[*]
7. Unknown ECFP_2 feature: 1616402542: [*]CN(C)C(=[*])[*]
8. Unknown ECFP_2 feature: -1818486371: [*]NC(=O)C1([*])[*][*]1
9. Unknown ECFP_2 feature: 1310940530: [*]C([*])O[c](:[*]):[*]
10. Unknown ECFP_2 feature: -253227249: [*]:[c](:[*])[c]1:n:[*]:[*]:s:1
11. Unknown ECFP_2 feature: 733491677: [*]:[c](:[*])C(C)C
12. Unknown ECFP_2 feature: 2121941848: [*]NS(=O)(=O)C1[*][*]1
13. Unknown ECFP_2 feature: -622223421: [*]S(=[*])(=[*])C1CC1

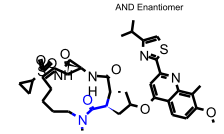
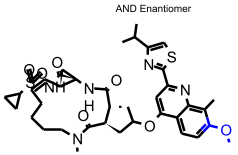
Feature Contribution

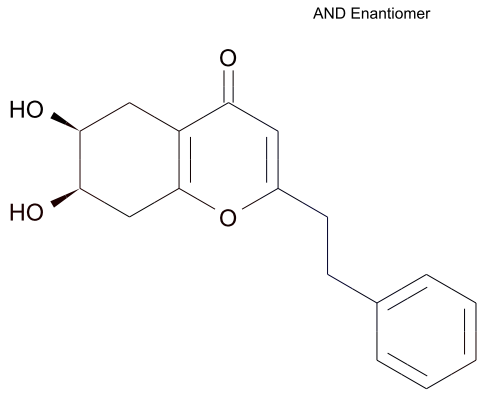
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-2095963820	 [*][C@@H]1[*][*]C[C@H]1C(=[*])[*]	0.891	12 out of 14
ECFP_6	2082767335	 [*][c](:[*]):[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1:[*]	0.617	2 out of 2
ECFP_6	-1331088410	 [*]CCC=[*]	0.442	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	2077607946	 [*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-1.15	0 out of 7

ECFP_6	1526862590	 [*]C([*])C(=O)N([*])[*]	-0.638	1 out of 9
ECFP_6	1307307440	 [*]:[c](:[*])OC	-0.558	4 out of 25



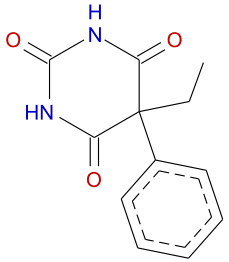
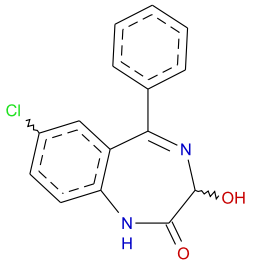
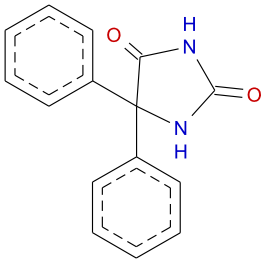
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.419
Enrichment: 1.02
Bayesian Score: 0.425
Mahalanobis Distance: 8.75
Mahalanobis Distance p-value: 0.113

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenobarbital	Oxazepam	Phenytoin
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.601	0.605	0.635
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

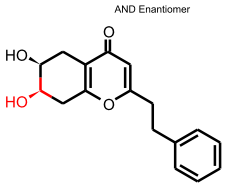
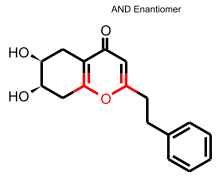
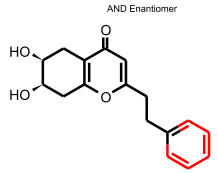
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1875238785: [*]C\C(=C\[*])\O[*]
3. Unknown ECFP_2 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
4. Unknown ECFP_2 feature: 1794461805: [*]CC(=C([*])[*])C(=[*])[*]
5. Unknown ECFP_2 feature: 1650944136: [*]CC(=C([*])[*])O[*]

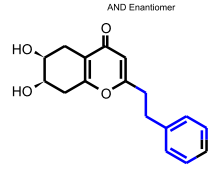
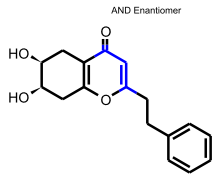
Feature Contribution

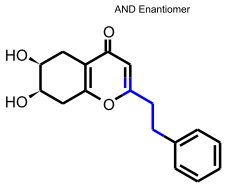
Top features for positive contribution

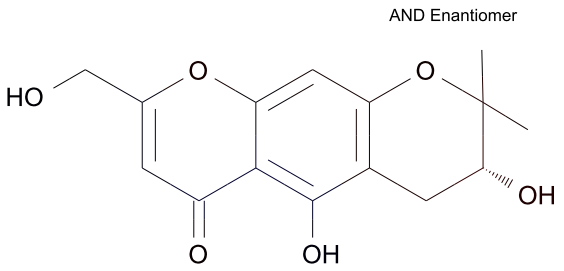
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set

ECFP_4	2024749573	 [*]C([*])O	0.494	8 out of 11
ECFP_4	-560785749	 [*]C(=[*])OC(=[*])[*]	0.351	1 out of 1
ECFP_4	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.283	9 out of 16

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	1205550831	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.545	1 out of 6
ECFP_4	464808839	 [*]C(=CC(=[*])[*])[*]	-0.352	2 out of 8

ECFP_4	-1795525632	 [*]CCC(=[*])[*]	-0.352	2 out of 8



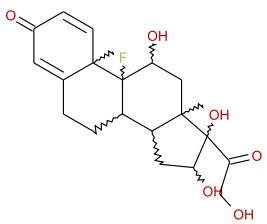
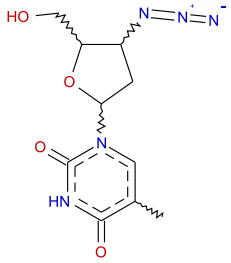
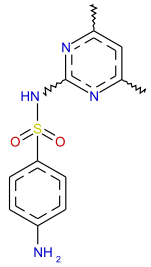
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.405
Enrichment: 0.987
Bayesian Score: 0.0887
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.00148

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Triamcinolone	Zidovudine	Sulfamethazine
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.670	0.672	0.704
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

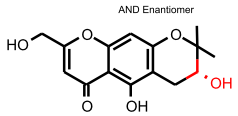
Model Applicability

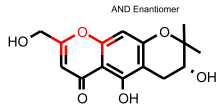
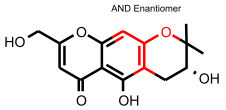
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1875238785: [*]C\C(=C\[*])\O[*]
3. Unknown ECFP_2 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
4. Unknown ECFP_2 feature: 1778376725: [*]OC(C)(C)C([*])[*]

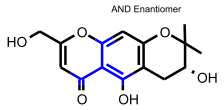
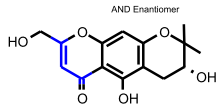
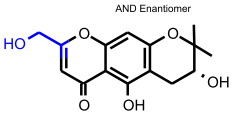
Feature Contribution

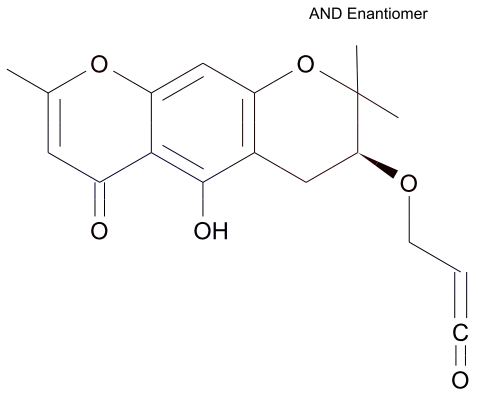
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	2024749573	 [*]C([*])O	0.494	8 out of 11

ECFP_4	-560785749	 [*]C(=[*])OC(=[*])[*]	0.351	1 out of 1
ECFP_4	-570915357	 [*]O[c](:[cH]:[*]):[c](:[*]):[*]	0.351	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	-813997308	 [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.597	0 out of 2
ECFP_4	464808839	 [*]C(=CC(=[*])[*])[*]	-0.352	2 out of 8
ECFP_4	769217534	 [*]C(=[*])CO	-0.342	0 out of 1



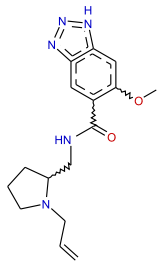
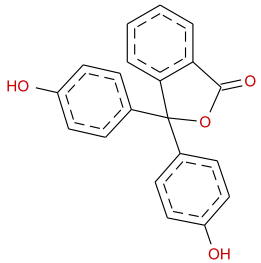
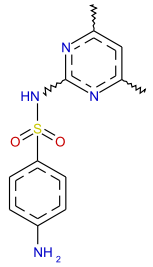
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.361
Enrichment: 0.879
Bayesian Score: -1.09
Mahalanobis Distance: 13.9
Mahalanobis Distance p-value: 7.96e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Alizapride	Phenolphthalein	Sulfamethazine
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.633	0.657	0.665
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

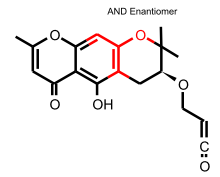
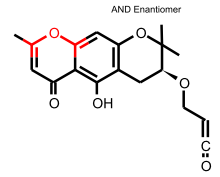
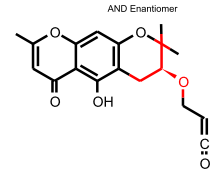
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -331149802: [*]O\C(=C/[*])\C
3. Unknown ECFP_2 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
4. Unknown ECFP_2 feature: 1778376725: [*]OC(C)(C)C([*])[*]
5. Unknown ECFP_2 feature: -1591590376: [*]C=C=O
6. Unknown ECFP_2 feature: 2106995136: [*]=C=O
7. Unknown ECFP_2 feature: -91536905: [*]CC=C=[*]
8. Unknown ECFP_2 feature: -1688150664: [*]OCC=[*]
9. Unknown ECFP_2 feature: -1250019913: [*]COC([*])[*]

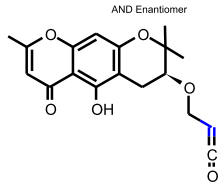
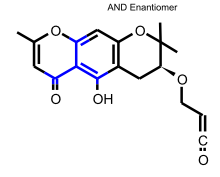
Feature Contribution

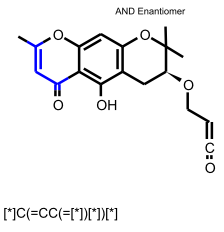
Top features for positive contribution

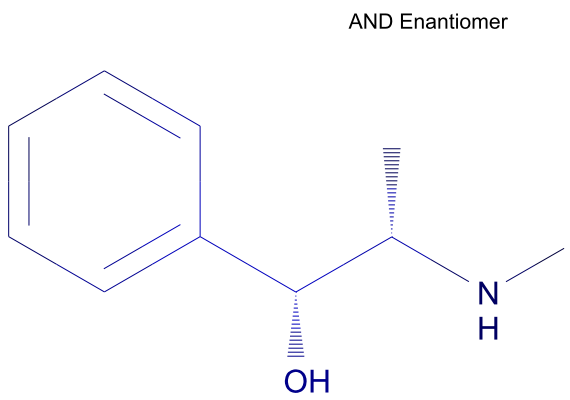
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set

ECFP_4	-570915357	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.351	1 out of 1
ECFP_4	-560785749	 [*]C(=[*])OC(=[*])[*]	0.351	1 out of 1
ECFP_4	-2124995946	 [*]C[C@H](O[*])C([*]) ([*])[*]	0.351	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	-1925046727	 [*]C=[*]	-0.605	2 out of 11
ECFP_4	-813997308	 [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.597	0 out of 2

ECFP_4	464808839	 [*]C(=CC(=[*])[*])[*]	-0.352	2 out of 8



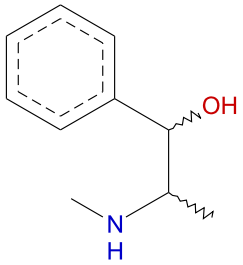
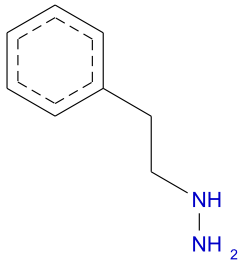
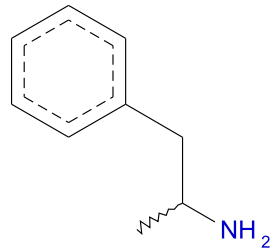
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.216
Enrichment: 0.735
Bayesian Score: -4
Mahalanobis Distance: 6.95
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Phenelzine	Amphetamine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.000	0.484	0.487
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

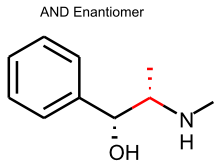
Model Applicability

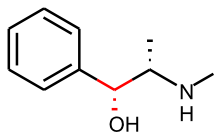
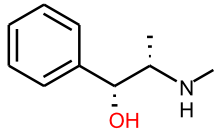
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

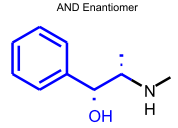
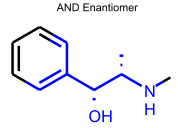
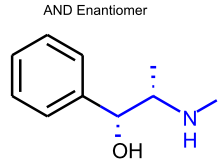
Feature Contribution

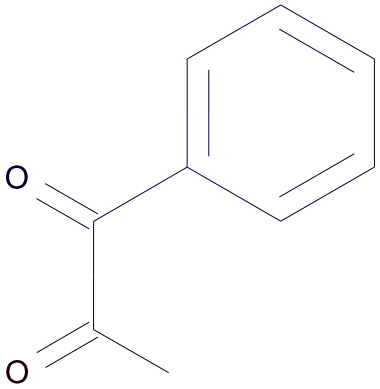
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	136597326	 [*]C([*])C	0.155	49 out of 159

FCFP_6	0	AND Enantiomer  [*]C([*])[*]	0.114	90 out of 305
FCFP_6	3	AND Enantiomer  [*]O	0.105	62 out of 212

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1083860676	AND Enantiomer  [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[cH]:[cH] :[cH]:1	-0.719	0 out of 4
FCFP_6	-239801958	AND Enantiomer  [*]N[C@@H](C)[C@H](O) [c](:[cH]:[*]):[cH]: [*]	-0.423	0 out of 2
FCFP_6	223071155	AND Enantiomer  [*]C([*])[C@H](C)NC	-0.423	0 out of 2



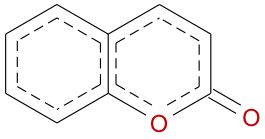
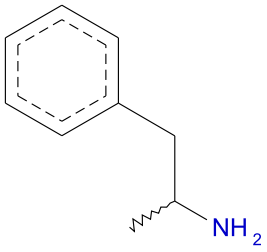
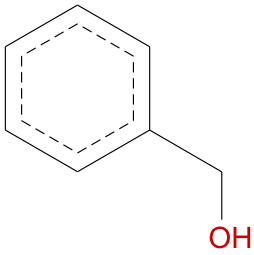
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Carcinogen
Probability: 0.312
Enrichment: 1.06
Bayesian Score: 0.23
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 0.252

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Coumarin	Amphetamine	Benzyl alcohol
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.520	0.524	0.538
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

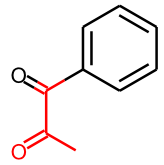
Model Applicability

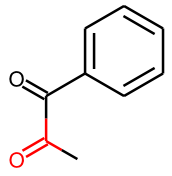
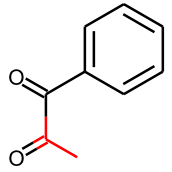
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

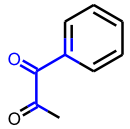
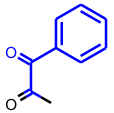
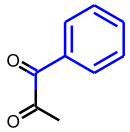
Feature Contribution

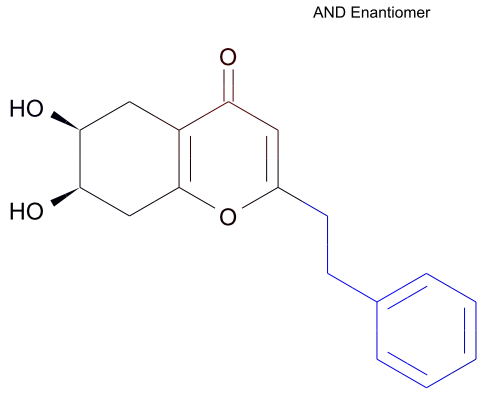
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.245	7 out of 20

FCFP_6	1872154524	 [*]C(=O)[*]	0.205	69 out of 213
FCFP_6	136597326	 [*]C([*])C	0.155	49 out of 159

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1549192822	 [*]C(=[*])C(=O)[c](:[*]):[*]	-0.489	3 out of 21
FCFP_6	975909016	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.233	0 out of 1
FCFP_6	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.22	15 out of 72



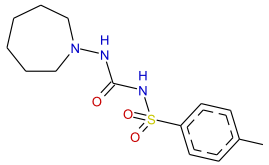
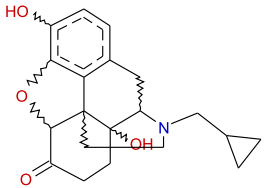
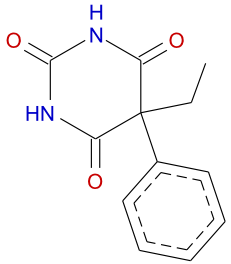
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.209
Enrichment: 0.712
Bayesian Score: -4.4
Mahalanobis Distance: 11.4
Mahalanobis Distance p-value: 0.0288

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Tolazamide	Naltrexone	Phenobarbital
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.524	0.581	0.584
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

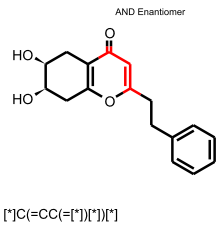
Model Applicability

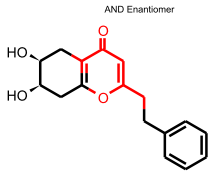
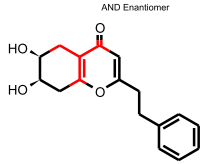
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC25 out of range. Value: -3.472. Training min, max, SD, explained variance: -3.0329, 3.839, 1.061, 0.0118.

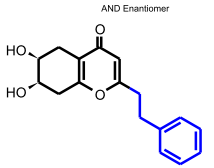
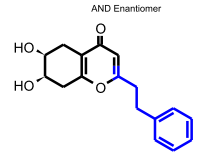
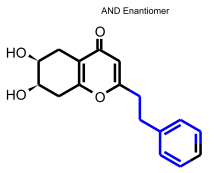
Feature Contribution

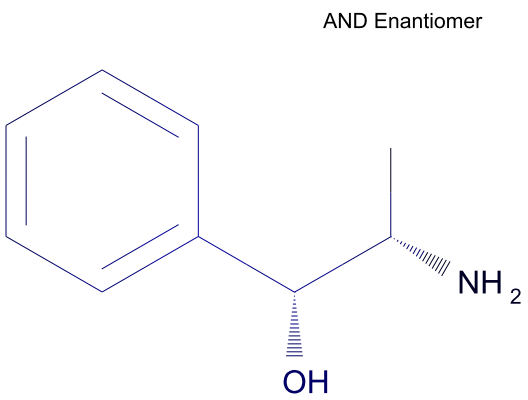
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48

FCFP_6	859241794	 [*]CC1=CC(=O)C(=[*]O1)[*]	0.46	1 out of 1
FCFP_6	436886043	 [*]CC(=C([*])[*])C(=[*])[*]	0.306	15 out of 41

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1981711554	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.42	0 out of 12
FCFP_6	1388176727	 [*]C(=[*])CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.21	0 out of 9
FCFP_6	-497728148	 [*]CC[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.96	2 out of 26



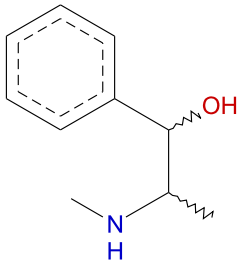
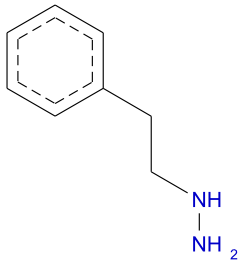
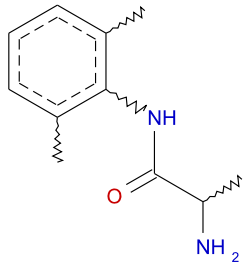
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.257
Enrichment: 0.872
Bayesian Score: -1.99
Mahalanobis Distance: 7.67
Mahalanobis Distance p-value: 0.996

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Phenelzine	Tocainide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.308	0.429	0.458
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

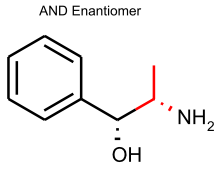
Model Applicability

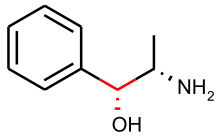
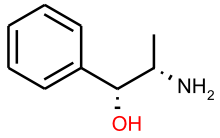
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

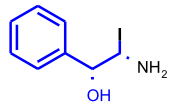
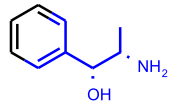
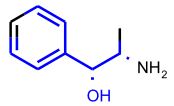
Feature Contribution

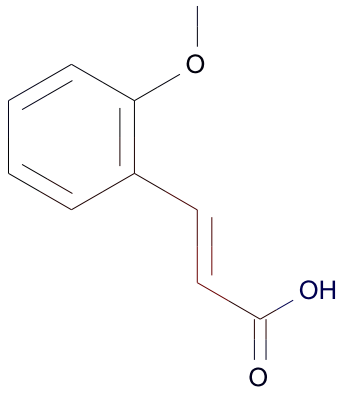
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	136597326	 [*]C([*])C	0.155	49 out of 159

FCFP_6	0	AND Enantiomer  [*]C([*])[*]	0.114	90 out of 305
FCFP_6	3	AND Enantiomer  [*]O	0.105	62 out of 212

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1083860676	AND Enantiomer  [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[cH]:[cH] :[cH]:1	-0.719	0 out of 4
FCFP_6	-239801958	AND Enantiomer  [*]N[C@@H](C)[C@H](O) [c](:[cH]:[*]):[cH]: [*]	-0.423	0 out of 2
FCFP_6	-1931573337	AND Enantiomer  [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	-0.387	2 out of 13



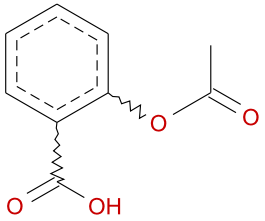
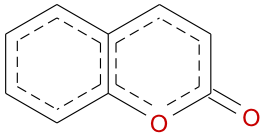
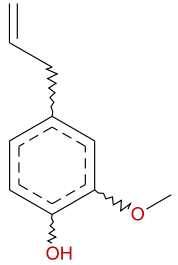
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.265
Enrichment: 0.901
Bayesian Score: -1.62
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.207

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aspirin	Coumarin	Eugenol
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.466	0.498	0.529
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

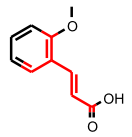
Model Applicability

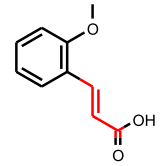
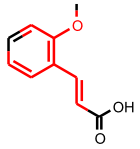
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

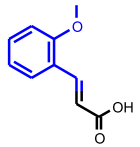
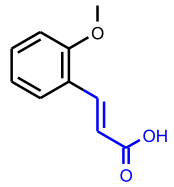
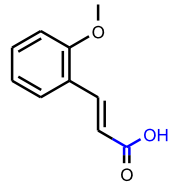
Feature Contribution

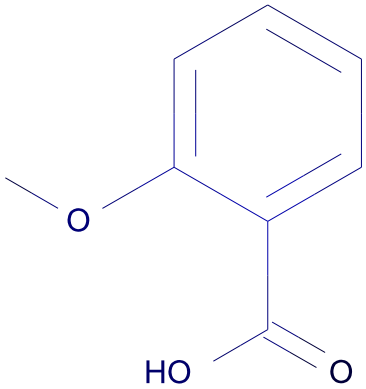
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	0.676	2 out of 2

FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48
FCFP_6	-395337223	 [*]O[c]1:[cH]:[*]:[cH]:[cH]:[c]:1\C=C\[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1458856986	 [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	-0.439	1 out of 8
FCFP_6	-2107131107	 [*]\C=C\C(=O)O	-0.423	0 out of 2
FCFP_6	-548632217	 [*]C(=[*])O	-0.383	9 out of 52



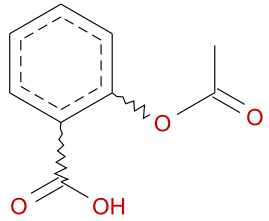
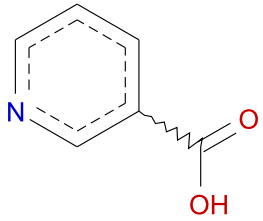
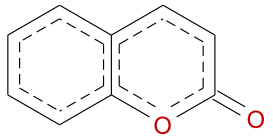
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.234
Enrichment: 0.797
Bayesian Score: -3.04
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.0485

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aspirin	Nicotinic acid	Coumarin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.271	0.528	0.551
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

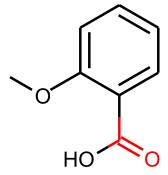
Model Applicability

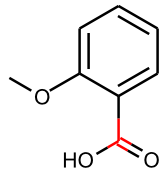
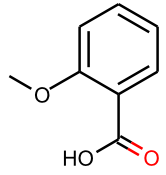
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

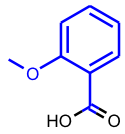
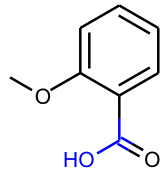
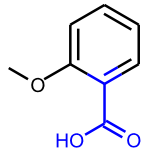
Feature Contribution

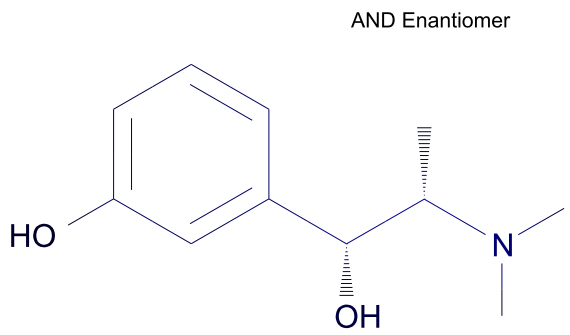
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1872154524	 [*]C(=O)[*]	0.205	69 out of 213

FCFP_6	0	 [*]C([*])[*]	0.114	90 out of 305
FCFP_6	1	 [*]=O	0.0783	76 out of 267

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1458856986	 [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	-0.439	1 out of 8
FCFP_6	-548632217	 [*]C(=[*])O	-0.383	9 out of 52
FCFP_6	-1549222613	 [*]:[c](:[*])C(=O)O	-0.351	1 out of 7



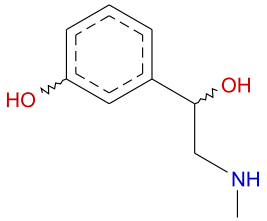
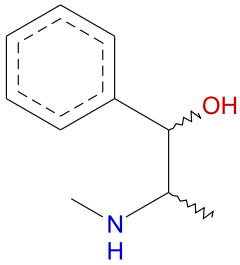
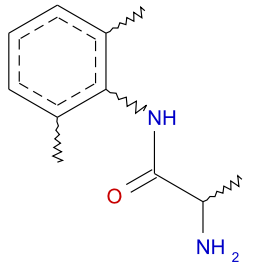
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.2
Enrichment: 0.678
Bayesian Score: -5.01
Mahalanobis Distance: 11.5
Mahalanobis Distance p-value: 0.0255

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenylephrine	Ephedrine	Tocainide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.458	0.463	0.534
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

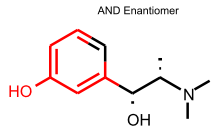
Model Applicability

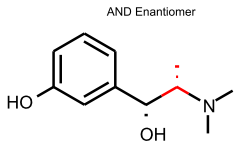
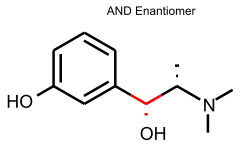
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

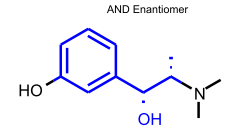
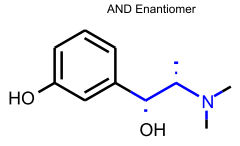
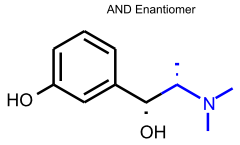
Feature Contribution

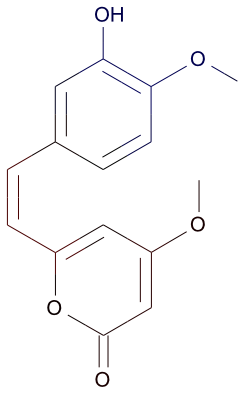
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-158888774	 [*][c]1:[*]:[cH]:[cH] :[c](O):[cH]:1	0.367	5 out of 12

FCFP_6	136597326	 [*]C([*])C	0.155	49 out of 159
FCFP_6	0	 [*]C([*])[*]	0.114	90 out of 305

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1083860676	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[cH]:[cH] :[cH]:1	-0.719	0 out of 4
FCFP_6	-1946918893	 [*]C([*])[C@H](C)N([*])[*]	-0.692	2 out of 19
FCFP_6	-1343180157	 [*]C([*])N(C)C	-0.533	9 out of 61



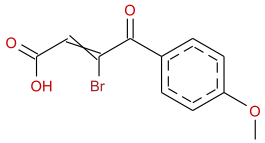
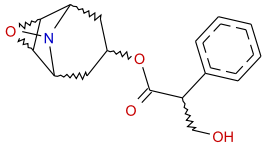
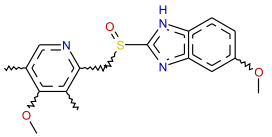
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.295
Enrichment: 1
Bayesian Score: -0.404
Mahalanobis Distance: 13.1
Mahalanobis Distance p-value: 0.000195

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Cytembena	Scopolamine	Omeprazole
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.489	0.578	0.607
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

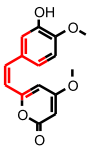
Model Applicability

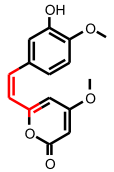
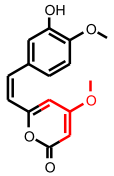
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

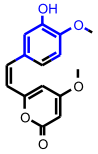
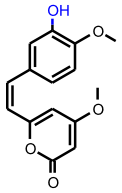
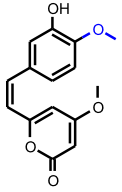
Feature Contribution

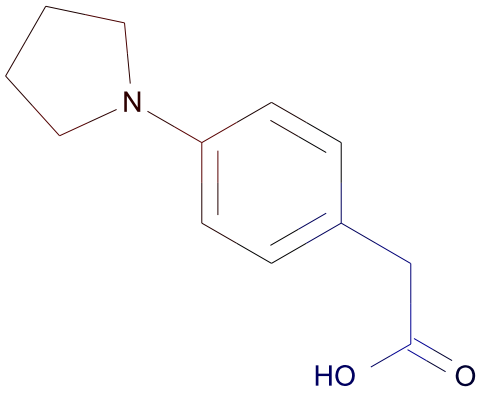
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	0.676	2 out of 2

FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48
FCFP_6	-165061478	 [*]\C=C(\OC)/C=[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	523826990	 [*]O[c]1:[cH]:[*]:[c]([*]):[cH]:[c]:1O	-0.423	0 out of 2
FCFP_6	7	 [*]O	-0.308	15 out of 79
FCFP_6	136627117	 [*]OC	-0.252	10 out of 50



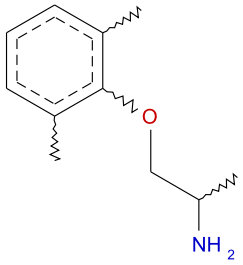
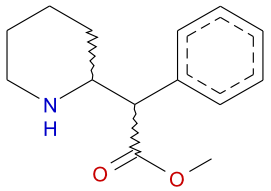
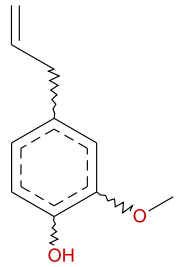
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.239
Enrichment: 0.813
Bayesian Score: -2.79
Mahalanobis Distance: 10.5
Mahalanobis Distance p-value: 0.192

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Mexiletine	Methylphenidate	Eugenol
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.522	0.523	0.526
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

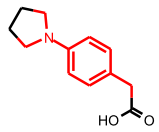
Model Applicability

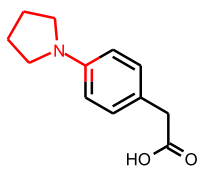
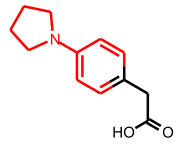
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

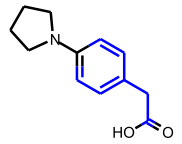
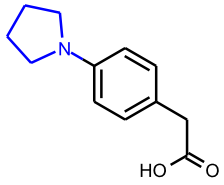
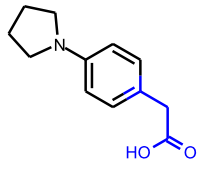
Feature Contribution

Top features for positive contribution

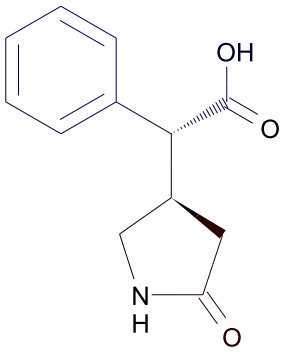
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1033057683	 [*]C[c]1:[cH]:[cH]:[c](:[cH]:[cH]:1)N2C[*][*]C2	0.805	3 out of 3

FCFP_6	-822674211	 [*]:[c](:[*])N1CCCC1	0.517	2 out of 3
FCFP_6	243588731	 [*]1:[cH]:[cH]:[c](:[cH]:[cH]:1)N2CCCC2	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-497728148	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.96	2 out of 26
FCFP_6	-98332825	 [*]N1CCCC1	-0.793	1 out of 13
FCFP_6	2039196715	 [*]:[c](:[*])CC(=O)O	-0.582	0 out of 3

AND Enantiomer



$C_{12}H_{13}NO_3$

Molecular Weight: 219.237

ALogP: 0.621

Rotatable Bonds: 3

Acceptors: 3

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.28

Enrichment: 0.953

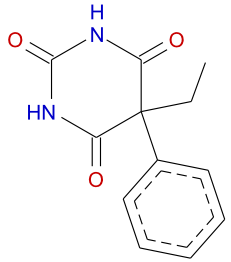
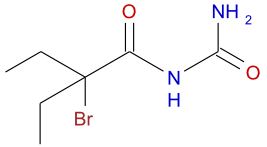
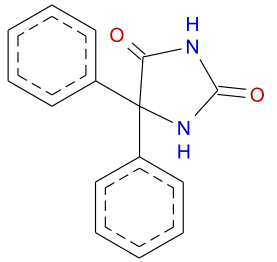
Bayesian Score: -0.973

Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.0923

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenobarbital	Carbromal	Phenytoin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.452	0.548	0.561
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

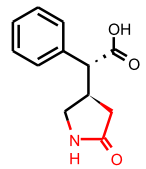
Model Applicability

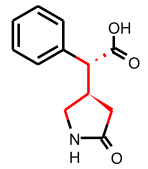
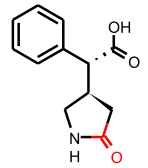
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

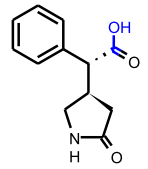
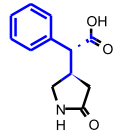
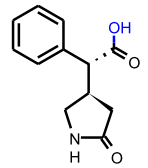
Feature Contribution

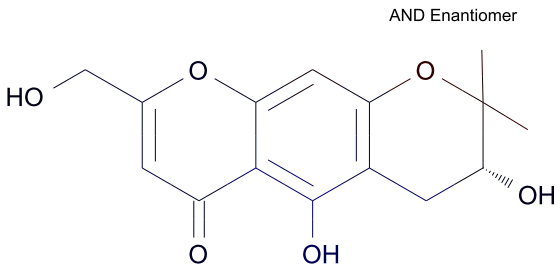
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	566058135	AND Enantiomer  O=C1C[*][*]N1	0.447	17 out of 40

FCFP_6	-1043339860	AND Enantiomer  [*]C([*])C1C[*][*]C1	0.383	24 out of 61
FCFP_6	1872154524	AND Enantiomer  [*]C(=O)[*]	0.205	69 out of 213

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-548632217	AND Enantiomer  [*]C(=[*])O	-0.383	9 out of 52
FCFP_6	-1716639150	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.351	1 out of 7
FCFP_6	7	AND Enantiomer  [*]O	-0.308	15 out of 79



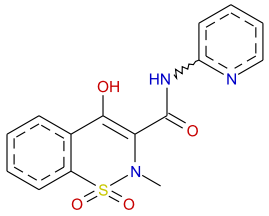
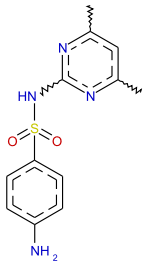
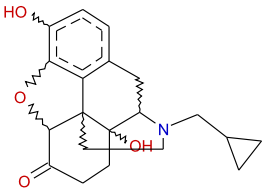
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.247
Enrichment: 0.839
Bayesian Score: -2.43
Mahalanobis Distance: 13.5
Mahalanobis Distance p-value: 4.52e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Piroxicam	Sulfamethazine	Naltrexone
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.642	0.668	0.679
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

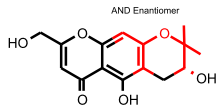
Model Applicability

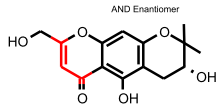
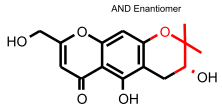
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

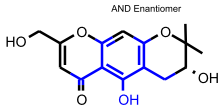
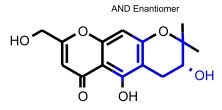
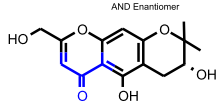
Feature Contribution

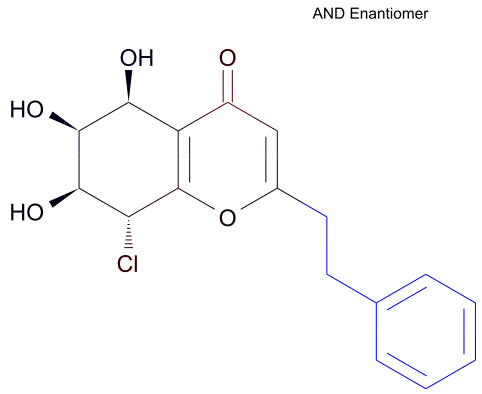
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1979984779	 [*][C@@H]1[*][c](:[*]):[c](OC1(C)C):[cH]:[*]	0.668	3 out of 4

FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48
FCFP_6	-415216134	 [*]OC(C)(C)C([*])[*]	0.306	5 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1601875224	 [*]C[c]1:[c]([*]):[*] :[c]([*]):[c](C(=[*])[*]):[c]:1O	-0.582	0 out of 3
FCFP_6	-2006448698	 [*][c](:[*]):[c]1C[C@ @H](O)C([*])([*])[*] [c]:1:[*]	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]C(=[*])C(=O)[c](:[*]):[*]	-0.489	3 out of 21



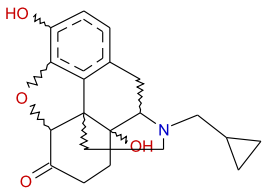
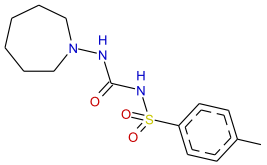
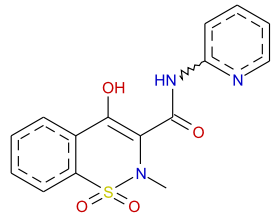
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.209
Enrichment: 0.71
Bayesian Score: -4.43
Mahalanobis Distance: 12.7
Mahalanobis Distance p-value: 0.000809

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Naltrexone	Tolazamide	Piroxicam
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.612	0.626	0.631
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

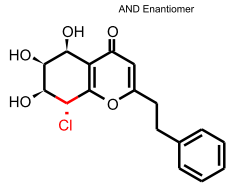
Model Applicability

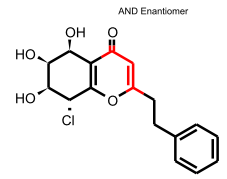
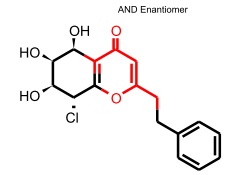
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC20 out of range. Value: 4.5013. Training min, max, SD, explained variance: -4.4914, 4.1353, 1.159, 0.0141.
2. OPS PC25 out of range. Value: -3.0813. Training min, max, SD, explained variance: -3.0329, 3.839, 1.061, 0.0118.

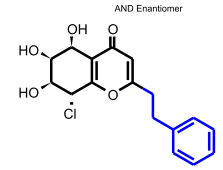
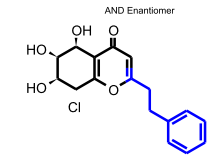
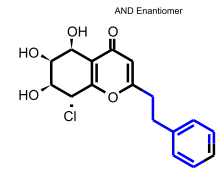
Feature Contribution

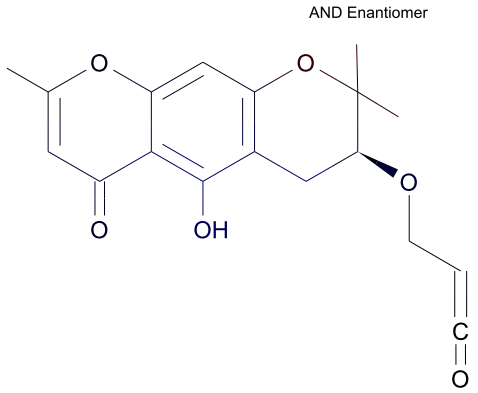
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	71953198	 [*]C([*])Cl	0.612	12 out of 23

FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48
FCFP_6	859241794	 [*]CC1=CC(=O)C(=[*]O1)[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1981711554	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.42	0 out of 12
FCFP_6	1388176727	 [*]C(=[*])CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.21	0 out of 9
FCFP_6	-497728148	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.96	2 out of 26



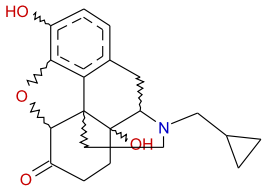
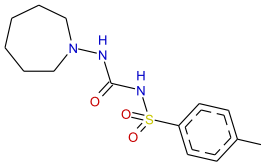
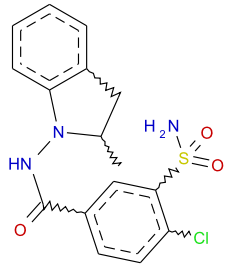
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.193
Enrichment: 0.655
Bayesian Score: -5.45
Mahalanobis Distance: 13
Mahalanobis Distance p-value: 0.000256

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Naltrexone	Tolazamide	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.523	0.554	0.570
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

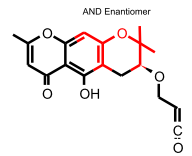
Model Applicability

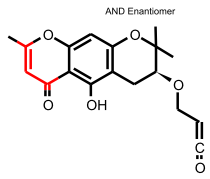
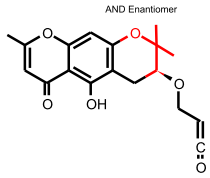
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 444624378: [*]C=C=O

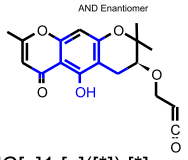
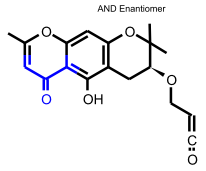
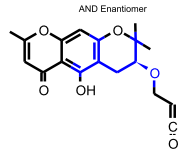
Feature Contribution

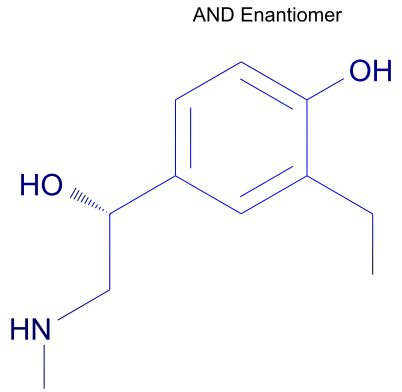
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1979984779	 [*][C@@H]1[*][c](:[*]):[c](OC1(C)C):[cH]:[*]	0.668	3 out of 4

FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48
FCFP_6	-415216134	 [*]OC(C)(C)C([*])[*]	0.306	5 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1601875224	 [*]C[c]1:[c]([*]):[*]:[c]([*]):[c](C(=[*])[*]):[c]:1O	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]C(=[*])C(=O)[c](:[*]):[*]	-0.489	3 out of 21
FCFP_6	-2000646149	 [*]O[C@H]1C[c](:[c]([*]):[*]):[c](:[*])[*]C1([*])[*]	-0.423	0 out of 2



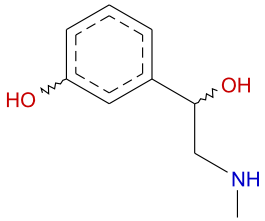
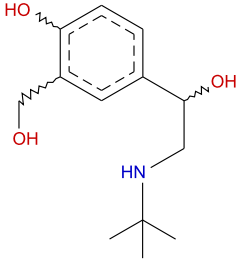
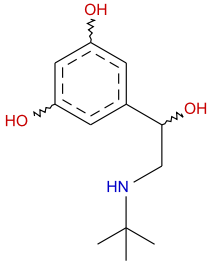
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.173
Enrichment: 0.589
Bayesian Score: -6.85
Mahalanobis Distance: 8.56
Mahalanobis Distance p-value: 0.925

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenylephrine	Albuterol	Terbutaline
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.376	0.429	0.494
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

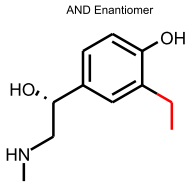
Model Applicability

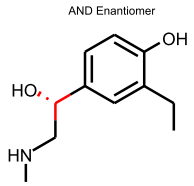
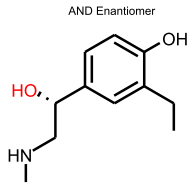
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

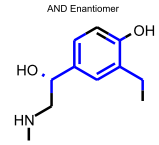
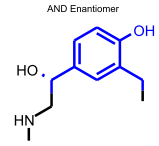
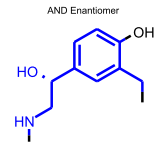
Feature Contribution

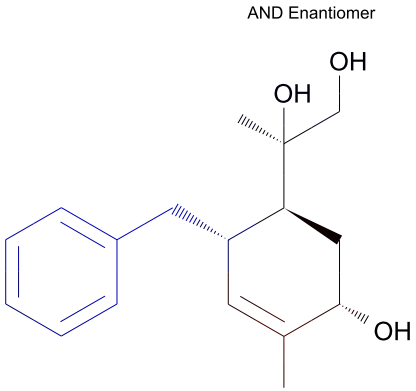
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	136597326	 [*]C([*])C	0.155	49 out of 159

FCFP_6	0	 [*]C([*])[*]	0.114	90 out of 305
FCFP_6	3	 [*]O	0.105	62 out of 212

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-451251206	 [*]C[c]1:[cH]:[c](:[cH]:[*]:[c]:1[*])C([*])[*]	-0.731	1 out of 12
FCFP_6	-306804326	 [*]C[c]1:[cH]:[c](:[cH]:[cH]:[c]:1O)C([*])[*]	-0.582	0 out of 3
FCFP_6	-616387365	 [*]C[c]1:[cH]:[c](:[cH]:[cH]:[c]:1[*])[C@@H](O)CN[*]	-0.582	0 out of 3



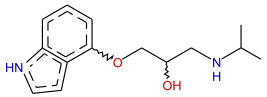
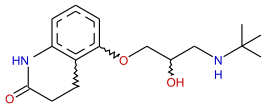
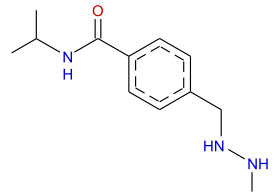
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.237
Enrichment: 0.806
Bayesian Score: -2.9
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 0.00166

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Pindolol	Carteolol	Procarbazine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.576	0.590	0.609
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

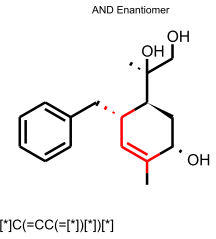
Model Applicability

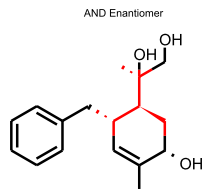
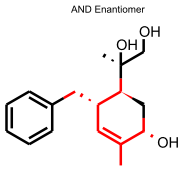
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

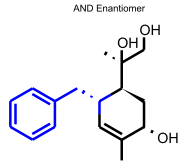
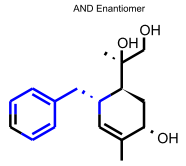
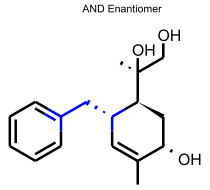
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48

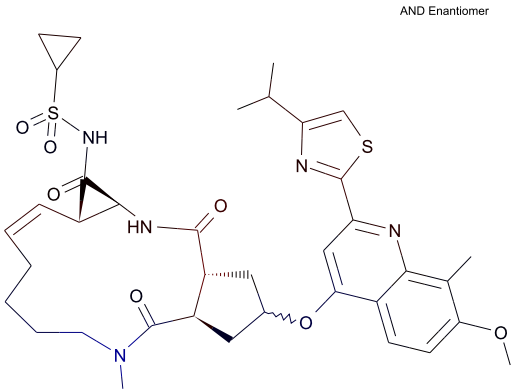
FCFP_6	-1043339860	 [*]C([*])C1C[*][*]C1	0.383	24 out of 61
FCFP_6	1000913921	 [*]C[C@H]1C=C(C)[C@@H]([*])[*][C@@H]1[*]	0.38	2 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1981711554	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.42	0 out of 12
FCFP_6	-497728148	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.96	2 out of 26
FCFP_6	907007053	 [*]CC[c](:[*]):[*]	-0.366	11 out of 62

Simeprevir

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.258

Enrichment: 0.877

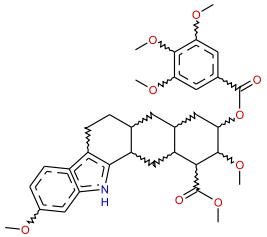
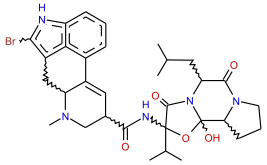
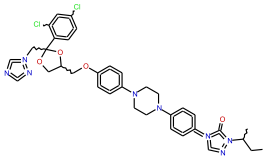
Bayesian Score: -1.92

Mahalanobis Distance: 18.3

Mahalanobis Distance p-value: 6.22e-015

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Reserpine	Bromocriptine	Itraconazole
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.775	0.832	0.865
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

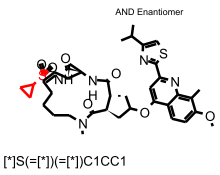
Model Applicability

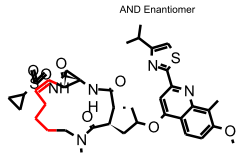
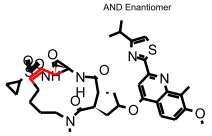
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 690481386: [*]:[c](:[*])[c]1:n:[*]:[*]:s:1

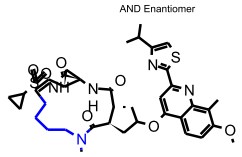
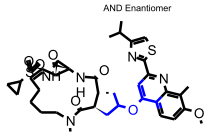
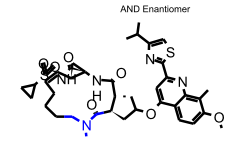
Feature Contribution

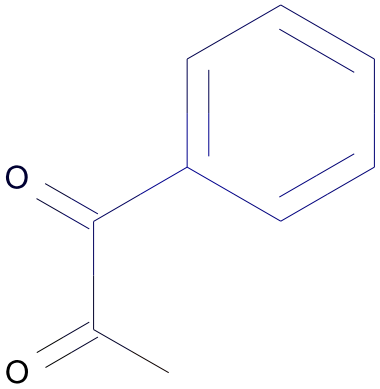
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-55265897	 [*]S(=[*])(=[*])C1CC1	0.594	17 out of 34

FCFP_6	-1289661876	 [*]CCC\C=C/[*]	0.517	2 out of 3
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-98332825	 [*]N1CCCC1	-0.793	1 out of 13
FCFP_6	-1972798083	 [*]:[cH]:[c](OC1C[*][*]C1):[c](:[*]):[*]	-0.582	0 out of 3
FCFP_6	-1553874037	 [*]CN(C)C(=[*])[*]	-0.45	5 out of 32



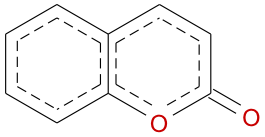
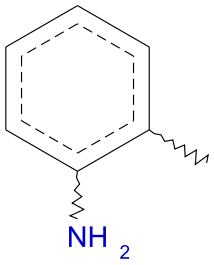
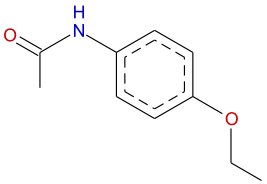
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.175
Enrichment: 0.581
Bayesian Score: -4.25
Mahalanobis Distance: 7.52
Mahalanobis Distance p-value: 0.418

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Coumarin	Toluidine; o-	Phenacetin
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.528	0.637	0.642
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

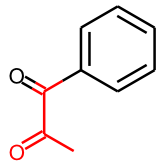
Model Applicability

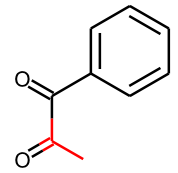
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

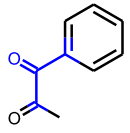
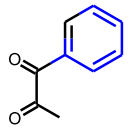
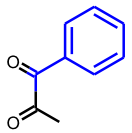
Feature Contribution

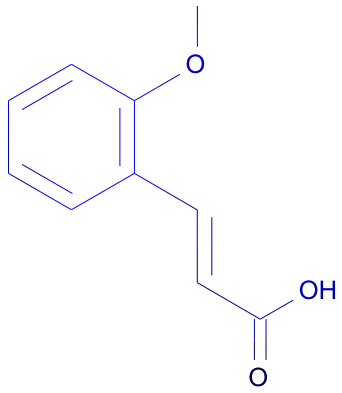
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.168	3 out of 7

FCFP_12	136597326	 [*]C(=[*])C	0.0722	18 out of 49

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1549192822	 [*]C(=[*])C(=O)[c](:[*]):[*]	-0.704	0 out of 3
FCFP_12	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.562	5 out of 28
FCFP_12	-1698724694	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.423	3 out of 15



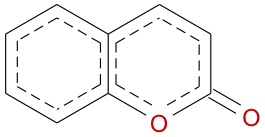
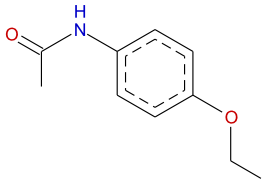
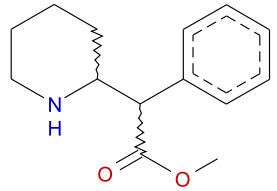
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.148
Enrichment: 0.491
Bayesian Score: -10.9
Mahalanobis Distance: 9.52
Mahalanobis Distance p-value: 0.0383

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Coumarin	Phenacetin	Methylphenidate
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.548	0.559	0.580
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

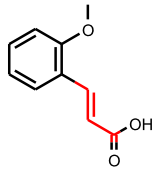
Model Applicability

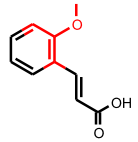
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC8 out of range. Value: 4.4612. Training min, max, SD, explained variance: -3.6394, 4.2146, 1.519, 0.0430.

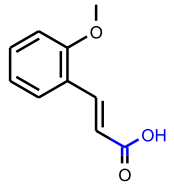
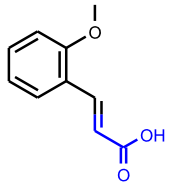
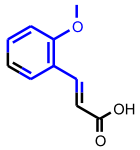
Feature Contribution

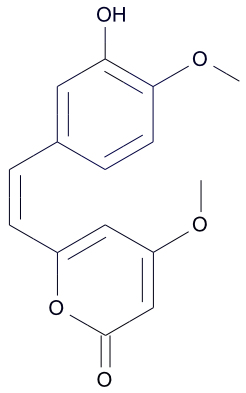
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	451847724	 [*]\C=C\C(=[*])[*]	0.3	10 out of 21

FCFP_12	-1977641857	 [*][c](:[*]):[c](OC):[cH]:[*]	0.105	2 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-548632217	 [*]C(=[*])O	-1.4	0 out of 9
FCFP_12	-1176841573	 [*]=CC(=O)O	-1.31	0 out of 8
FCFP_12	-1099193755	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	-1.11	0 out of 6



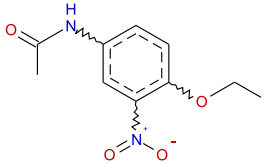
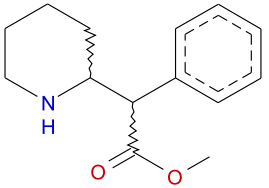
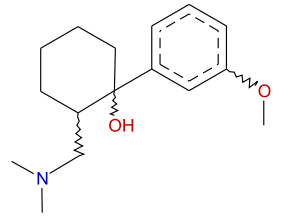
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.157
Enrichment: 0.521
Bayesian Score: -6.14
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.0024

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Nitroacetophenetide	Methylphenidate	Tramadol
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.660	0.681	0.705
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

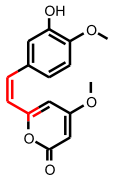
Model Applicability

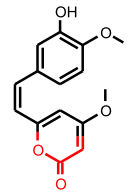
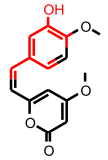
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

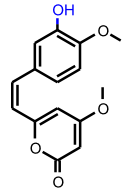
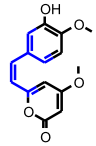
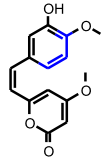
Feature Contribution

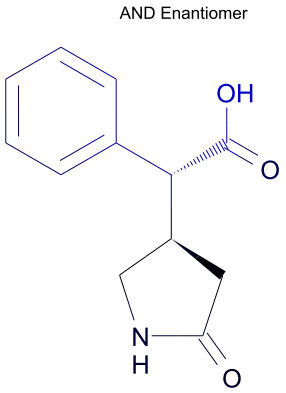
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	451847724	 [*]\C=C\C(=[*])[*]	0.3	10 out of 21

FCFP_12	565998553	 [*]OC(=O)C=[*]	0.194	6 out of 14
FCFP_12	949015626	 [*][c]1:[*]:[cH]:[c](C=[*]):[cH]:[c]:1O	0.174	1 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	7	 [*]O	-0.71	2 out of 15
FCFP_12	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	-0.519	0 out of 2
FCFP_12	1618154665	 [*]:[cH]:[cH]:[cH]:[*]	-0.409	13 out of 59



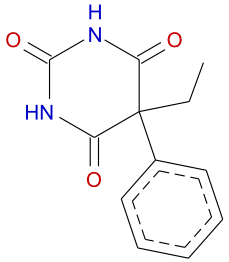
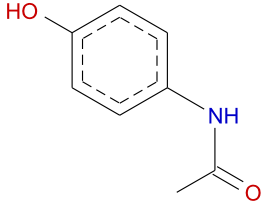
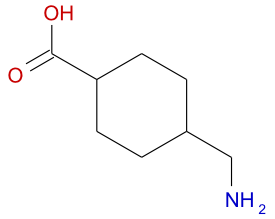
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.146
Enrichment: 0.484
Bayesian Score: -9.75
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 0.00023

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenobarbital	Acetaminophen	Tranexamic acid
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.469	0.633	0.644
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

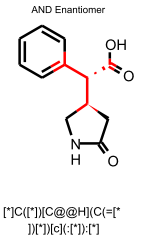
Model Applicability

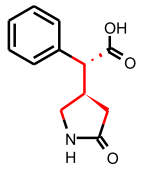
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: -3.5894. Training min, max, SD, explained variance: -3.5268, 3.8048, 1.733, 0.0560.

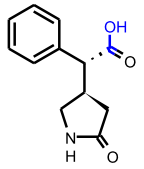
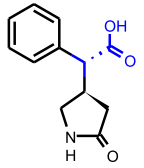
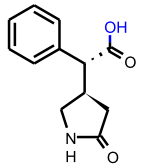
Feature Contribution

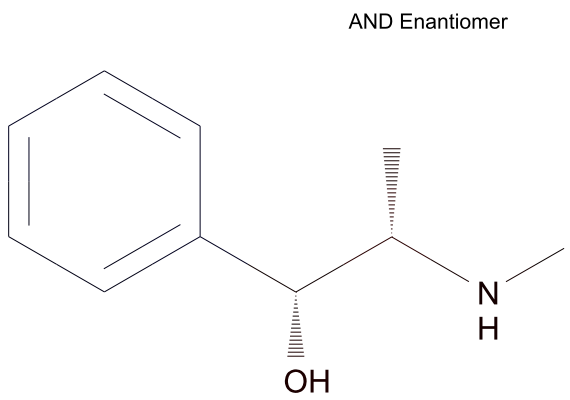
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1186303932	 [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]	0.497	4 out of 6

FCFP_12	-1043339860	AND Enantiomer  [*]C([*])C1C[*][*]C1	0.349	12 out of 24

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-548632217	AND Enantiomer  [*]C(=[*])O	-1.4	0 out of 9
FCFP_12	-1176841573	AND Enantiomer  [*]=CC(=O)O	-1.31	0 out of 8
FCFP_12	7	AND Enantiomer  [*]O	-0.71	2 out of 15



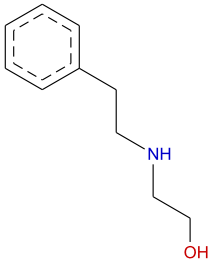
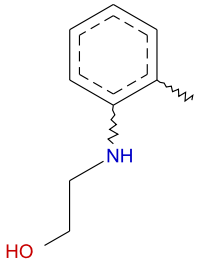
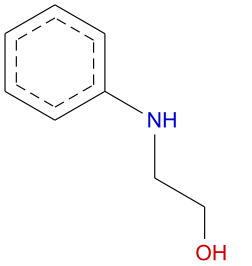
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Moderate_Severe
Probability: 0.827
Enrichment: 1.2
Bayesian Score: 0.297
Mahalanobis Distance: 8.03
Mahalanobis Distance p-value: 0.917

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ETHANOL;2-(PHENETHYLAMINO)-	0-TOLUIDINO-ETHANOL	Ethanol; 2-anilino-
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild
Distance	0.430	0.439	0.486
Reference	AMIHBC 10;61;54	JIHTAB 31;60;49	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;691;86

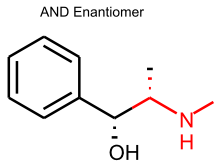
Model Applicability

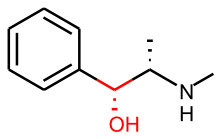
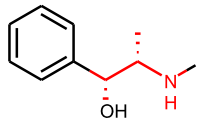
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

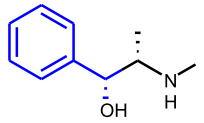
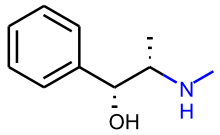
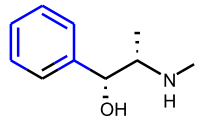
Feature Contribution

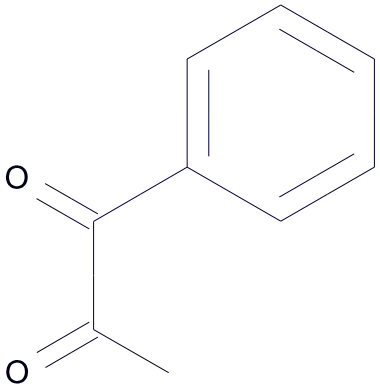
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-885550502	 [*]C([*])NC	0.239	54 out of 64

FCFP_10	1070061035	AND Enantiomer  [*]C([*])O	0.239	284 out of 338
FCFP_10	-1043250487	AND Enantiomer  [*]N[C@@H](C)C([*])[*]	0.22	62 out of 75

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1698724694	AND Enantiomer  [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	136686699	AND Enantiomer  [*]NC	-0.243	6 out of 12
FCFP_10	991735244	AND Enantiomer  [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.185	130 out of 237



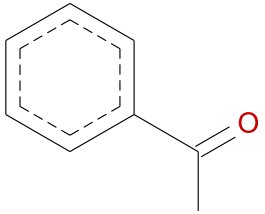
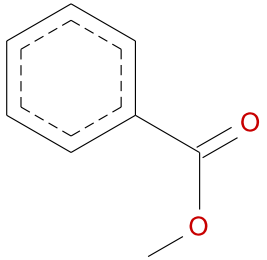
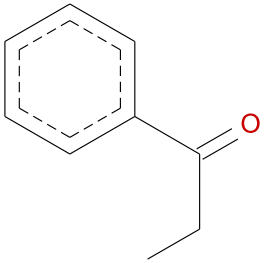
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Mild
Probability: 0.781
Enrichment: 1.13
Bayesian Score: -1.64
Mahalanobis Distance: 6.22
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACETOPHENONE	METHYL BENZOATE	Propiophenone
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.312	0.382	0.394
Reference	AJOPAA 29;1363;46	AMIHBC 10;61;54	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;291;86

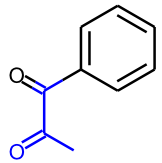
Model Applicability

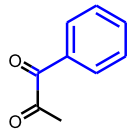
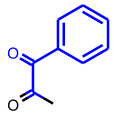
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

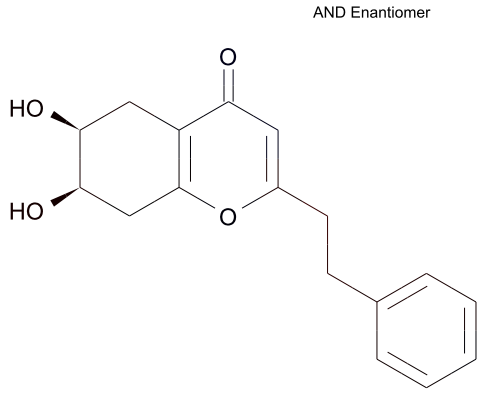
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	565968762	 [*]C(=[*])C(=O)C	-0.372	17 out of 38

FCFP_10	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	975909016	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.216	3 out of 6



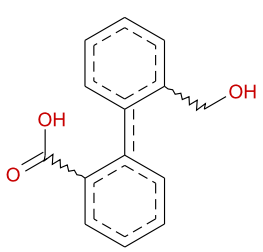
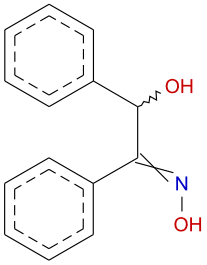
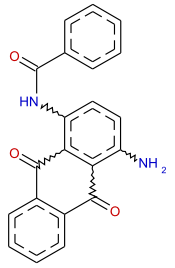
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Moderate_Severe
Probability: 0.829
Enrichment: 1.2
Bayesian Score: 0.444
Mahalanobis Distance: 10.7
Mahalanobis Distance p-value: 0.0129

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2-BIPHENYLCARBOXYLIC ACID; 2'-HYDROXYMETHYL-	BENZOIN; OXIME	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.593	0.613	0.621
Reference	IHFCAY 6;1;67	28ZPAK-;111;72	28ZPAK-;124;72

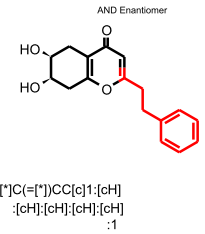
Model Applicability

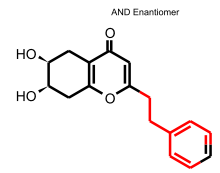
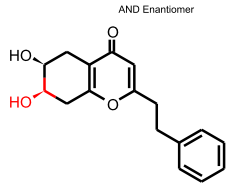
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

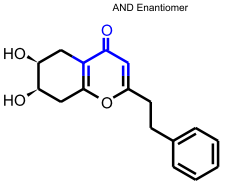
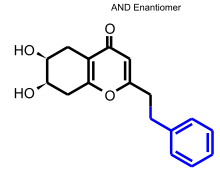
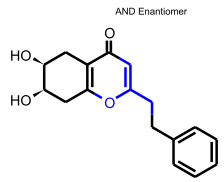
Feature Contribution

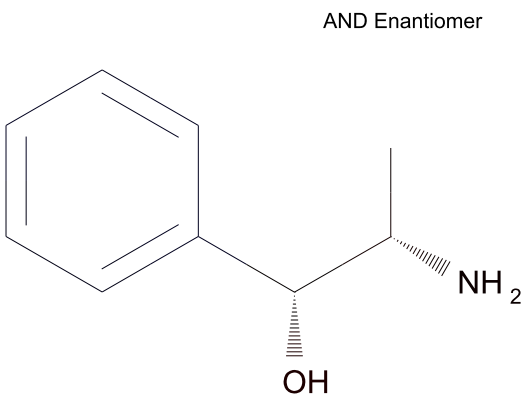
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1388176727	 [*]C(=[*])CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.389	19 out of 19

FCFP_10	-497728148	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.356	24 out of 25
FCFP_10	1070061035	 [*]C([*])O	0.239	284 out of 338

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	565968762	 [*]C(=[*])C(=O)C	-0.372	17 out of 38
FCFP_10	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	436915834	 [*]C\C(=C\[*])\O[*]	-0.194	2 out of 4



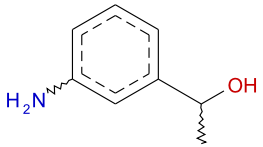
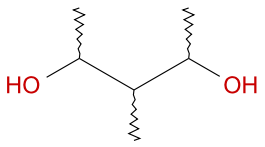
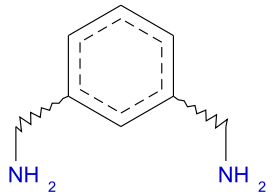
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Moderate_Severe
Probability: 0.825
Enrichment: 1.2
Bayesian Score: 0.171
Mahalanobis Distance: 5.12
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3-AMINO-2-METHYL BENZYL ALCOHOL	1;2-DIMETHYL;1;3-BUTANEDIOL	METHYLAMINE;M-PHENYLENEBIS-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe
Distance	0.341	0.459	0.461
Reference	AMIHBC 10;61;54	AIHAAP 23;95;62	28ZPAK-;64;72

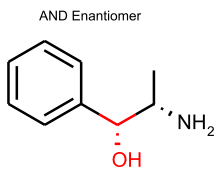
Model Applicability

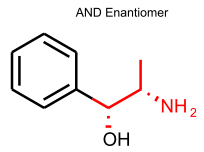
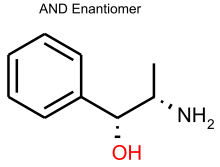
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

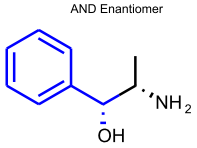
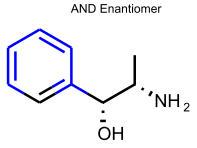
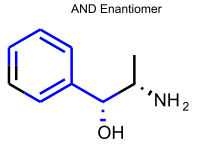
Feature Contribution

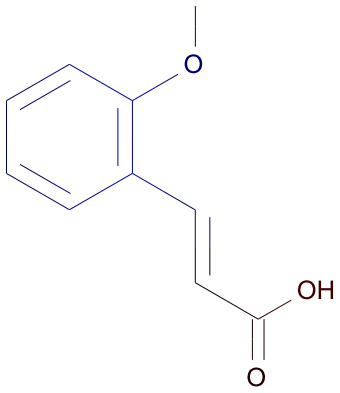
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1070061035	 [*]C([*])O	0.239	284 out of 338

FCFP_10	-1043250487	 [*]N[C@@H](C)C([*])[*]]	0.22	62 out of 75
FCFP_10	3	 [*]O	0.165	383 out of 491

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.185	130 out of 237
FCFP_10	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.13	153 out of 264



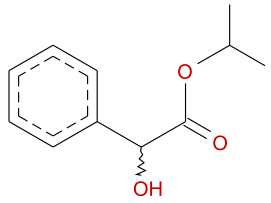
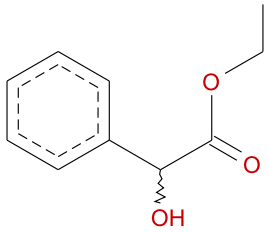
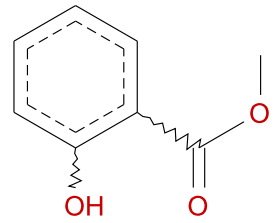
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.756
Enrichment: 1.1
Bayesian Score: -2.36
Mahalanobis Distance: 8.17
Mahalanobis Distance p-value: 0.882

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	MANDELIC ACID; ISOPROPYL ESTER	MANDELIC ACID; ETHYL ESTER	SALICYCLIC ACID; METHYL ESTER
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild
Distance	0.469	0.474	0.475
Reference	AJOPAA 29;1363;46	AJOPAA 29;1363;46	28ZPAK-;106;72

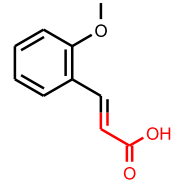
Model Applicability

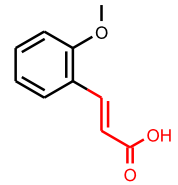
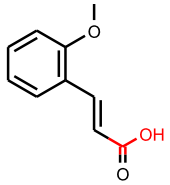
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

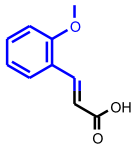
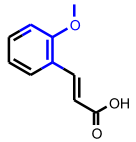
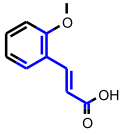
Feature Contribution

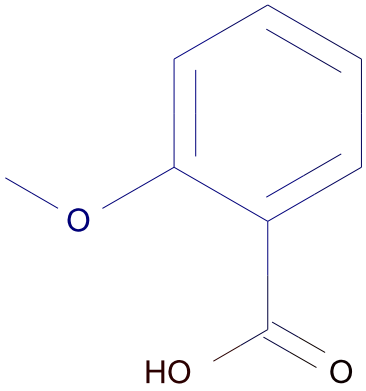
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1176841573	 [*]=CC(=O)O	0.356	41 out of 43

FCFP_10	-2107131107	 [*]\C=C\C(=O)O	0.344	6 out of 6
FCFP_10	-548632217	 [*]C(=[*])O	0.319	54 out of 59

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1458856986	 [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	-0.842	0 out of 2
FCFP_10	-1977641857	 [*][c](:[*]):[c](OC):[cH]:[*]	-0.78	4 out of 15
FCFP_10	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	-0.507	0 out of 1



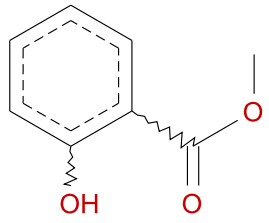
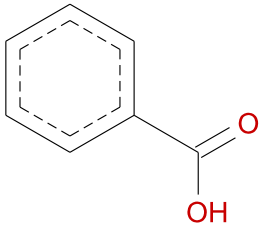
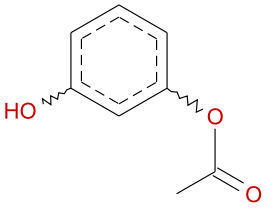
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.772
Enrichment: 1.12
Bayesian Score: -1.94
Mahalanobis Distance: 9.76
Mahalanobis Distance p-value: 0.159

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	SALICYCLIC ACID; METHYL ESTER	BENZOIC ACID	RESORCINOL; MONOACETATE
Structure			
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe
Distance	0.355	0.370	0.386
Reference	28ZPAK-;106;72	BIOFX* 28-4/73	JAPMA8 46;185;57

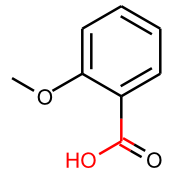
Model Applicability

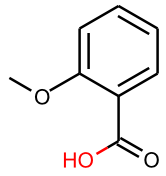
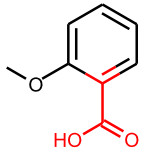
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

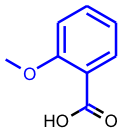
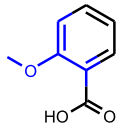
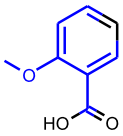
Feature Contribution

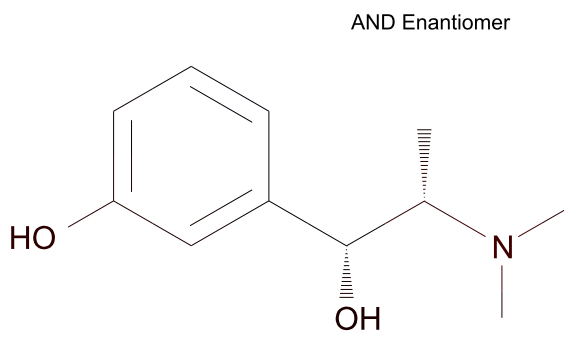
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-548632217	 [*]C(=[*])O	0.319	54 out of 59

FCFP_10	7	 [*]O	0.219	117 out of 142
FCFP_10	-1549222613	 [*]:[c](:[*])C(=O)O	0.157	11 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1458856986	 [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	-0.842	0 out of 2
FCFP_10	-1977641857	 [*][c](:[*]):[c](OC):[cH]:[*]	-0.78	4 out of 15
FCFP_10	-1099193755	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	-0.361	2 out of 5



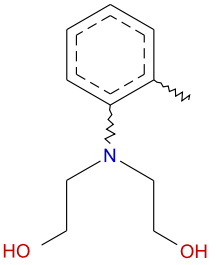
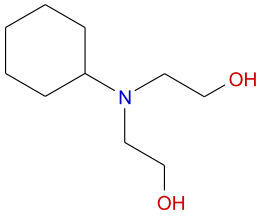
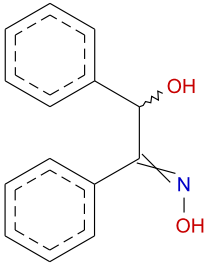
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Moderate_Severe
Probability: 0.864
Enrichment: 1.25
Bayesian Score: 2.28
Mahalanobis Distance: 14.7
Mahalanobis Distance p-value: 1.53e-012

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	o-Toluidine; N;N-bis(2-hydroxyethyl)-	ETHANOL; 2;2'-CYCLOHEXYLIMINODI-	BENZOIN; OXIME
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild
Distance	0.482	0.513	0.526
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;697;86	34ZIAG 61;69	28ZPAK-;111;72

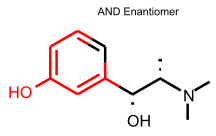
Model Applicability

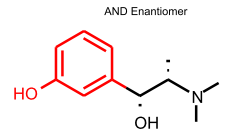
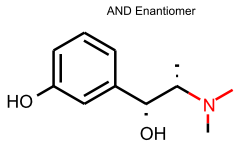
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

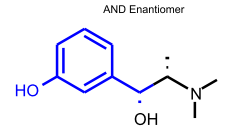
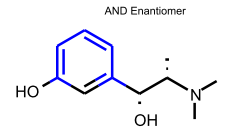
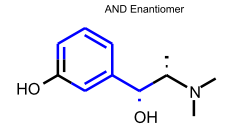
Feature Contribution

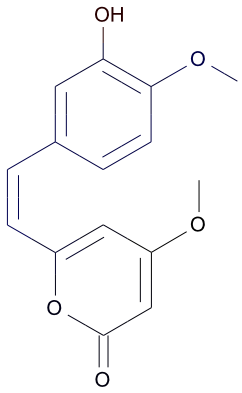
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-158888774	 [*][c]1:[*]:[cH]:[cH]:[c](O):[cH]:1	0.356	24 out of 25

FCFP_10	-1849700223	 [*][c]1:[cH]:[cH]:[cH]:[c](O):[cH]:1	0.344	6 out of 6
FCFP_10	136388789	 [*]N([*])C	0.331	28 out of 30

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1694796561	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[c](O):[cH]:1	-0.6	1 out of 4
FCFP_10	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.185	130 out of 237
FCFP_10	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.13	153 out of 264



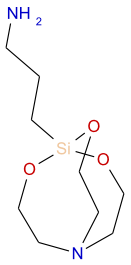
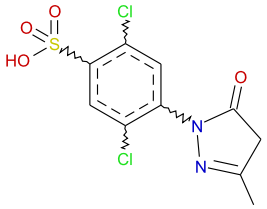
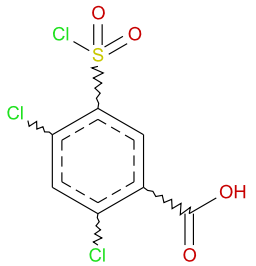
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.761
Enrichment: 1.1
Bayesian Score: -2.23
Mahalanobis Distance: 9.91
Mahalanobis Distance p-value: 0.116

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2;8;9-TRIOXA-5-AZA-1-SILABICYCLO(3.3.3)UNDECANE;1-(3-AMINOPROPYL)-	2;5-DICHLORO-4(3'-METHYL-5' PYRAZOLON-1'-YL)BENZENE SULFONIC ACID	BENZOIC ACID; 5-(CHLOROSULFONYL)-2;4-DICHLORO-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Moderate_Severe
Predicted Endpoint	Moderate_Severe	Mild	Moderate_Severe
Distance	0.587	0.648	0.664
Reference	28ZPAK-;220;72	28ZPAK-;186;72	FCTOD7 20;573;82

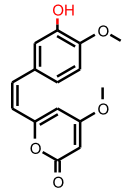
Model Applicability

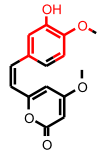
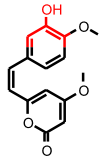
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

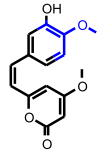
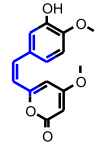
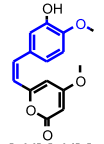
Feature Contribution

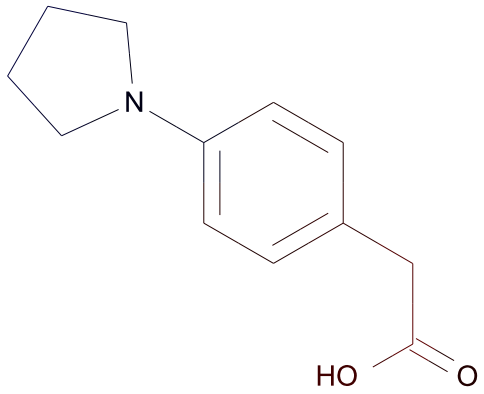
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	7	 [*]O	0.219	117 out of 142

FCFP_10	523826990	 [*]O[c]1:[cH]:[*]:[c]([*]):[cH]:[c]:1O	0.186	1 out of 1
FCFP_10	74595001	 [*]:[cH]:[c](O):[cH]:[*]	0.16	63 out of 81

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1977641857	 [*][c](:[*]):[c](OC):[cH]:[*]	-0.78	4 out of 15
FCFP_10	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	-0.507	0 out of 1
FCFP_10	-747977902	 [*]O[c]1:[cH]:[cH]:[c](\C=C/[*]):[cH]:[c]:1[*]	-0.507	0 out of 1



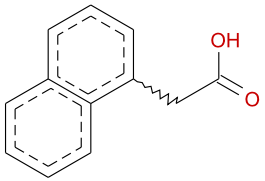
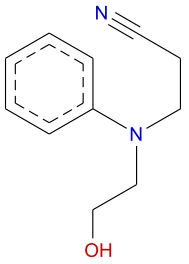
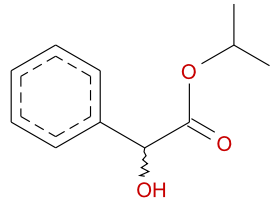
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Moderate_Severe
Probability: 0.837
Enrichment: 1.22
Bayesian Score: 1.1
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.00124

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	1-NAPHTHALENEACETIC ACID	Propionitrile; 3-(N-(2-hydroxyethyl)anilino)-	MANDELIC ACID; ISOPROPYL ESTER
Structure			
Actual Endpoint	Moderate_Severe	Mild	Moderate_Severe
Predicted Endpoint	Moderate_Severe	Mild	Moderate_Severe
Distance	0.492	0.492	0.505
Reference	PESTC* 9;10;80	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;918;86	AJOPAA 29;1363;46

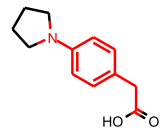
Model Applicability

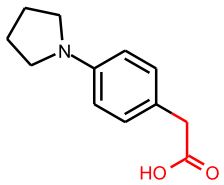
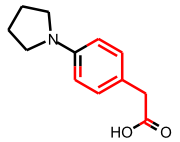
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

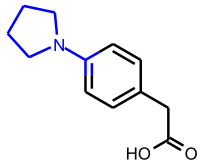
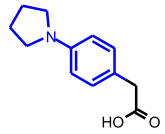
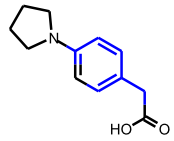
Feature Contribution

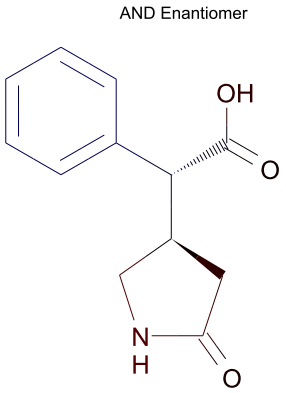
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1985089045	 [*]N([*])[c]1:[cH]:[cH]:[c](CC(=[*])[*]):[cH]:[cH]:1	0.385	16 out of 16

FCFP_10	-1176841573	 [*]=CC(=O)O	0.356	41 out of 43
FCFP_10	-497728148	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.356	24 out of 25

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-822674211	 [*]:[c](:[*])N1CCCC1	-0.842	0 out of 2
FCFP_10	202105689	 [*][c]1:[cH]:[cH]:[c] (:[cH]:[cH]:1)N2CCCC 2	-0.361	2 out of 5
FCFP_10	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.13	153 out of 264



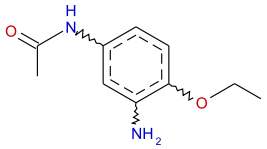
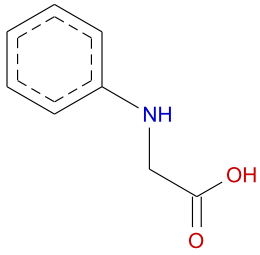
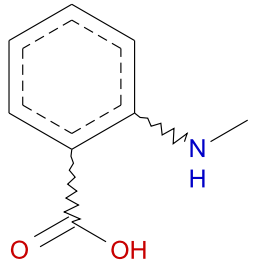
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Moderate_Severe
Probability: 0.832
Enrichment: 1.21
Bayesian Score: 0.66
Mahalanobis Distance: 9.42
Mahalanobis Distance p-value: 0.293

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACETANILIDE; 3'-AMINO-4'-ETHOXY-	ACETIC ACID; CYCLOHEXYLAMINO-	Anthranilic acid; N-methyl-
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Mild
Distance	0.496	0.508	0.518
Reference	28ZPAK 115;72	28ZPAK-;64;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;741;86

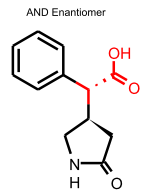
Model Applicability

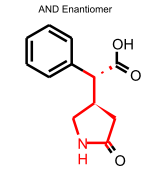
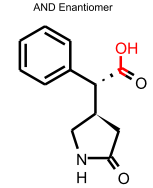
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

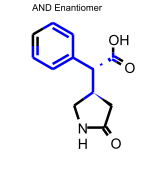
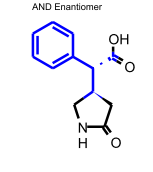
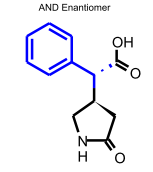
Feature Contribution

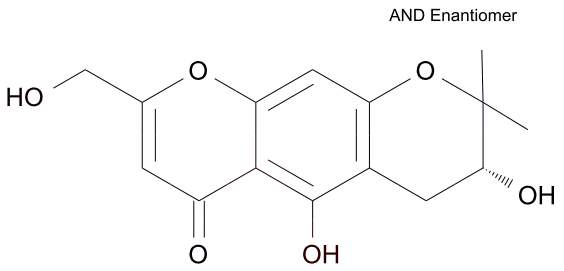
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1176841573	 [*]=CC(=O)O	0.356	41 out of 43

FCFP_10	-416918913	 [*]C([*])[C@@H]1CNC(=[*])C1	0.332	5 out of 5
FCFP_10	-548632217	 [*]C(=[*])O	0.319	54 out of 59

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1133295320	 [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.507	0 out of 1
FCFP_10	-1716639150	 [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.507	0 out of 1
FCFP_10	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.284	53 out of 107



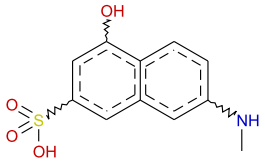
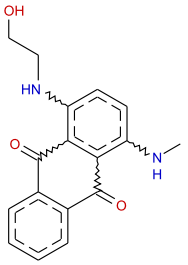
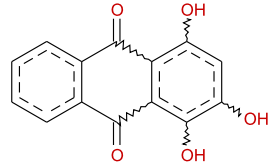
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Moderate_Severe
Probability: 0.836
Enrichment: 1.21
Bayesian Score: 0.995
Mahalanobis Distance: 9.89
Mahalanobis Distance p-value: 0.121

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2-NAPHTHALENESULFONIC ACID; 4-HYDROXY-7-(METHYLAMINO)-	ANTHRAQUINONE; 1-((2-HYDROXYETHYL)AMINO)-4-(METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.614	0.647	0.650
Reference	28ZPAK 190;72	28ZPAK 245;72	28ZPAK-;103;7

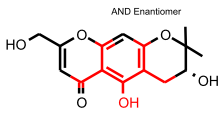
Model Applicability

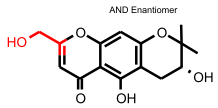
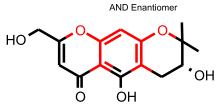
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

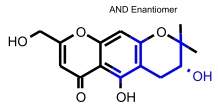
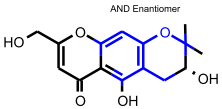
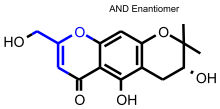
Feature Contribution

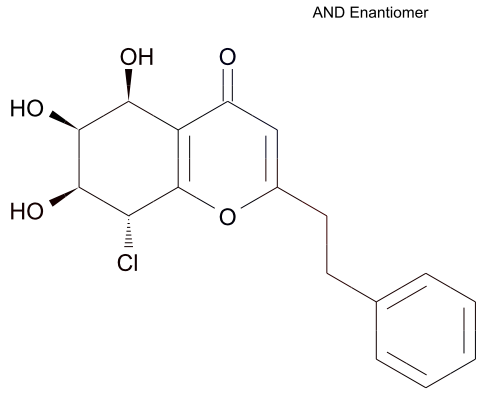
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1601875224	 [*]C[c]1:[c]([*]):[*]:[c]([*]):[c](C(=[*])[*]):[c]:1O	0.352	7 out of 7

FCFP_10	-1272709286	 [*][C@H]1[*][*]NC1	0.285	234 out of 266
FCFP_10	1679744180	 [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]: 1[*]	0.256	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-2006448698	 [*][c](:[*]):[c]1C[C@ @H](O)C([*])([*])[*] [c]:1:[*]	-0.507	0 out of 1
FCFP_10	-1099193755	 [*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:1OC	-0.361	2 out of 5
FCFP_10	436915834	 [*]C\C(=C\[*])\O[*]	-0.194	2 out of 4



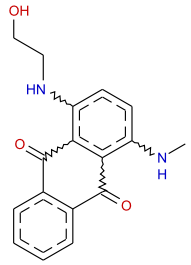
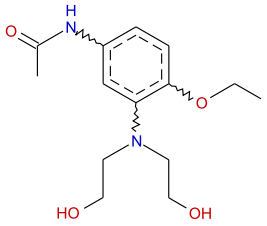
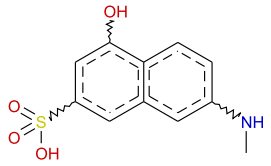
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Moderate_Severe
Probability: 0.836
Enrichment: 1.21
Bayesian Score: 0.958
Mahalanobis Distance: 12.2
Mahalanobis Distance p-value: 1.98e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1-((2-HYDROXYETHYL)AMINO)-4-(METHYLAMINO)-	p-Acetophenetidide; 3'-(bis(2-hydroxyethyl)amino)-	2-NAPHTHALENESULFONIC ACID; 4-HYDROXY-7-(METHYLAMINO)-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.542	0.671	0.687
Reference	28ZPAK 245;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;645;86	28ZPAK 190;72

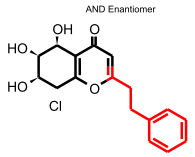
Model Applicability

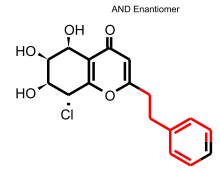
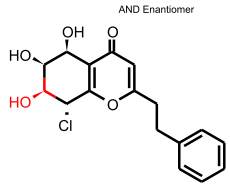
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

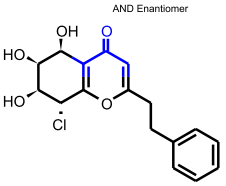
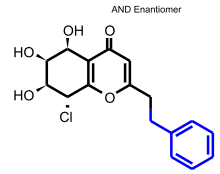
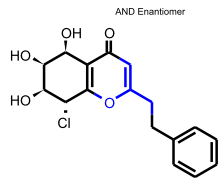
Feature Contribution

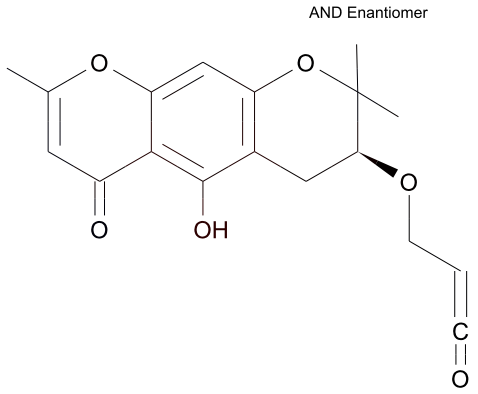
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1388176727	 [*]C(=[*])CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.389	19 out of 19

FCFP_10	-497728148	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.356	24 out of 25
FCFP_10	1070061035	 [*]C([*])O	0.239	284 out of 338

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	565968762	 [*]C(=[*])C(=O)C	-0.372	17 out of 38
FCFP_10	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	436915834	 [*]C\C(=C\[*])\O[*]	-0.194	2 out of 4



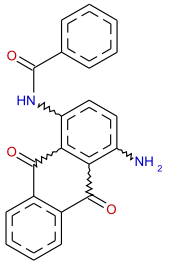
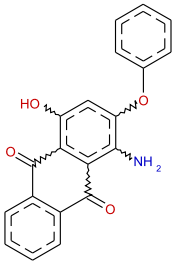
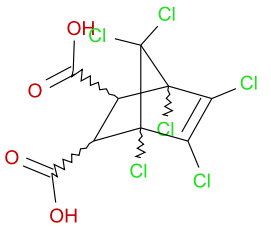
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Moderate_Severe
Probability: 0.829
Enrichment: 1.2
Bayesian Score: 0.464
Mahalanobis Distance: 8.79
Mahalanobis Distance p-value: 0.625

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.608	0.610	0.625
Reference	28ZPAK-;124;72	28ZPAK 239;72	28ZPAK-;92;72

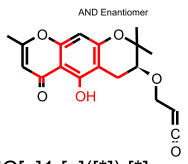
Model Applicability

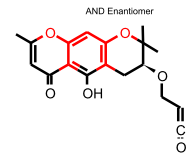
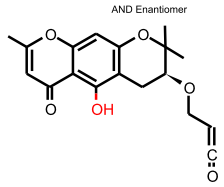
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 444624378: [*]C=C=O

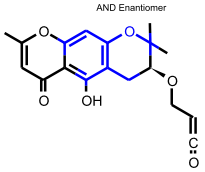
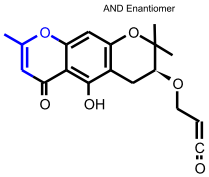
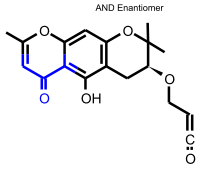
Feature Contribution

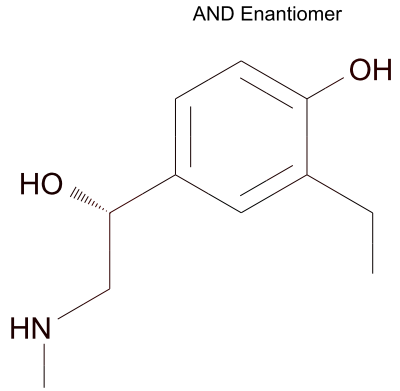
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1601875224	 [*]C[c]1:[c]([*]):[*]:[c]([*]):[c](C(=[*])[*]):[c]:1O	0.352	7 out of 7

FCFP_10	1679744180	 [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]:1[*]	0.256	2 out of 2
FCFP_10	7	 [*]O	0.219	117 out of 142

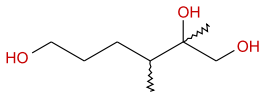
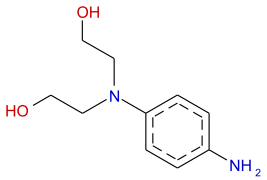
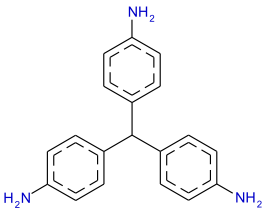
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1099193755	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	-0.361	2 out of 5
FCFP_10	436915834	 [*]C\C(=C\[*])\O[*]	-0.194	2 out of 4
FCFP_10	-1549192822	 [*]C(=[*])C(=O)[c](:[*]):[*]	-0.11	7 out of 12



$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Structural Similar Compounds

Name	2;3-DIMETHYL;1;2;6-HEXANETRIOL	p-Phenylenediamine; N;N-bis(2-hydroxyethyl)-; sulfate (1:1)	METHANE;TRIS(4-AMINOPHENYL)-
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.625	0.653	0.659
Reference	AIHAAP 23;95;62	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;695;86	28ZPAK-;73;72

Model Prediction

Prediction: Moderate_Severe
Probability: 0.848
Enrichment: 1.23
Bayesian Score: 1.84
Mahalanobis Distance: 7.65
Mahalanobis Distance p-value: 0.974

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

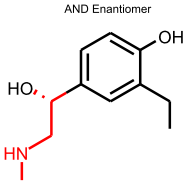
Model Applicability

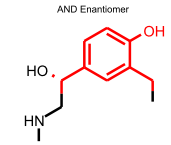
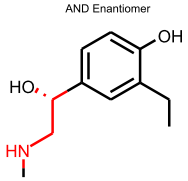
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

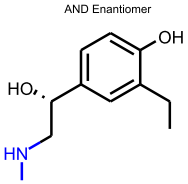
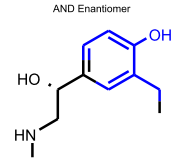
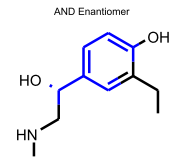
Feature Contribution

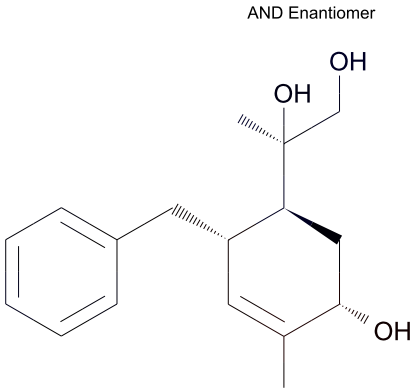
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	226619748	 [*]C([*])CNC	0.294	3 out of 3

FCFP_10	-306804326	 [*]C[c]1:[cH]:[c](:[cH]:[cH]:[c]:1O)C([*])[*]	0.294	3 out of 3
FCFP_10	-1272709286	 [*][C@H]1[*][*]NC1	0.285	234 out of 266

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	136686699	 [*]NC	-0.243	6 out of 12
FCFP_10	-1604301295	 [*]C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1O	-0.164	15 out of 27
FCFP_10	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.13	153 out of 264



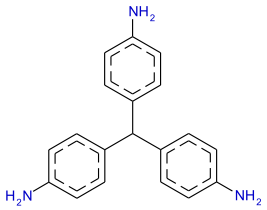
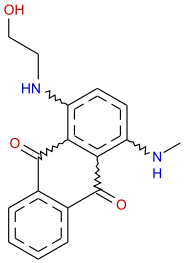
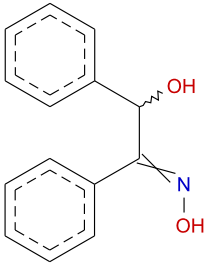
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Moderate_Severe
Probability: 0.824
Enrichment: 1.2
Bayesian Score: 0.142
Mahalanobis Distance: 9.04
Mahalanobis Distance p-value: 0.49

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	METHANE;TRIS(4-AMINOPHENYL)-	ANTHRAQUINONE; 1-((2-HYDROXYETHYL)AMINO)-4-(METHYLAMINO)-	BENZOIN; OXIME
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.565	0.653	0.680
Reference	28ZPAK-;73;72	28ZPAK 245;72	28ZPAK-;111;72

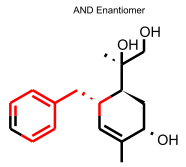
Model Applicability

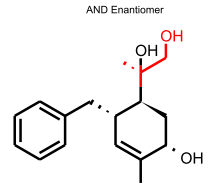
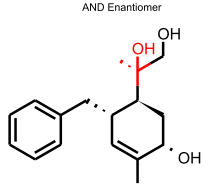
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

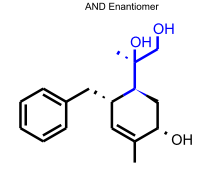
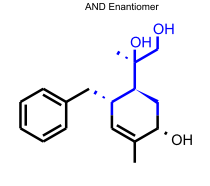
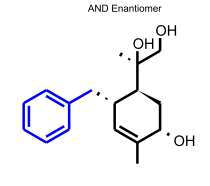
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-497728148	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.356	24 out of 25

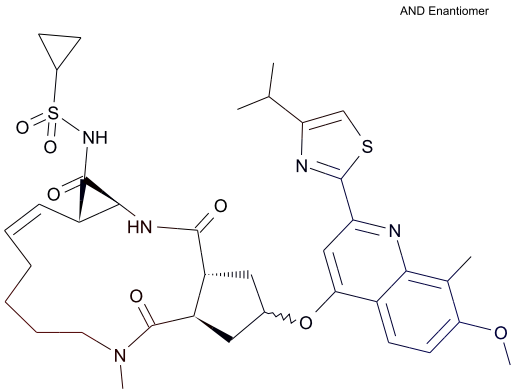
FCFP_10	-1272709286	 [*][C@H]1[*][*]NC1	0.285	234 out of 266
FCFP_10	1070061035	 [*]C([*])O	0.239	284 out of 338

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-2097085266	 [*]C([*])[C@@](C)(O)C O	-0.507	0 out of 1
FCFP_10	-255810400	 [*]C[C@@H](C([*])[*]) [C@@](C)(O)CO	-0.507	0 out of 1
FCFP_10	-1698724694	 [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.284	53 out of 107

Simeprevir

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.777

Enrichment: 1.13

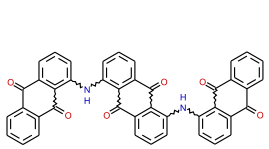
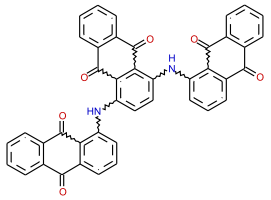
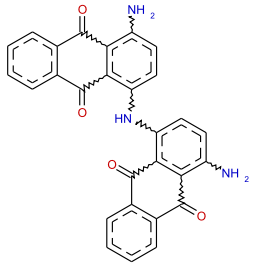
Bayesian Score: -1.77

Mahalanobis Distance: 17.1

Mahalanobis Distance p-value: 9.62e-022

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Anthraquinone; 1;1'-(anthraquinon-1;5-ylenediimino)di-	Anthraquinone; 1;1'-(anthraquinon-1;4-ylenediimino)di-	4;4'-DIAMINO-1;1'-DIANTHRIMIDE
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.875	0.875	1.102
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	28ZPAK-;125;72

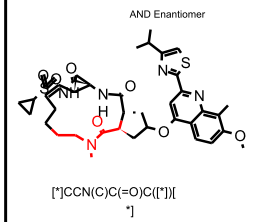
Model Applicability

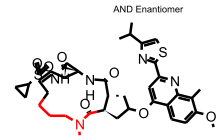
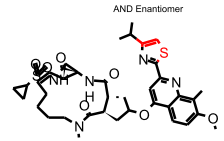
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC15 out of range. Value: 5.5683. Training min, max, SD, explained variance: -4.4073, 5.1625, 1.138, 0.0158.

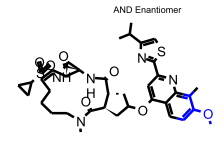
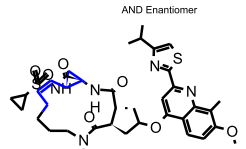
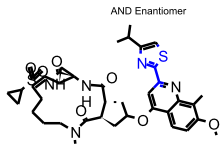
Feature Contribution

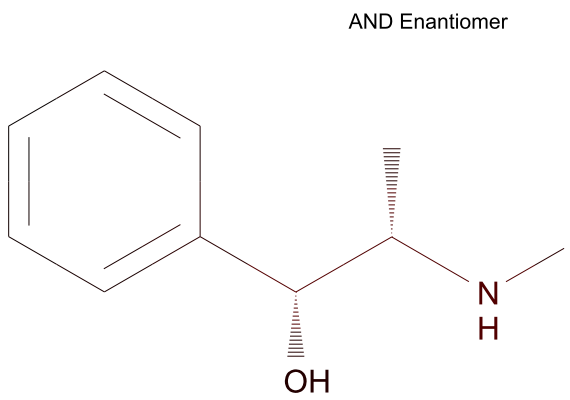
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-2015019763	 [*]CCN(C)C(=O)C([*])[*]	0.294	3 out of 3

FCFP_10	-1474971978	 [*]CCCN(C)C(=[*])[*]	0.259	14 out of 16
FCFP_10	-124655670	 [*][c]1:[*]:[*]:s:[cH]:1	0.259	14 out of 16

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1977641857	 [*][c](:[*]):[c](OC):[cH]:[*]	-0.78	4 out of 15
FCFP_10	-528918648	 [*]C\C=C/[C@@H]1CC1([*])[*]	-0.651	4 out of 13
FCFP_10	690481386	 [*]:[c](:[*])[c]1:n:[*]:[*]:s:1	-0.6	1 out of 4



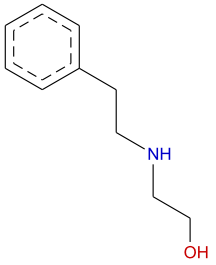
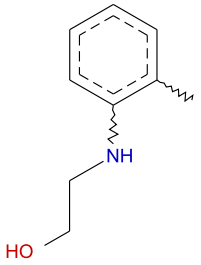
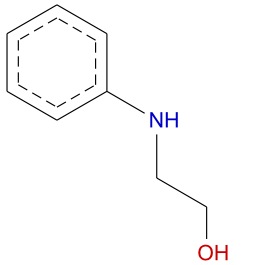
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Severe
Probability: 0.779
Enrichment: 1.26
Bayesian Score: 2.84
Mahalanobis Distance: 4.78
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ETHANOL;2-(PHENETHYLAMINO)-	0-TOLUIDINO-ETHANOL	Ethanol; 2-anilino-
Structure			
Actual Endpoint	Severe	Severe	Moderate
Predicted Endpoint	Severe	Moderate	Moderate
Distance	0.420	0.449	0.493
Reference	AMIHBC 10;61;54	JIHTAB 31;60;49	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;691;86

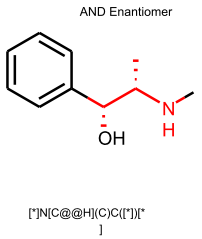
Model Applicability

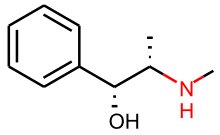
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

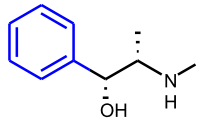
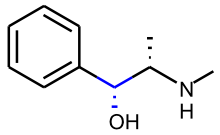
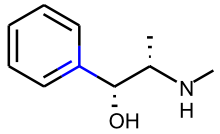
Feature Contribution

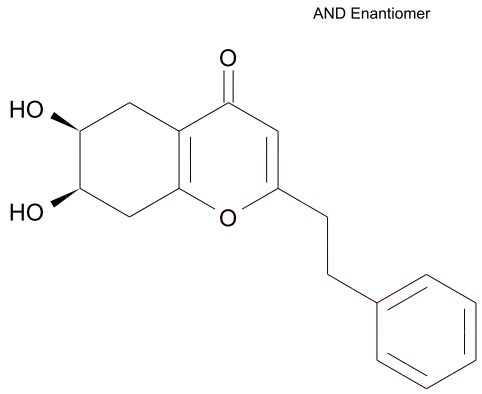
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1430588017		0.469	21 out of 21

SCFP_12	5	AND Enantiomer  [*]N[*]	0.337	129 out of 151
SCFP_12	1702724181	AND Enantiomer  [*]C([*])[C@H](O)[c](:[*]):[*]	0.303	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1653911926	AND Enantiomer  [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0644	33 out of 58
SCFP_12	0	AND Enantiomer  [*]C([*])[*]	0	463 out of 727
SCFP_12	3	AND Enantiomer  [*][c](:[*]):[*]	0	162 out of 280



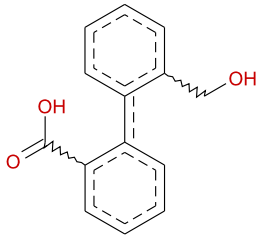
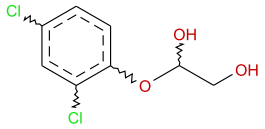
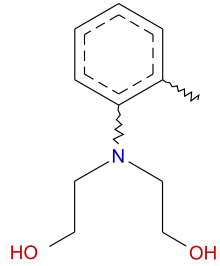
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Severe
Probability: 0.718
Enrichment: 1.16
Bayesian Score: 0.89
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 0.0253

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2-BIPHENYLCARBOXYLIC ACID; 2'-HYDROXYMETHYL-	1;2-ETHANEDIOL;2;4-DICHLOROPHENOXY-	o-Toluidine; N;N-bis(2-hydroxyethyl)-
Structure			
Actual Endpoint	Moderate	Severe	Severe
Predicted Endpoint	Moderate	Severe	Moderate
Distance	0.594	0.633	0.649
Reference	IHFCAY 6;1;67	AMIHBC 4;119;51	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;697;86

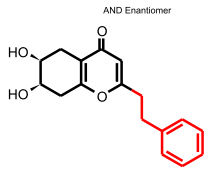
Model Applicability

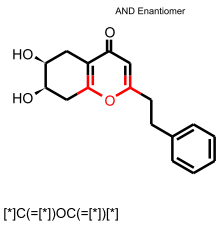
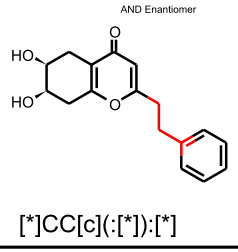
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

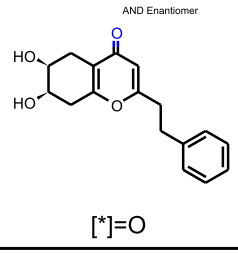
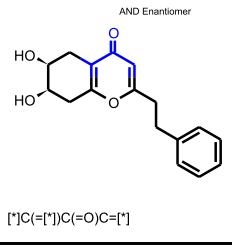
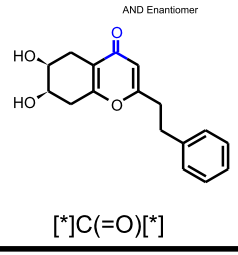
Feature Contribution

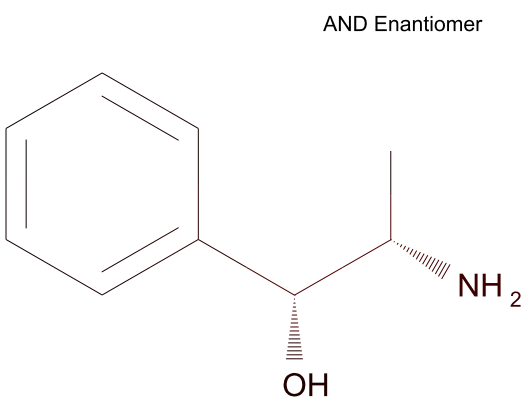
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1640858361	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.376	4 out of 4

SCFP_12	2019034095		0.345	15 out of 17
SCFP_12	-1272709286		0.231	24 out of 31

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	13		-0.105	185 out of 338
SCFP_12	-1980361709		-0.103	1 out of 2
SCFP_12	1311071855		-0.0866	146 out of 262



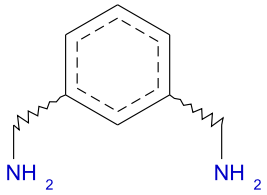
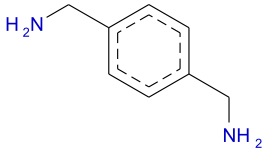
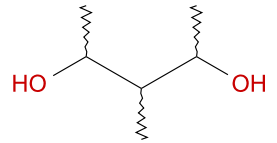
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Severe
Probability: 0.746
Enrichment: 1.2
Bayesian Score: 1.95
Mahalanobis Distance: 4.82
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	METHYLAMINE;M-PHENYLENEBIS-	METHYLAMINE;P-PHENYLENEBIS-	1;2-DIMETHYL;1;3-BUTANEDIOL
Structure			
Actual Endpoint	Severe	Severe	Moderate
Predicted Endpoint	Severe	Severe	Severe
Distance	0.476	0.484	0.494
Reference	28ZPAK-;64;72	28ZPAK-;64;72	AIHAAP 23;95;62

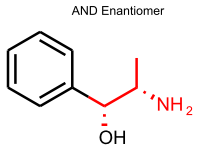
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

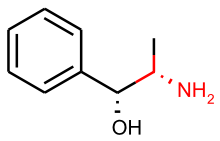
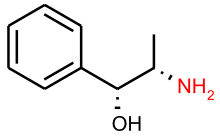
1. All properties and OPS components are within expected ranges.

Feature Contribution

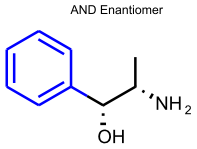
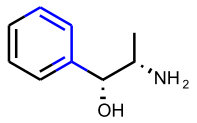
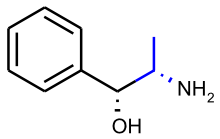
Top features for positive contribution

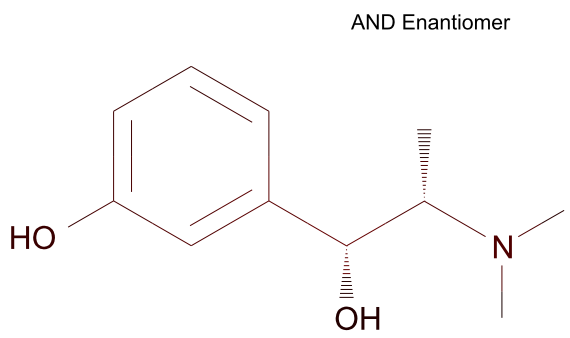
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1430588017		0.469	21 out of 21

[*]N[C@@H](C)C([*])[*]
]

SCFP_12	260714409	AND Enantiomer  [*]C([*])N	0.409	49 out of 53
SCFP_12	5	AND Enantiomer  [*]N[*]	0.337	129 out of 151

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1653911926	AND Enantiomer  [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0644	33 out of 58
SCFP_12	-496409612	AND Enantiomer  [*][c](:[*]):[cH]:[cH]:[*]	0	158 out of 272
SCFP_12	136597326	AND Enantiomer  [*]C([*])C	0	236 out of 373



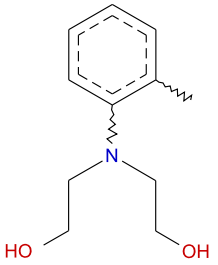
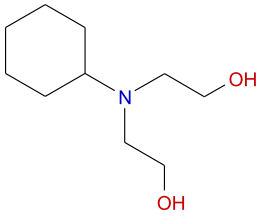
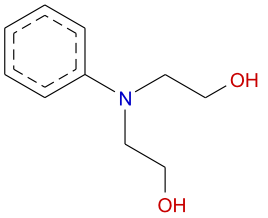
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Severe
Probability: 0.809
Enrichment: 1.3
Bayesian Score: 3.7
Mahalanobis Distance: 4.73
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	o-Toluidine; N;N-bis(2-hydroxyethyl)-	ETHANOL; 2;2'-CYCLOHEXYLIMINODI-	ETHANOL; 2;2'-(PHENYLIMINO)DI-
Structure			
Actual Endpoint	Severe	Severe	Severe
Predicted Endpoint	Moderate	Severe	Moderate
Distance	0.476	0.509	0.531
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;697;86	34ZIAG 61;69	UCDS** 6/13/60

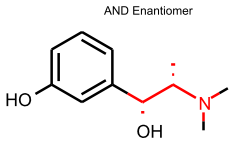
Model Applicability

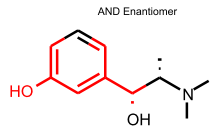
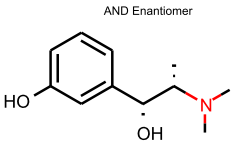
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

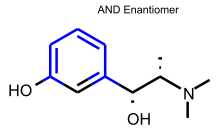
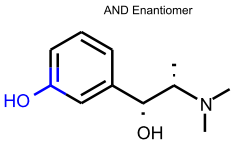
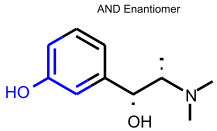
Feature Contribution

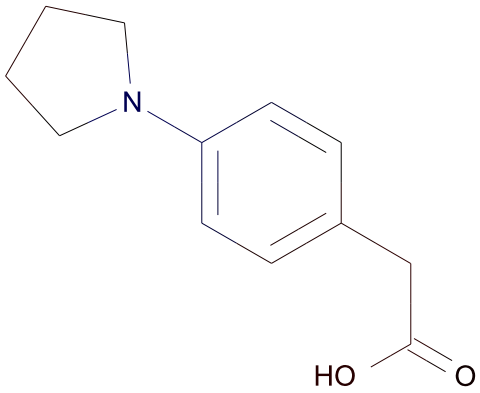
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1430588017	 [*]N[C@@H](C)C([*])[*]	0.469	21 out of 21

SCFP_12	125999298	 [*]C([*])[c]1:[cH]:[*]:[cH]:[c](O):[cH]:1	0.42	7 out of 7
SCFP_12	5	 [*]N[*]	0.337	129 out of 151

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1379591900	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	77 out of 130
SCFP_12	-424425761	 [*]:[c](:[*])O	0	40 out of 63
SCFP_12	560173167	 [*]:[cH]:[c](O):[cH]:[*]	0	68 out of 110



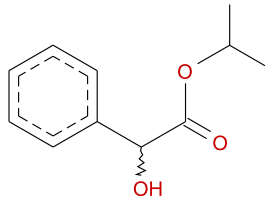
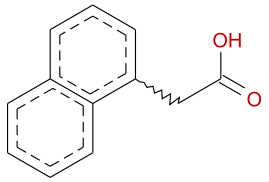
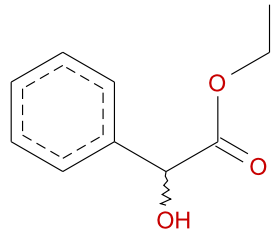
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Severe
Probability: 0.682
Enrichment: 1.1
Bayesian Score: -0.266
Mahalanobis Distance: 6.53
Mahalanobis Distance p-value: 0.999

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	MANDELIC ACID; ISOPROPYL ESTER	1-NAPHTHALENEACETIC ACID	MANDELIC ACID; ETHYL ESTER
Structure			
Actual Endpoint	Moderate	Severe	Moderate
Predicted Endpoint	Moderate	Severe	Moderate
Distance	0.481	0.485	0.534
Reference	AJOPAA 29;1363;46	PESTC* 9;10;80	AJOPAA 29;1363;46

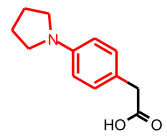
Model Applicability

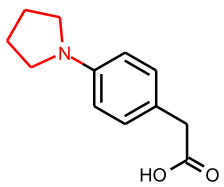
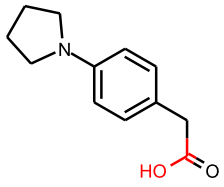
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

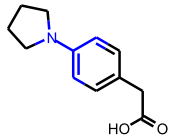
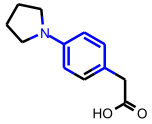
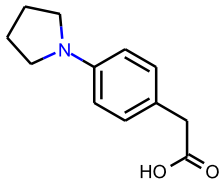
Feature Contribution

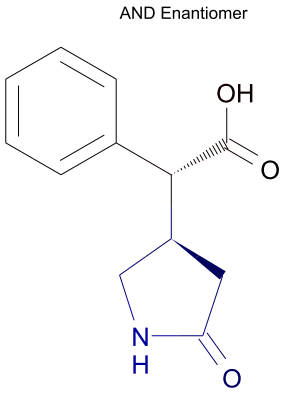
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-658389839	 [*][c]1:[cH]:[cH]:[c](:[cH]:[cH]:1)N2CCCC2	0.303	2 out of 2

SCFP_12	-2056996303	 [*]N1CCCC1	0.303	2 out of 2
SCFP_12	-424485343	 [*]C(=[*])O	0.263	43 out of 54

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1334669481	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.685	28 out of 93
SCFP_12	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.506	26 out of 72
SCFP_12	10	 [*]N([*])[*]	-0.451	43 out of 112



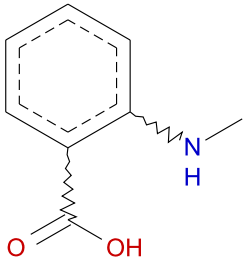
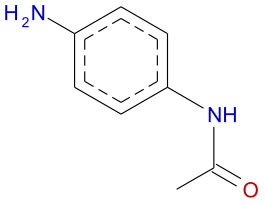
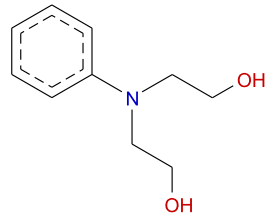
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Moderate
Probability: 0.549
Enrichment: 0.886
Bayesian Score: -3.24
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 0.000134

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Anthranilic acid; N-methyl-	ACETANILIDE. 4'-AMINO-	ETHANOL; 2;2'-(PHENYLIMINO)DI-
Structure			
Actual Endpoint	Moderate	Moderate	Severe
Predicted Endpoint	Moderate	Moderate	Moderate
Distance	0.557	0.596	0.609
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;741;86	28ZPAK-;130;72	UCDS** 6/13/60

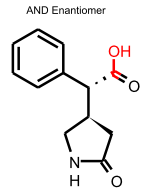
Model Applicability

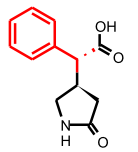
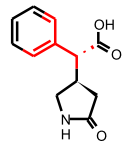
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

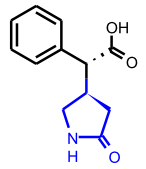
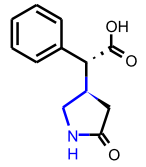
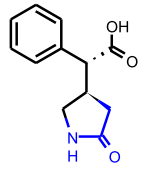
Feature Contribution

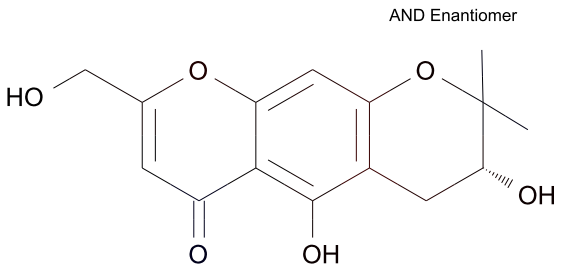
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-424485343	 [*]C(=[*])O	0.263	43 out of 54

SCFP_12	-1631132401	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.104	19 out of 28
SCFP_12	2109165795	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.0868	79 out of 119

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	2005026407	AND Enantiomer  [*][C@@H]1CNC(=O)C1	-0.796	0 out of 2
SCFP_12	-587569116	AND Enantiomer  [*][C@H]1[*][*]NC1	-0.619	11 out of 35
SCFP_12	1256995004	AND Enantiomer  O=C1C[*][*]N1	-0.483	12 out of 33



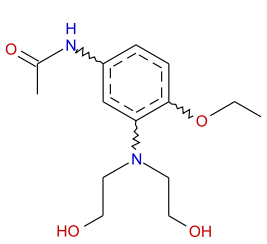
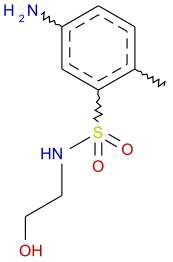
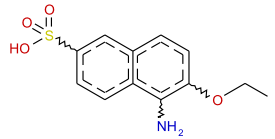
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Moderate
Probability: 0.654
Enrichment: 1.05
Bayesian Score: -1.03
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 0.0213

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	p-Acetophenetidide; 3'-(bis(2-hydroxyethyl)amino)-	O-TOLUENESULFONAMIDE; 4-AMINO-N-(2-HYDROXYETHYL)-	2-NAPHTHALENESULFONIC ACID;5-AMINO-6-ETHOXY-
Structure			
Actual Endpoint	Moderate	Moderate	Moderate
Predicted Endpoint	Moderate	Moderate	Moderate
Distance	0.694	0.745	0.746
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;645;86	28ZPAK-;200;72	28ZPAK-;191;72

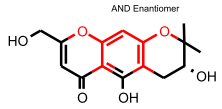
Model Applicability

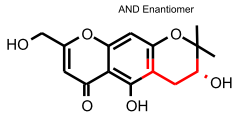
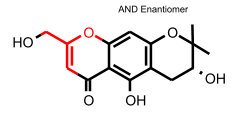
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

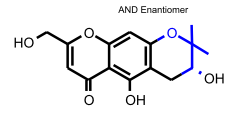
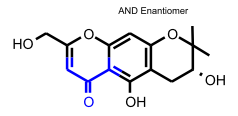
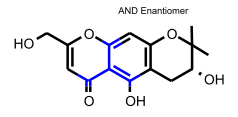
Feature Contribution

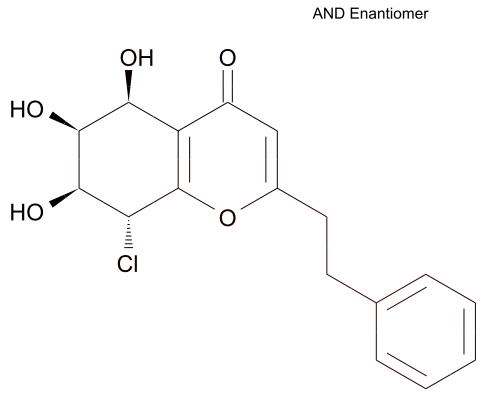
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	130348166	 [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]:1[*]	0.376	4 out of 4

SCFP_12	-1272709286	 [*]CC[c](:[*]):[*]	0.231	24 out of 31
SCFP_12	616547045	 [*]C\C(=C\[*])\O[*]	0.218	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1416196903	 [*]OC(C)(C)C([*])[*]	-0.383	3 out of 8
SCFP_12	-1980302127	 [*]=CC(=O)[c](:[*]):[*]	-0.345	1 out of 3
SCFP_12	-2056718782	 [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.163	20 out of 39



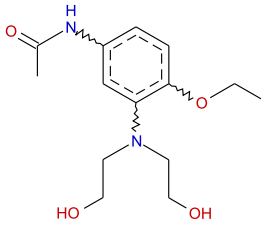
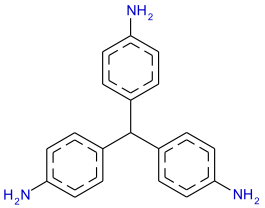
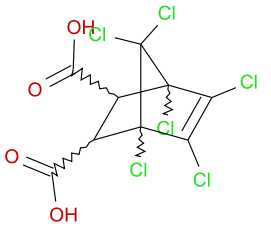
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Severe
Probability: 0.728
Enrichment: 1.17
Bayesian Score: 1.23
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 0.000106

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	p-Acetophenetidide; 3'-(bis(2-hydroxyethyl)amino)-	METHANE;TRIS(4-AMINOPHENYL)-	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Moderate	Moderate	Severe
Predicted Endpoint	Moderate	Moderate	Severe
Distance	0.665	0.730	0.759
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;645;86	28ZPAK-;73;72	28ZPAK-;92;72

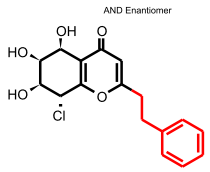
Model Applicability

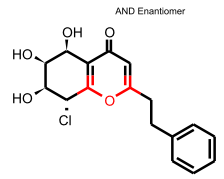
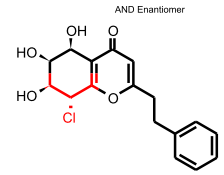
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

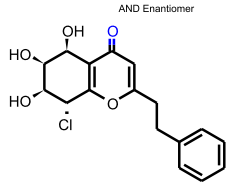
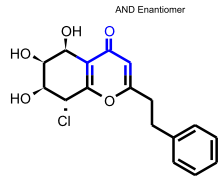
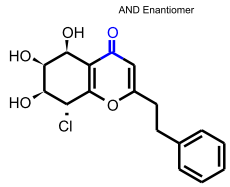
Feature Contribution

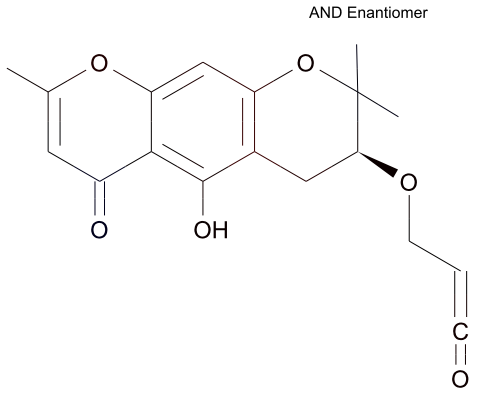
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1640858361	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.376	4 out of 4

SCFP_12	2019034095	 [*]C(=[*])OC(=[*])[*]	0.345	15 out of 17
SCFP_12	540920128	 [*]C([*])[C@@H](Cl)C(=[*])[*]	0.303	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	13	 [*]=O	-0.105	185 out of 338
SCFP_12	-1980361709	 [*]C(=[*])C(=O)C=[*]	-0.103	1 out of 2
SCFP_12	1311071855	 [*]C(=O)[*]	-0.0866	146 out of 262



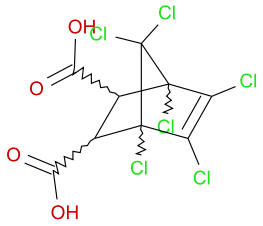
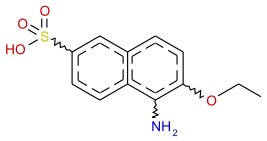
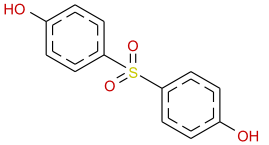
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Severe
Probability: 0.709
Enrichment: 1.14
Bayesian Score: 0.548
Mahalanobis Distance: 13
Mahalanobis Distance p-value: 4.98e-008

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	2-NAPHTHALENESULFONIC ACID;5-AMINO-6-ETHOXY-	PHENOL; 4;4'-SULFONYLDI-
Structure			
Actual Endpoint	Severe	Moderate	Moderate
Predicted Endpoint	Severe	Moderate	Moderate
Distance	0.622	0.676	0.697
Reference	28ZPAK-;92;72	28ZPAK-;191;72	BIOFX* 601-05501;74

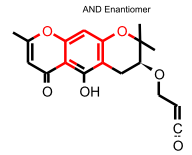
Model Applicability

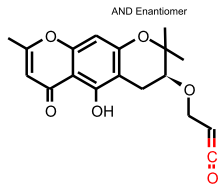
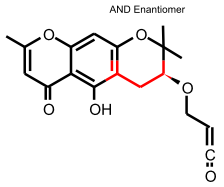
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

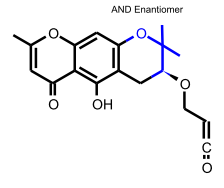
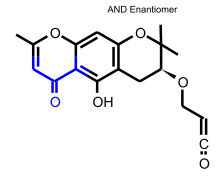
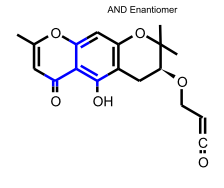
Feature Contribution

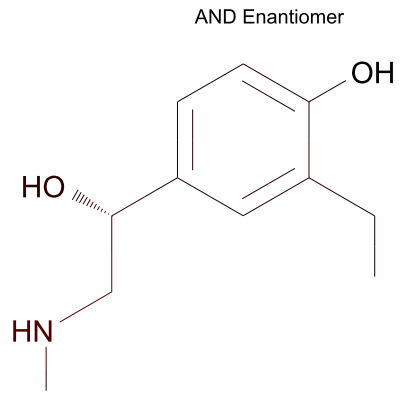
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	130348166	 [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]:1[*]	0.376	4 out of 4

SCFP_12	1311101646	 [*]=C=O	0.311	7 out of 8
SCFP_12	-1272709286	 [*]CC[c](:[*]):[*]	0.231	24 out of 31

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1416196903	 [*]OC(C)(C)C([*])[*]	-0.383	3 out of 8
SCFP_12	-1980302127	 [*]=CC(=O)[c](:[*]):[*]	-0.345	1 out of 3
SCFP_12	-2056718782	 [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.163	20 out of 39



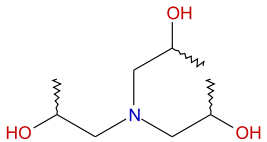
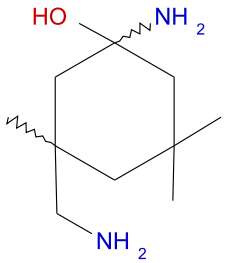
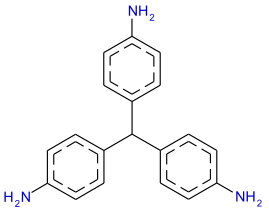
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Severe
Probability: 0.769
Enrichment: 1.24
Bayesian Score: 2.56
Mahalanobis Distance: 5.02
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2-PROPANOL; 1;1';1''-NITRILOTRI-	1-AMINO-3-AMINOMETHYL-3;5;5-TRIMETHYLCYCLOHEXANOL	METHANE;TRIS(4-AMINOPHENYL)-
Structure			
Actual Endpoint	Severe	Severe	Moderate
Predicted Endpoint	Severe	Severe	Moderate
Distance	0.651	0.660	0.673
Reference	AJOPAA 29;1363;46	AIHAAP 30;470;69	28ZPAK-;73;72

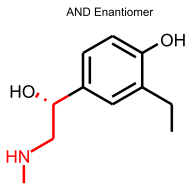
Model Applicability

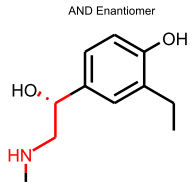
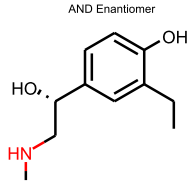
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

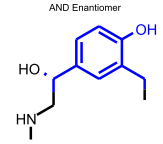
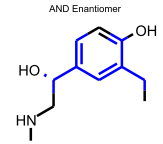
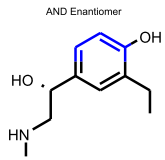
Feature Contribution

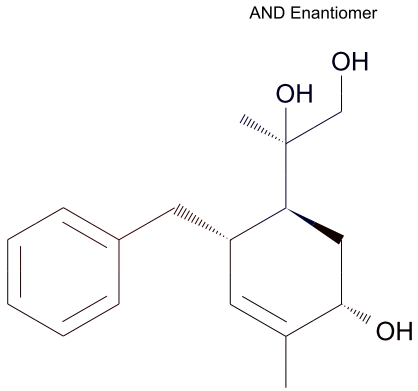
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	696607952	 [*]C([*])CNC	0.348	3 out of 3

SCFP_12	2088704928	 [*]NCC([*])[*]	0.342	110 out of 128
SCFP_12	5	 [*]N[*]	0.337	129 out of 151

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1848446841	 [*]C[c]1:[cH]:[c](:[cH]:[cH]:[c]:1O)C([*])[*]	-0.345	1 out of 3
SCFP_12	-350503170	 [*]C[c]1:[cH]:[c](:[cH]:[*]:[c]:1[*])C([*])[*]	-0.0702	9 out of 16
SCFP_12	-496409612	 [*][c](:[*]):[cH]:[cH]:[*]	0	158 out of 272



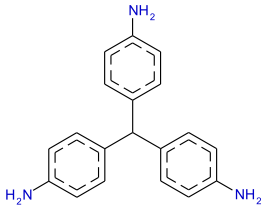
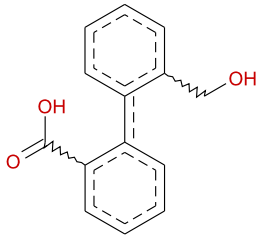
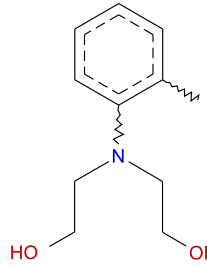
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Moderate
Probability: 0.675
Enrichment: 1.09
Bayesian Score: -0.469
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 6.26e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	METHANE;TRIS(4-AMINOPHENYL)-	2-BIPHENYLCARBOXYLIC ACID; 2'-HYDROXYMETHYL-	o-Toluidine; N;N-bis(2-hydroxyethyl)-
Structure			
Actual Endpoint	Moderate	Moderate	Severe
Predicted Endpoint	Moderate	Moderate	Moderate
Distance	0.579	0.696	0.713
Reference	28ZPAK-;73;72	IHFCAY 6;1;67	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;697;86

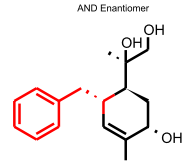
Model Applicability

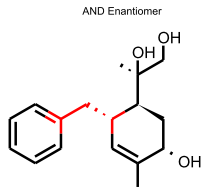
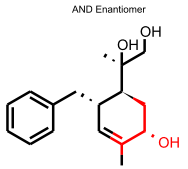
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

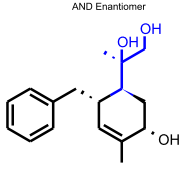
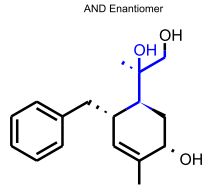
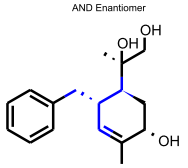
Feature Contribution

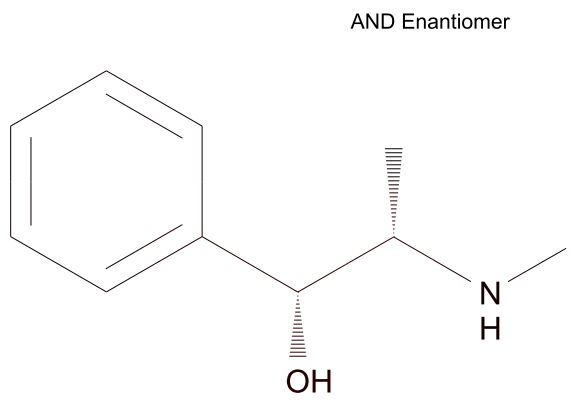
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1640858361	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.376	4 out of 4

SCFP_12	-1272709286	 [*]CC[c](:[*]):[*]	0.231	24 out of 31
SCFP_12	1702664599	 [*]C([*])[C@@H](O)C(=[*])[*]	0.212	7 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1847692671	 [*]C([*])[C@@](C)(O)CO	-0.475	0 out of 1
SCFP_12	1416196903	 [*]OC(C)(C)C([*])[*]	-0.383	3 out of 8
SCFP_12	-1043310069	 [*]C[C@@H](C=[*])C([*])[*]	-0.246	15 out of 32



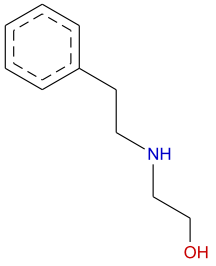
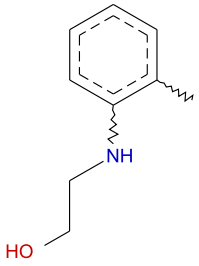
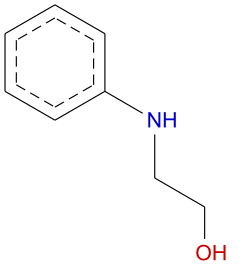
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.47
Mahalanobis Distance: 5.04
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ETHANOL;2-(PHENETHYLAMINO)-	0-TOLUIDINO-ETHANOL	Ethanol; 2-anilino-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.425	0.436	0.484
Reference	AMIHBC 10;61;54	JIHTAB 31;60;49	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;691;86

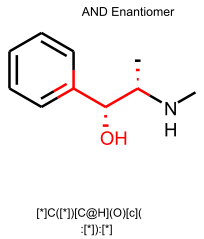
Model Applicability

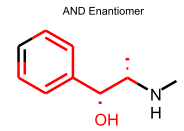
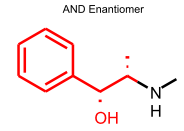
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

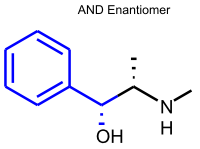
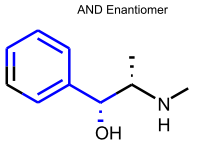
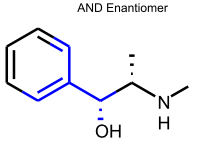
Feature Contribution

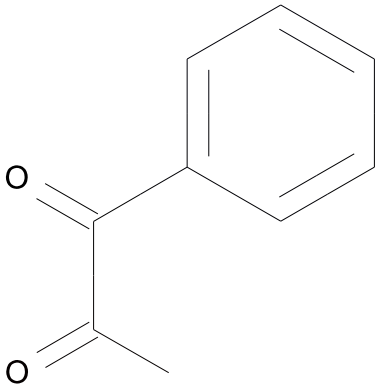
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1186393305	 [*]C([*])[C@H](O)[c](:[*]):[*]	0.197	13 out of 13

FCFP_12	-1931573337	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	0.195	12 out of 12
FCFP_12	-1083860676	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[cH]:[cH] :[cH]:1	0.192	10 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0	264 out of 323
FCFP_12	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382



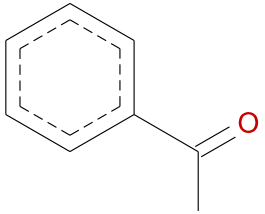
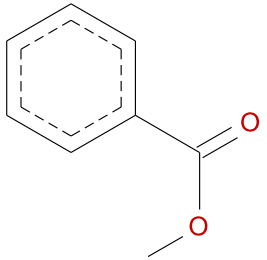
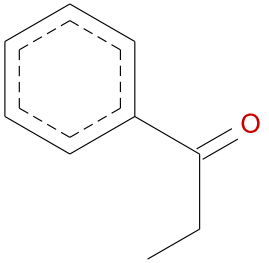
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Irritant
Probability: 0.976
Enrichment: 1.15
Bayesian Score: 0.192
Mahalanobis Distance: 3.35
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACETOPHENONE	METHYL BENZOATE	Propiophenone
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.310	0.380	0.389
Reference	AJOPAA 29;1363;46	AMIHBC 10;61;54	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;291;86

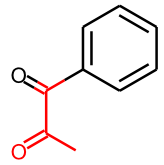
Model Applicability

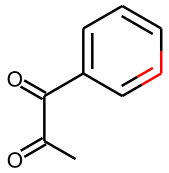
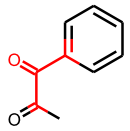
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

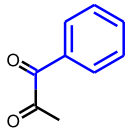
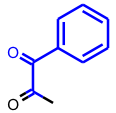
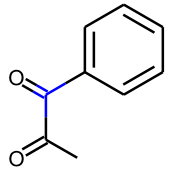
Feature Contribution

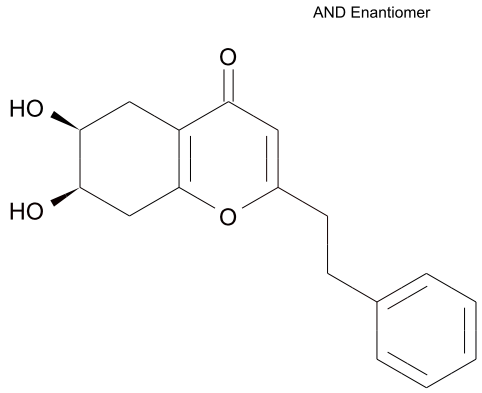
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.11	38 out of 42

FCFP_12	16	 [*][c](:[*]):[*]	0.0586	442 out of 516
FCFP_12	-1549192822	 [*]C(=[*])C(=O)[c](:[*]):[*]	0.0547	12 out of 14

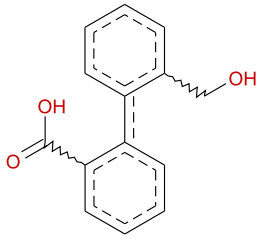
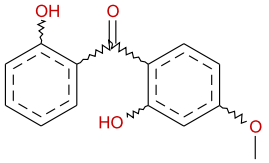
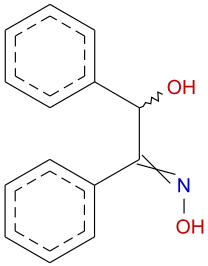
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	975909016	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0639	6 out of 8
FCFP_12	0	 [*]C([*])[*]	0	1184 out of 1397



$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Structural Similar Compounds

Name	2-BIPHENYLCARBOXYLIC ACID; 2'-HYDROXYMETHYL-	2;2'-Dihydroxy-4-methoxybenzophenone	BENZOIN; OXIME
Structure			
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.585	0.585	0.602
Reference	IHFCAY 6;1;67	J. Am. Coll. Toxicol. 2(5):35;1983	28ZPAK-;111;72

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 0.52
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.0363

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

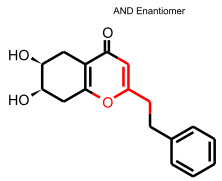
Model Applicability

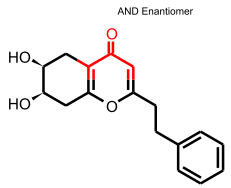
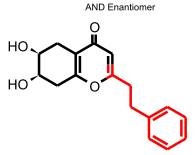
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

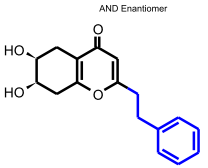
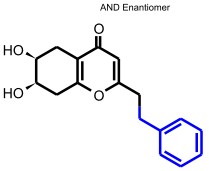
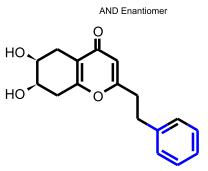
Feature Contribution

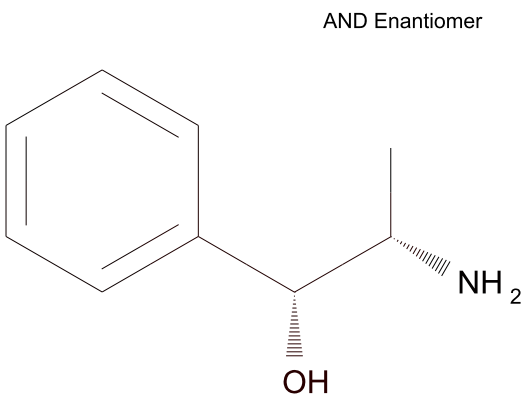
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.167	4 out of 4

FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.11	38 out of 42
FCFP_12	1388176727	 [*]C(=[*])CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.107	19 out of 21

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1981711554	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.103	5 out of 7
FCFP_12	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291



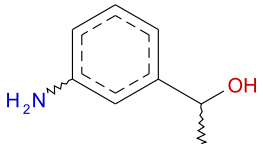
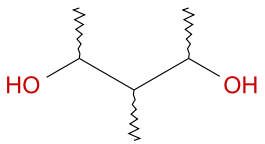
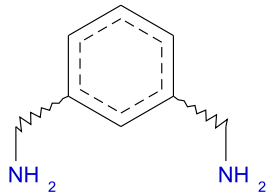
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.14
Mahalanobis Distance: 4.97
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3-AMINO-2-METHYL BENZYL ALCOHOL	1;2-DIMETHYL;1;3-BUTANEDIOL	METHYLAMINE;M-PHENYLENEBIS-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.339	0.453	0.455
Reference	AMIHBC 10;61;54	AIHAAP 23;95;62	28ZPAK-;64;72

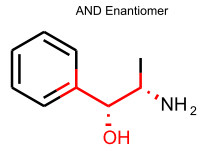
Model Applicability

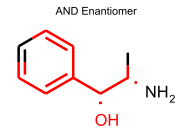
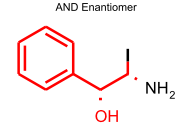
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

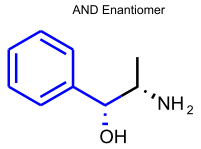
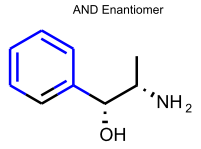
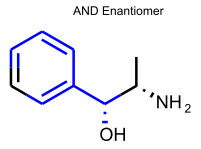
Feature Contribution

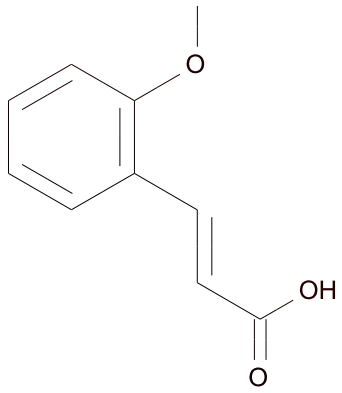
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1186393305	 [*]C([*])[C@H](O)[c](:[*]):[*]	0.197	13 out of 13

FCFP_12	-1931573337	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	0.195	12 out of 12
FCFP_12	-1083860676	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[cH]:[cH] :[cH]:1	0.192	10 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	991735244	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	0	237 out of 291
FCFP_12	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0	264 out of 323



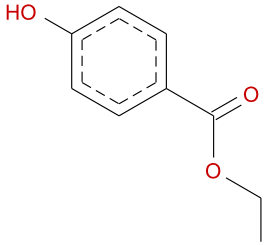
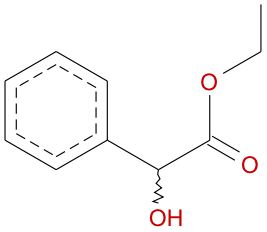
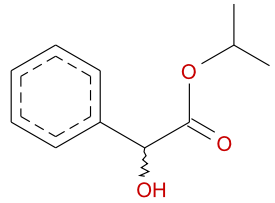
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.11
Mahalanobis Distance: 6.09
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ethyl Paraben	MANDELIC ACID; ETHYL ESTER	MANDELIC ACID; ISOPROPYL ESTER
Structure			
Actual Endpoint	Non-Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.458	0.467	0.467
Reference	J. Am. Coll. Toxicol. 3(5):147;1984	AJOPAA 29;1363;46	AJOPAA 29;1363;46

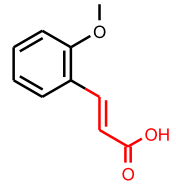
Model Applicability

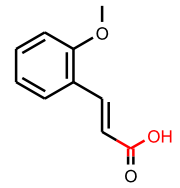
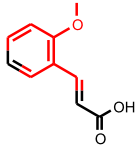
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

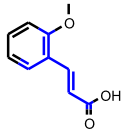
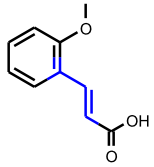
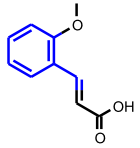
Feature Contribution

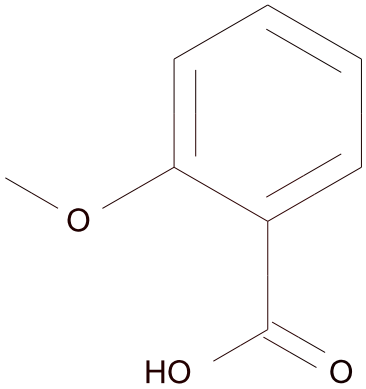
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-2107131107	 [*]\C=C\C(=O)O	0.18	6 out of 6

FCFP_12	-548632217	 [*]C(=[*])O	0.177	59 out of 61
FCFP_12	-1099193755	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	0.175	5 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	-0.268	1 out of 2
FCFP_12	451371068	 [*]\C=C\[c](:[*]):[*]	-0.167	6 out of 9
FCFP_12	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146



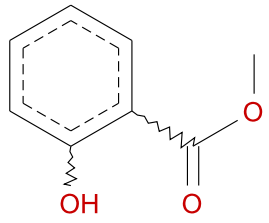
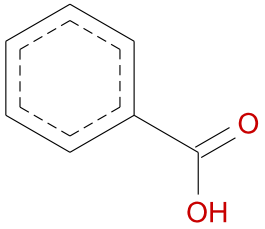
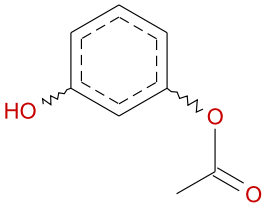
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.58
Mahalanobis Distance: 4.77
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	SALICYCLIC ACID; METHYL ESTER	BENZOIC ACID	RESORCINOL; MONOACETATE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.355	0.365	0.385
Reference	28ZPAK-;106;72	BIOFX* 28-4/73	JAPMA8 46;185;57

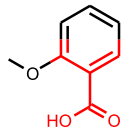
Model Applicability

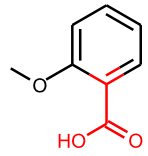
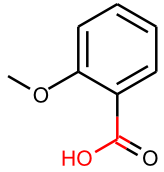
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

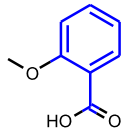
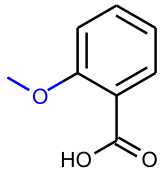
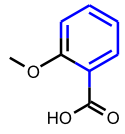
Feature Contribution

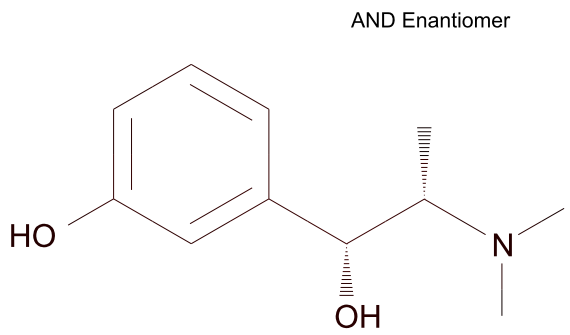
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	937923569	 [*][c](:[*]):[c](:[cH]:[*])C(=O)O	0.198	14 out of 14

FCFP_12	-1549222613	 [*]:[c](:[*])C(=O)O	0.198	14 out of 14
FCFP_12	-548632217	 [*]C(=[*])O	0.177	59 out of 61

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	136627117	 [*]OC	0	96 out of 113
FCFP_12	991735244	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	0	237 out of 291



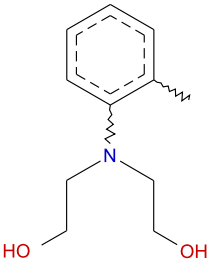
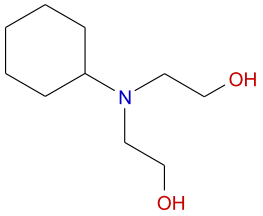
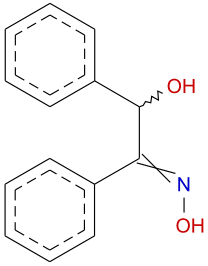
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 2.43
Mahalanobis Distance: 4.92
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	o-Toluidine; N;N-bis(2-hydroxyethyl)-	ETHANOL; 2;2'-CYCLOHEXYLIMINODI-	BENZOIN; OXIME
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.476	0.505	0.512
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;697;86	34ZIAG 61;69	28ZPAK-;111;72

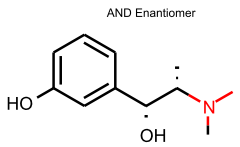
Model Applicability

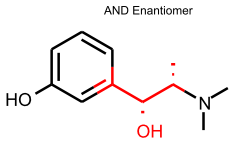
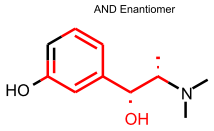
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

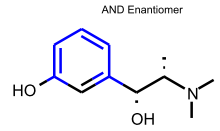
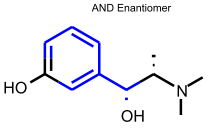
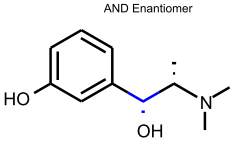
Feature Contribution

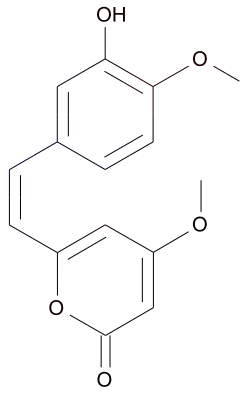
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	136388789	 [*]N([*])C	0.206	30 out of 30

FCFP_12	1186393305	 [*]C([*])[C@H](O)[c](:[*]):[*]	0.197	13 out of 13
FCFP_12	-1931573337	 [*]C([*])[C@H](O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.195	12 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0	264 out of 323
FCFP_12	0	 [*]C([*])[*]	0	1184 out of 1397



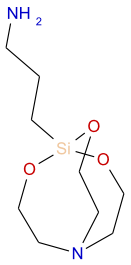
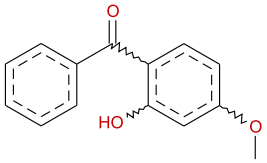
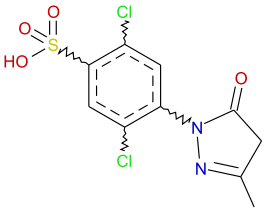
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 0.976
Enrichment: 1.15
Bayesian Score: 0.185
Mahalanobis Distance: 9
Mahalanobis Distance p-value: 0.512

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2;8;9-TRIOXA-5-AZA-1-SILABICYCLO(3.3.3)UNDECANE;1-(3-AMINOPROPYL)-	2-Hydroxy-4-methoxybenzophenone	2;5-DICHLORO-4(3'-METHYL-5' PYRAZOLON-1'-YL)BENZENE SULFONIC ACID
Structure			
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.579	0.634	0.647
Reference	28ZPAK-;220;72	J. Am. Coll. Toxicol. 2(5):35;1983	28ZPAK-;186;72

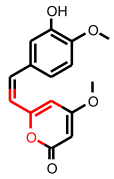
Model Applicability

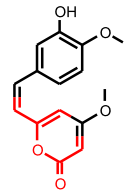
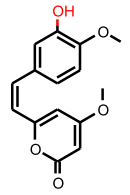
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

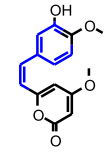
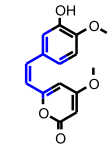
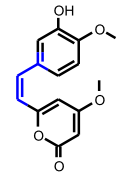
Feature Contribution

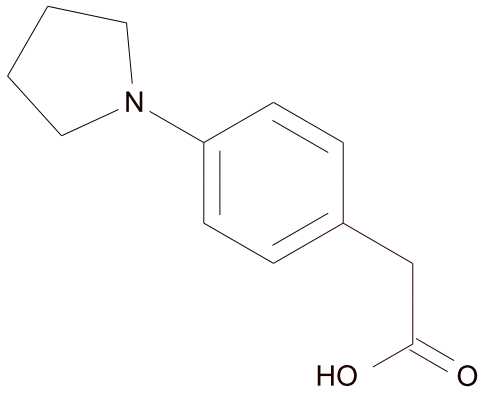
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.167	4 out of 4

FCFP_12	-166452859	 [*]=CC1=C[*]=CC(=O)O1	0.137	2 out of 2
FCFP_12	7	 [*]O	0.119	142 out of 156

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1078052987	 [*]\C=C/[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	-0.344	2 out of 4
FCFP_12	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	-0.268	1 out of 2
FCFP_12	451371068	 [*]\C=C\[c](:[*]):[*]	-0.167	6 out of 9



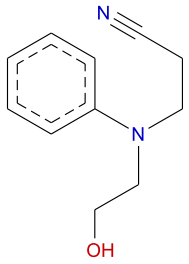
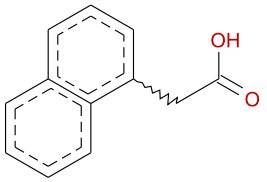
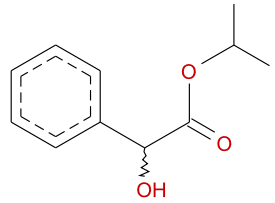
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.6
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.00337

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Propionitrile; 3-(N-(2-hydroxyethyl)anilino)-	1-NAPHTHALENEACETIC ACID	MANDELIC ACID; ISOPROPYL ESTER
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.482	0.487	0.501
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;918;86	PESTC* 9;10;80	AJOPAA 29;1363;46

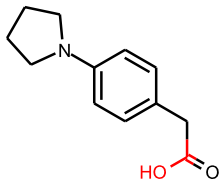
Model Applicability

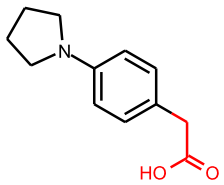
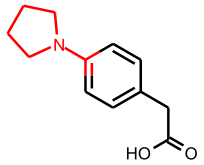
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

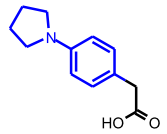
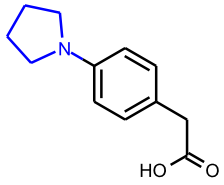
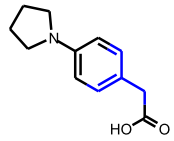
Feature Contribution

Top features for positive contribution

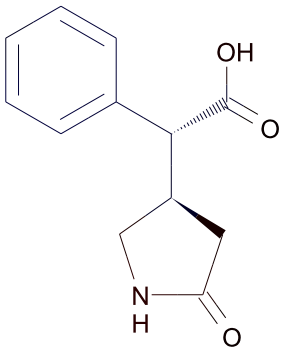
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-548632217	 [*]C(=[*])O	0.177	59 out of 61

FCFP_12	-1176841573	 [*]=CC(=O)O	0.164	43 out of 45
FCFP_12	-822674211	 [*]:[c](:[*])N1CCCC1	0.137	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	202105689	 [*][c]1:[cH]:[cH]:[c](:[cH]:[cH]:1)N2CCCC2	-0.103	5 out of 7
FCFP_12	-98332825	 [*]N1CCCC1	-0.0663	105 out of 139
FCFP_12	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382

AND Enantiomer



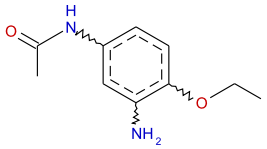
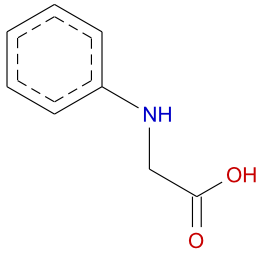
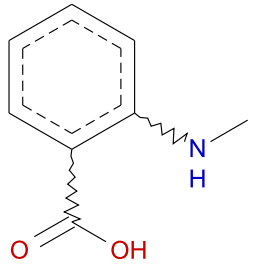
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 0.426
Mahalanobis Distance: 7.63
Mahalanobis Distance p-value: 0.976

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACETANILIDE; 3'-AMINO-4'-ETHOXY-	ACETIC ACID; CYCLOHEXYLAMINO-	Anthranilic acid; N-methyl-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.494	0.500	0.508
Reference	28ZPAK 115;72	28ZPAK-;64;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;741;86

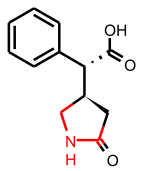
Model Applicability

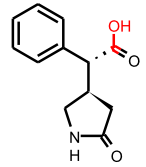
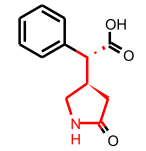
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

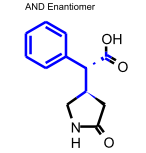
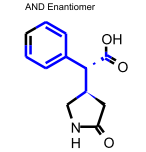
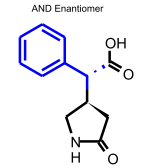
Feature Contribution

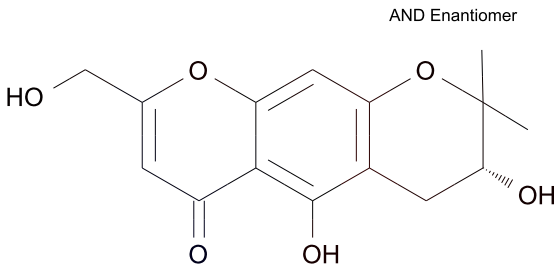
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-885550502	AND Enantiomer  [*]C([*])NC	0.18	64 out of 66

FCFP_12	-548632217	AND Enantiomer  [*]C(=[*])O	0.177	59 out of 61
FCFP_12	-416918913	AND Enantiomer  [*]C([*])[C@@H]1CNC(=[*])C1	0.175	5 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1716639150	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.537	1 out of 3
FCFP_12	-1133295320	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.537	1 out of 3
FCFP_12	-1698724694	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146



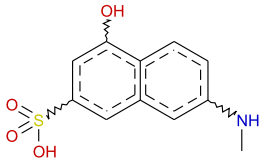
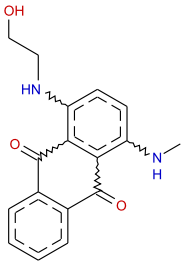
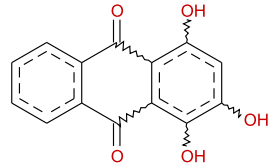
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.19
Mahalanobis Distance: 8.93
Mahalanobis Distance p-value: 0.551

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2-NAPHTHALENESULFONIC ACID; 4-HYDROXY-7-(METHYLAMINO)-	ANTHRAQUINONE; 1-((2-HYDROXYETHYL)AMINO)-4-(METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.606	0.636	0.638
Reference	28ZPAK 190;72	28ZPAK 245;72	28ZPAK-;103;7

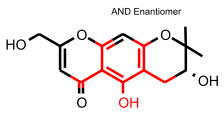
Model Applicability

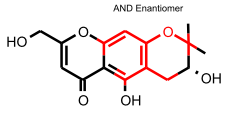
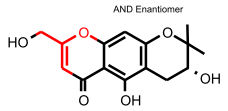
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

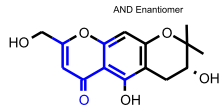
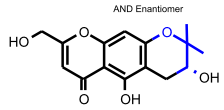
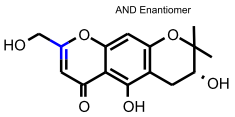
Feature Contribution

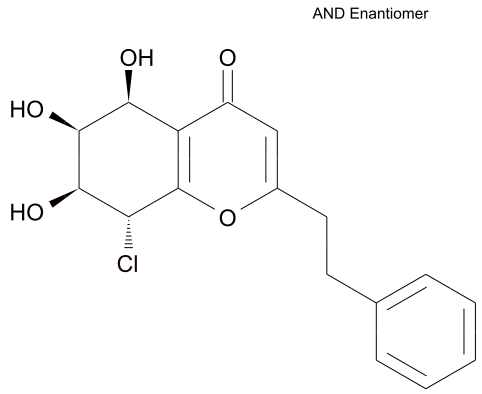
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1601875224	 [*]C[c]1:[c]([*]):[*]:[c]([*]):[c](C(=[*])[*]):[c]:1O	0.184	7 out of 7

FCFP_12	-1099193755	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	0.175	5 out of 5
FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.167	4 out of 4

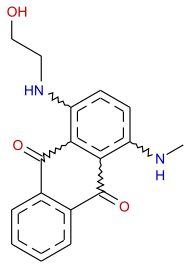
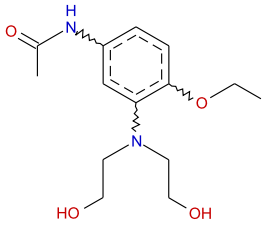
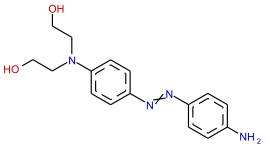
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1244036906	 [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*]	-0.592	0 out of 1
FCFP_12	-415216134	 [*]OC(C)(C)C([*])[*]	-0.0889	11 out of 15
FCFP_12	0	 [*]C([*])[*]	0	1184 out of 1397



$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Structural Similar Compounds

Name	ANTHRAQUINONE; 1-((2-HYDROXYETHYL)AMINO)-4-(METHYLAMINO)-	p-Acetophenetidide; 3'-(bis(2-hydroxyethyl)amino)-	Disperse Black 9
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.537	0.654	0.657
Reference	28ZPAK 245;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;645;86	J. Am. Coll. Toxicol. 5(3):205;1986

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.16
Mahalanobis Distance: 11
Mahalanobis Distance p-value: 0.00487

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

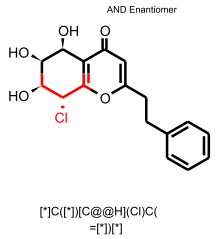
Model Applicability

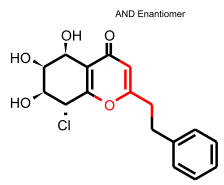
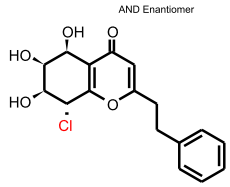
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

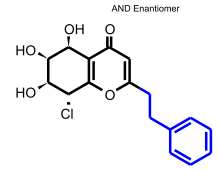
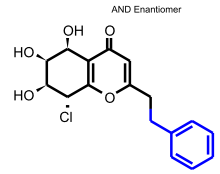
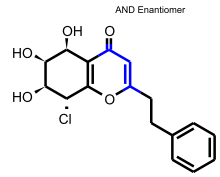
Feature Contribution

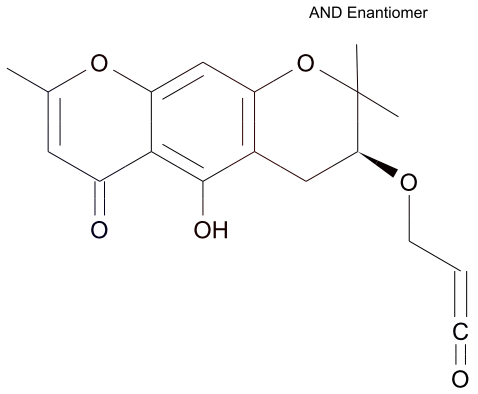
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-879019572	 [*]C([*])[C@@H](Cl)C(=[*])[*]	0.199	15 out of 15

FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.167	4 out of 4
FCFP_12	32	 [*]Cl	0.15	202 out of 215

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1981711554	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.103	5 out of 7
FCFP_12	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	451847724	 [*]C(=CC(=[*])[*])[*]	0	161 out of 192



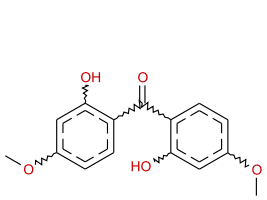
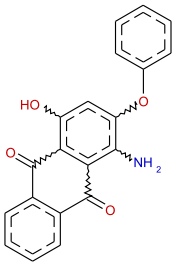
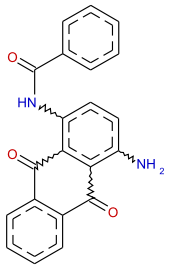
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 0.818
Mahalanobis Distance: 8.34
Mahalanobis Distance p-value: 0.829

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2;2';-Dihydroxy-4;4'-dimethoxybenzophenone	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE
Structure			
Actual Endpoint	Non-Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.568	0.595	0.599
Reference	J. Am. Coll. Toxicol. 2(5):35;1983	28ZPAK 239;72	28ZPAK-;124;72

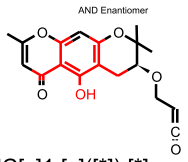
Model Applicability

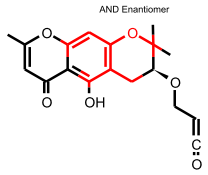
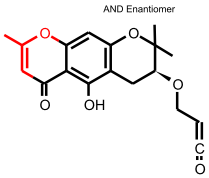
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 444624378: [*]C=C=O

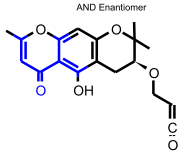
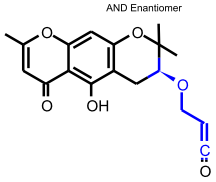
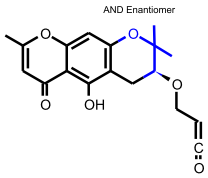
Feature Contribution

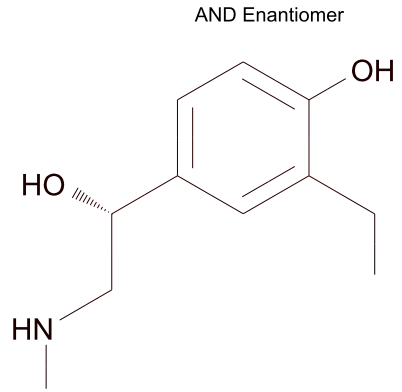
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1601875224	 [*]C[c]1:[c]([*]):[*]:[c]([*]):[c](C(=[*])[*]):[c]:1O	0.184	7 out of 7

FCFP_12	-1099193755	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	0.175	5 out of 5
FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.167	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1244036906	 [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*]	-0.592	0 out of 1
FCFP_12	353886735	 [*]C([*])OCC=C=[*]	-0.141	16 out of 23
FCFP_12	-415216134	 [*]OC(C)(C)C([*])[*]	-0.0889	11 out of 15



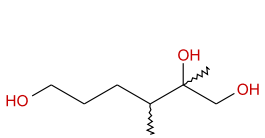
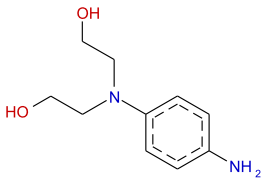
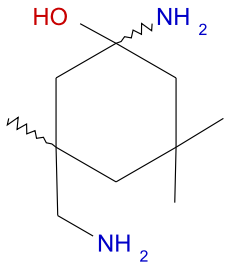
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 2.06
Mahalanobis Distance: 5.13
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2;3-DIMETHYL;1;2;6-HEXANETRIOL	p-Phenylenediamine; N;N-bis(2-hydroxyethyl)-; sulfate (1:1)	1-AMINO-3-AMINOMETHYL-3;5;5-TRIMETHYLCYCLOHEXANOL
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.610	0.639	0.641
Reference	AIHAAP 23;95;62	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;695;86	AIHAAP 30;470;69

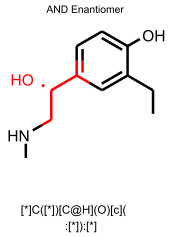
Model Applicability

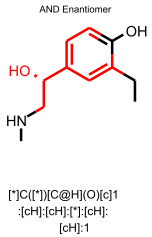
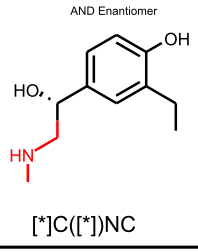
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

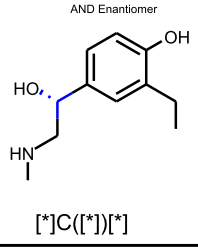
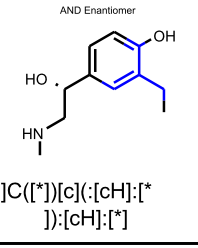
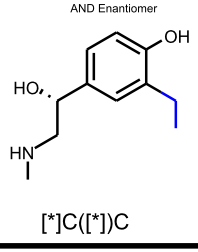
Feature Contribution

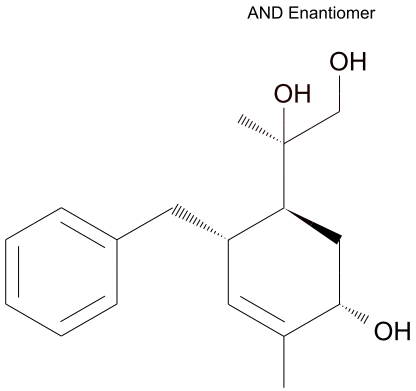
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1186393305	 [*]C([*])[C@H](O)[c](:[*]):[*]	0.197	13 out of 13

FCFP_12	-1931573337		0.195	12 out of 12
FCFP_12	-885550502		0.18	64 out of 66

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	0		0	1184 out of 1397
FCFP_12	203677720		0	319 out of 382
FCFP_12	136597326		0	612 out of 753



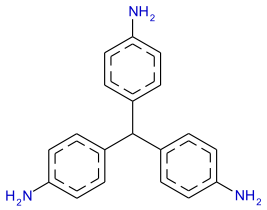
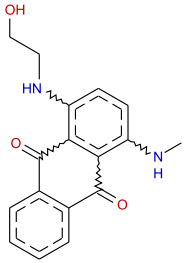
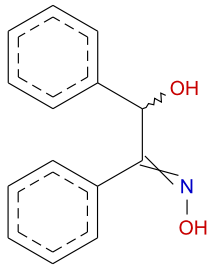
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 0.621
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 0.0104

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	METHANE;TRIS(4-AMINOPHENYL)-	ANTHRAQUINONE; 1-((2-HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	BENZOIN; OXIME
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.560	0.648	0.678
Reference	28ZPAK-;73;72	28ZPAK 245;72	28ZPAK-;111;72

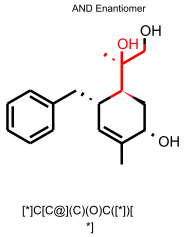
Model Applicability

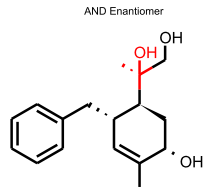
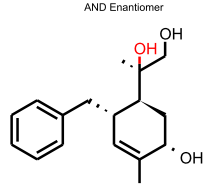
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

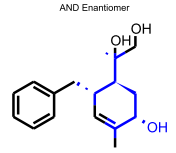
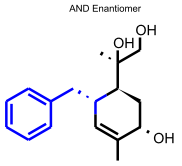
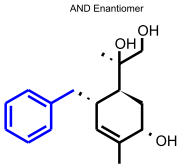
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-415156552	 [*]C[C@](C)(O)C([*])[*]	0.184	7 out of 7

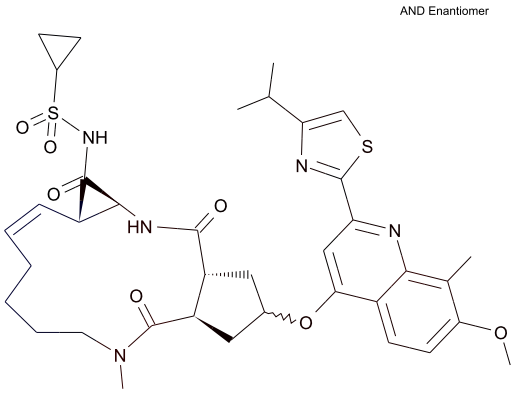
FCFP_12	1070061035	 [*]C([*])O	0.107	338 out of 376
FCFP_12	3	 [*]O	0.105	491 out of 547

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-59531427	 [*][C@H]1[*]=C([*])[C@@H](O)C[C@@H]1C([*])([*])[*]	-0.156	4 out of 6
FCFP_12	1981711554	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.103	5 out of 7
FCFP_12	-1698724694	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146

Simeprevir

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

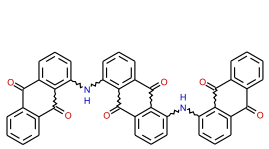
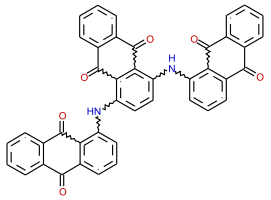
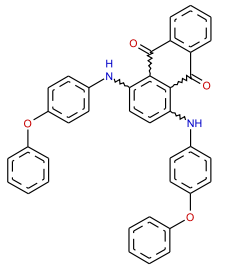
Bayesian Score: 1.81

Mahalanobis Distance: 16.5

Mahalanobis Distance p-value: 1.59e-019

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Anthraquinone; 1;1'-(anthraquinon-1;5-ylenediimino)di-	Anthraquinone; 1;1'-(anthraquinon-1;4-ylenediimino)di-	ANILINE;N;N'-1;4-ANTHRAQUINONYLENEBIS(4-PHENOXY-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.837	0.838	1.053
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	28ZPAK-;114;72

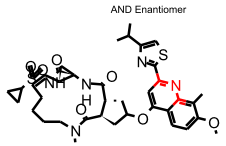
Model Applicability

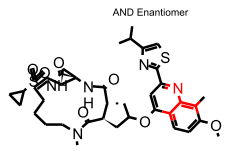
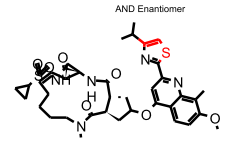
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

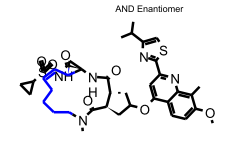
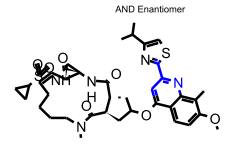
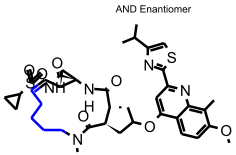
Feature Contribution

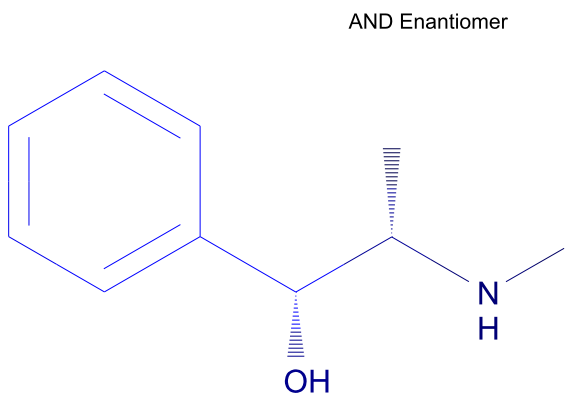
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*][c](:[*]):[c](:n:[*]):[c](:[*]):[*]	0.202	19 out of 19
FCFP_12	-124655670	 [*][c]1:[*]:[*]:s:[cH]:1	0.2	16 out of 16

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1812846456	 [*]CCCC\C=C/C1[*][*]1	-1.01	2 out of 9
FCFP_12	690511177	 [*]:[cH]:[c](:n:[*])[c](:[*]):[*]	-0.268	1 out of 2
FCFP_12	1175638033	 [*]CCCCC=[*]	-0.133	207 out of 293



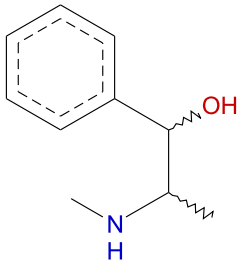
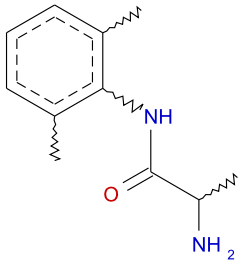
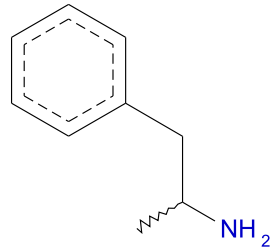
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.192
Enrichment: 0.595
Bayesian Score: -8.3
Mahalanobis Distance: 9.66
Mahalanobis Distance p-value: 0.53

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Tocainide	Amphetamine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.000	0.554	0.573
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

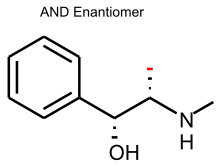
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

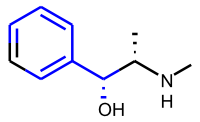
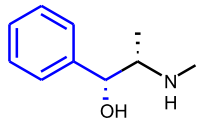
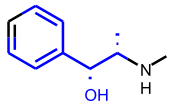
Feature Contribution

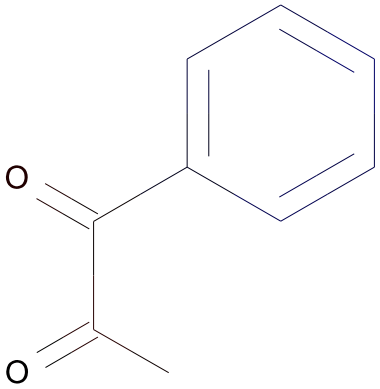
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	734603939	 [*]C	0.0966	92 out of 267

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-388186450	AND Enantiomer  [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-1.68	0 out of 14
ECFP_12	2014710090	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.56	0 out of 12
ECFP_12	-221133010	AND Enantiomer  [*]C([*])[C@H](O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.661	0 out of 3



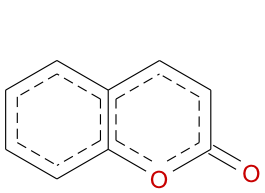
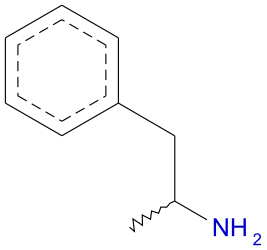
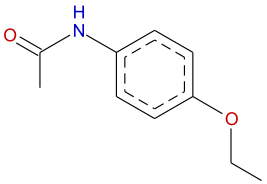
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Carcinogen
Probability: 0.298
Enrichment: 0.926
Bayesian Score: -0.0178
Mahalanobis Distance: 10.1
Mahalanobis Distance p-value: 0.321

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Coumarin	Amphetamine	Phenacetin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.570	0.579	0.590
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

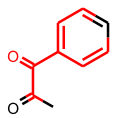
Model Applicability

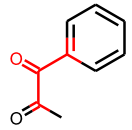
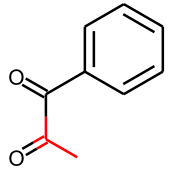
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -472613004: [*]C(=[*])C(=O)C

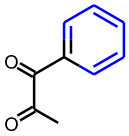
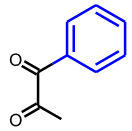
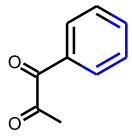
Feature Contribution

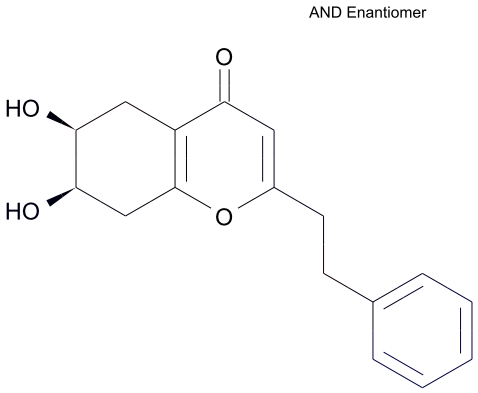
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	169261700	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.421	1 out of 1

ECFP_12	1432101658	 [*]C(=[*])C(=O)[c](:[*]):[*]	0.208	1 out of 2
ECFP_12	866218936	 [*]C(=[*])C	0.196	12 out of 31

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	-281505363	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	1997021792	 [*]:[cH]:[cH]:[cH]:[*]	-0.296	36 out of 156



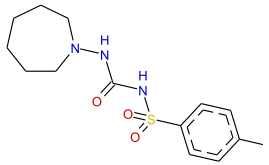
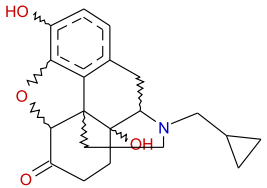
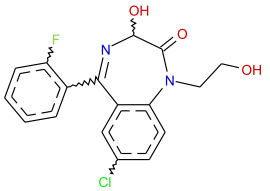
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.27
Enrichment: 0.838
Bayesian Score: -1.53
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 0.265

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Tolazamide	Naltrexone	Doxefazepam
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.566	0.611	0.616
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

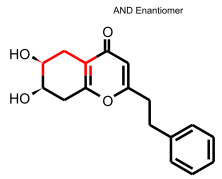
Model Applicability

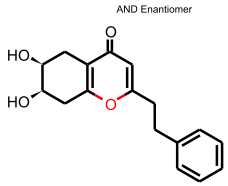
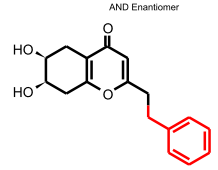
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1794461805: [*]CC(=C([*])[*])C(=[*])[*]
3. Unknown ECFP_2 feature: 1650944136: [*]CC(=C([*])[*])O[*]

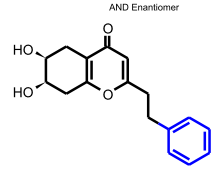
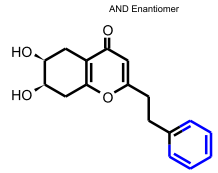
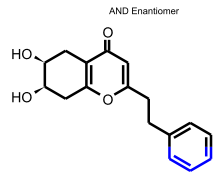
Feature Contribution

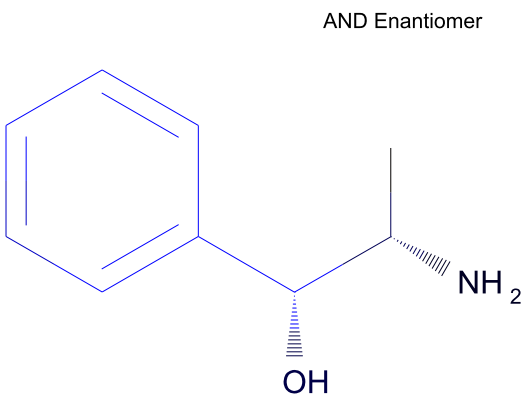
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	53207596	 [*]C([*])CC(=[*])[*]	0.459	8 out of 15

ECFP_12	683445015	 [*]O[*]	0.294	28 out of 66
ECFP_12	-1650219925	 [*]C[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.208	6 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-281505363	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	1997021792	 [*]:[cH]:[cH]:[cH]:[*]	-0.296	36 out of 156



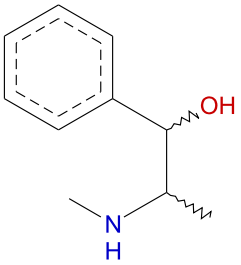
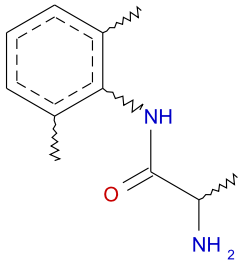
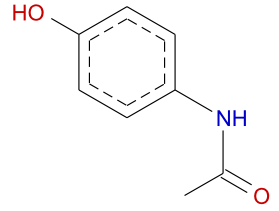
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.202
Enrichment: 0.627
Bayesian Score: -6.96
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 0.0982

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Tocainide	Acetaminophen
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.398	0.508	0.525
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

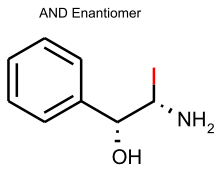
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

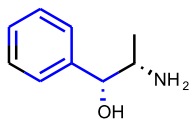
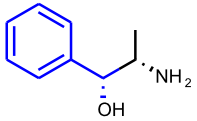
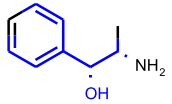
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -261056708: [*]C([*])[C@H](C)N

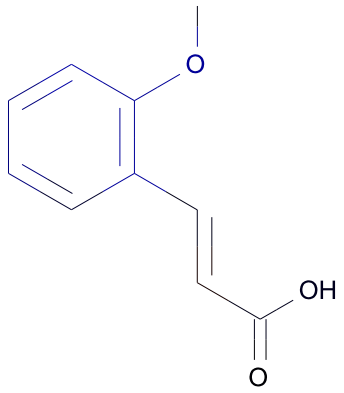
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	734603939	 [*]C	0.0966	92 out of 267

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-388186450	AND Enantiomer  [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-1.68	0 out of 14
ECFP_12	2014710090	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.56	0 out of 12
ECFP_12	-221133010	AND Enantiomer  [*]C([*])[C@H](O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.661	0 out of 3



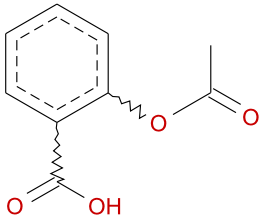
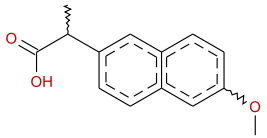
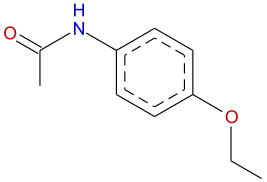
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.234
Enrichment: 0.728
Bayesian Score: -3.9
Mahalanobis Distance: 8.83
Mahalanobis Distance p-value: 0.871

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aspirin	Naproxen	Phenacetin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.532	0.564	0.565
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

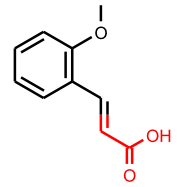
Model Applicability

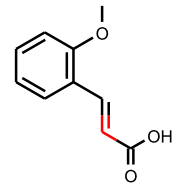
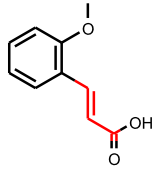
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1335702447: [*][c](:[*]):[c](C=[*]):c:[*]

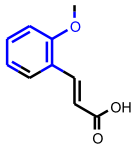
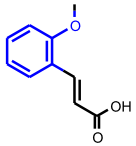
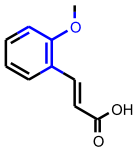
Feature Contribution

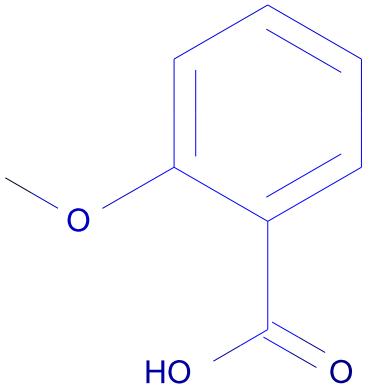
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1905093414	 [*]=CC(=O)O	0.437	2 out of 3

ECFP_12	-1925046727	 [*]C=[*]	0.407	16 out of 33
ECFP_12	-470416293	 [*]\C=C\C(=[*])[*]	0.158	2 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1634699529	 [*]O[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-1.25	0 out of 8
ECFP_12	-1961666573	 [*]O[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1[*]	-0.661	0 out of 3
ECFP_12	1408898974	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	-0.517	5 out of 29



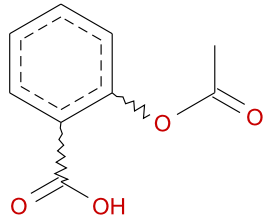
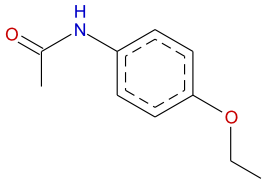
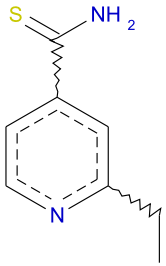
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.183
Enrichment: 0.568
Bayesian Score: -9.75
Mahalanobis Distance: 8.08
Mahalanobis Distance p-value: 0.984

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aspirin	Phenacetin	Ethionamide
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.351	0.601	0.606
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

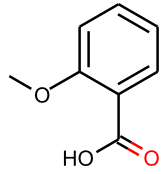
Model Applicability

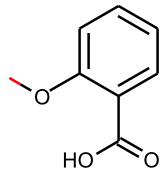
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

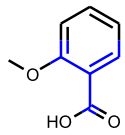
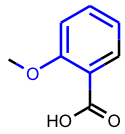
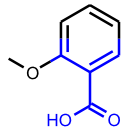
Feature Contribution

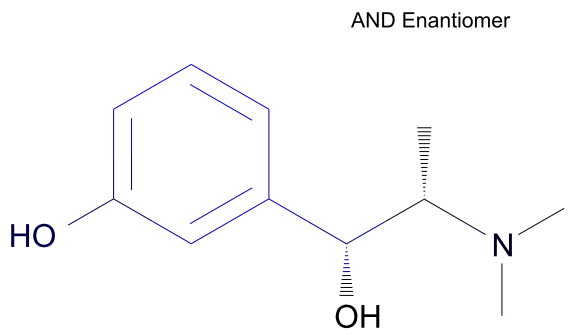
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1074141656	 [*]=O	0.103	86 out of 248

ECFP_12	734603939	 [*]C	0.0966	92 out of 267

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1337040050	 [*]C(=[*])[c](:[cH]:[*]):[c]([*]):[*]	-1.84	0 out of 17
ECFP_12	1634699529	 [*]O[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-1.25	0 out of 8
ECFP_12	989674687	 [*][c](:[*]):[c](:[cH]:[*])C(=O)O	-1.16	0 out of 7



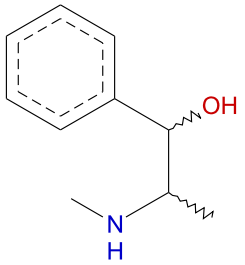
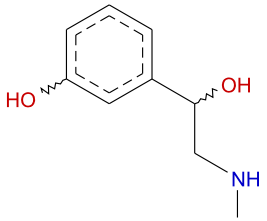
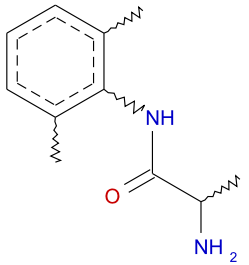
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.22
Enrichment: 0.682
Bayesian Score: -5.13
Mahalanobis Distance: 12.4
Mahalanobis Distance p-value: 0.00149

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Phenylephrine	Tocainide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.495	0.523	0.563
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

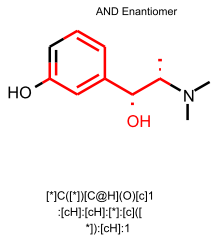
Model Applicability

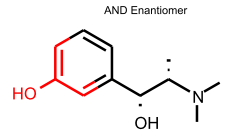
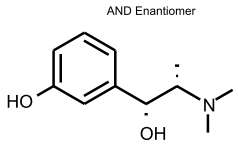
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1280034494: [*]C([*])[C@H](C)N([*])[*]

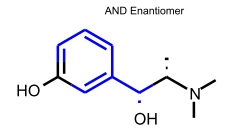
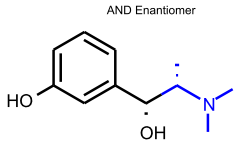
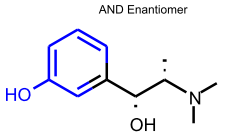
Feature Contribution

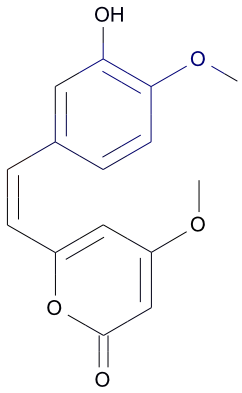
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1235570483	 [*]C([*])(C@H)(O)[c]1 :[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.421	1 out of 1

ECFP_12	-177786161	 [*]:[cH]:[c](O):[cH]:[*]	0.341	7 out of 15
ECFP_12	734603939	 [*]C	0.0966	92 out of 267

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-388186450	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-1.68	0 out of 14
ECFP_12	1064495017	 [*]C([*])N(C)C	-0.661	0 out of 3
ECFP_12	-783815036	 O[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.661	0 out of 3



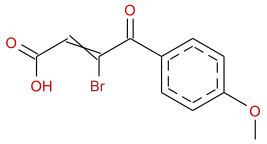
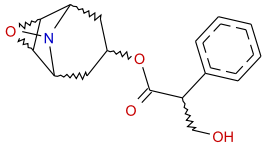
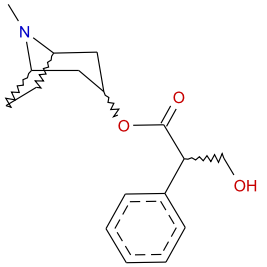
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.253
Enrichment: 0.787
Bayesian Score: -2.56
Mahalanobis Distance: 9.85
Mahalanobis Distance p-value: 0.44

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Cytembena	Scopolamine	Atropine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.554	0.605	0.610
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

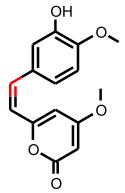
Model Applicability

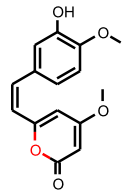
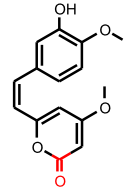
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC13 out of range. Value: -3.1324. Training min, max, SD, explained variance: -2.9952, 6.2862, 1.429, 0.0214.
2. Unknown ECFP_2 feature: -176483725: [*]=C[c](:c:[*]):c:[*]
3. Unknown ECFP_2 feature: -1053980253: [*]O\C(=C/[*])\C=[*]
4. Unknown ECFP_2 feature: -444332269: [*]O\C(=C\[*])\C=[*]

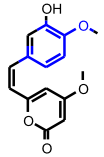
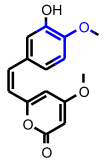
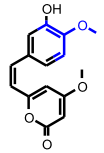
Feature Contribution

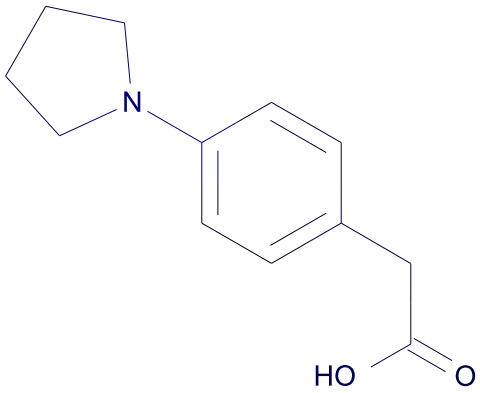
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1925046727	 [*]C=[*]	0.407	16 out of 33

ECFP_12	683445015	 [*]O[*]	0.294	28 out of 66
ECFP_12	2106656448	 [*]C(=O)[*]	0.141	30 out of 83

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	2077607946	 [*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-1.25	0 out of 8
ECFP_12	1408898974	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	-0.517	5 out of 29
ECFP_12	1680623188	 [*][c](:[*]):[c](OC): [cH]:[*]	-0.295	3 out of 14



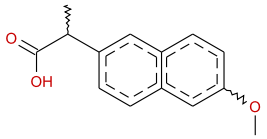
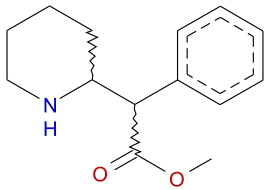
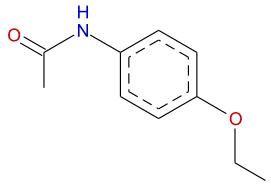
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.233
Enrichment: 0.724
Bayesian Score: -4.01
Mahalanobis Distance: 9.87
Mahalanobis Distance p-value: 0.431

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Naproxen	Methylphenidate	Phenacetin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.536	0.542	0.579
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

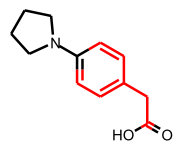
Model Applicability

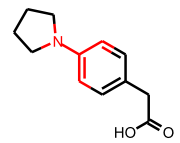
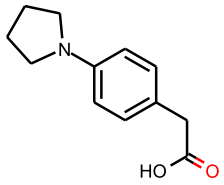
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

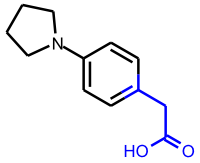
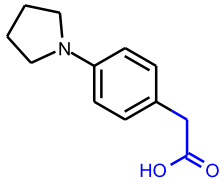
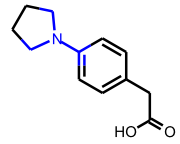
Feature Contribution

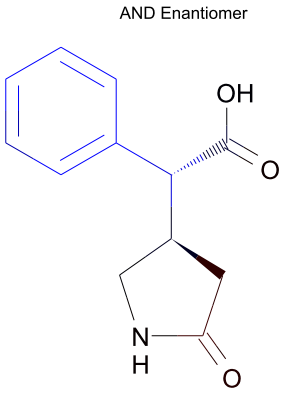
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1737023319	 [*]C(=[*])C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.613	2 out of 2

ECFP_12	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.158	2 out of 5
ECFP_12	-1074141656	 [*]=O	0.103	86 out of 248

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1251172932	 [*]:[c](:[*])CC(=O)O	-0.661	0 out of 3
ECFP_12	1731135544	 [*]CC(=O)O	-0.583	2 out of 14
ECFP_12	1951894094	 [*]:[c](:[*])N1C[*][*]C1	-0.505	3 out of 18



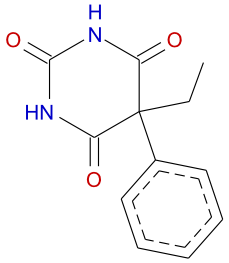
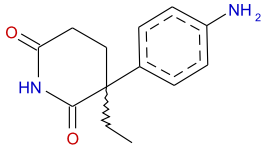
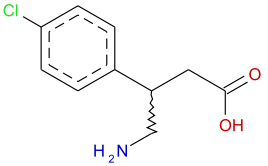
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.214
Enrichment: 0.665
Bayesian Score: -5.65
Mahalanobis Distance: 9.58
Mahalanobis Distance p-value: 0.568

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenobarbital	Aminoglutethimide	Baclofen
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.526	0.544	0.601
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

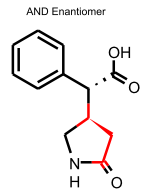
Model Applicability

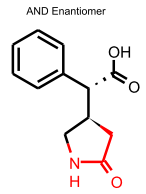
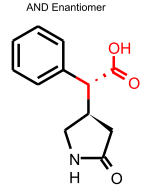
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -858846751: [*]C([*])C1C[*][*]C1

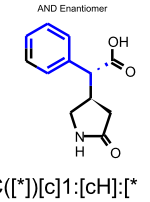
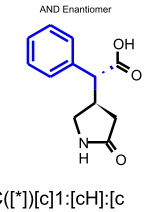
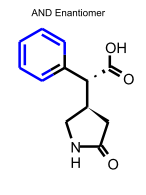
Feature Contribution

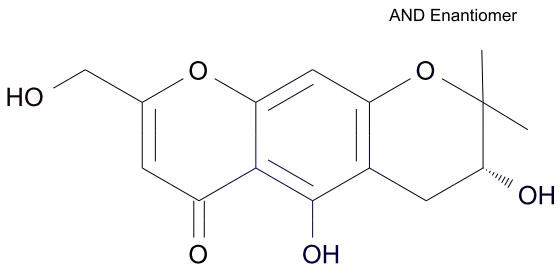
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	53207596	 [*]C([*])CC(=[*])[*]	0.459	8 out of 15

ECFP_12	-742538367	 O=C1C[*][*]N1	0.445	3 out of 5
ECFP_12	-1905455774	 [*]C([*])C(=O)O	0.357	4 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-388186450	 [*]C([*])[c]1:[cH]:[cH]:[*]:[cH]:[cH]:[cH]:1	-1.68	0 out of 14
ECFP_12	2014710090	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.56	0 out of 12
ECFP_12	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64



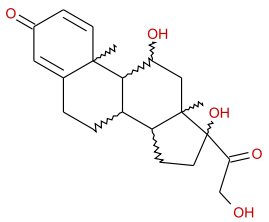
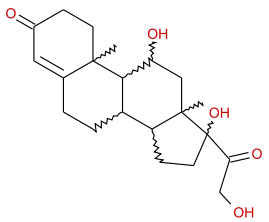
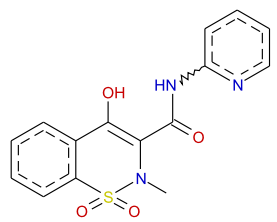
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.259
Enrichment: 0.803
Bayesian Score: -2.22
Mahalanobis Distance: 10.6
Mahalanobis Distance p-value: 0.162

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Prednisolone	Hydrocortisone	Piroxicam
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.645	0.651	0.674
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

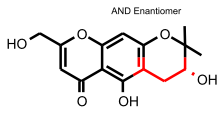
Model Applicability

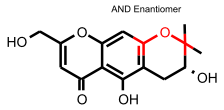
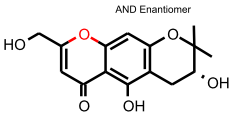
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

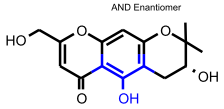
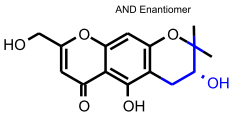
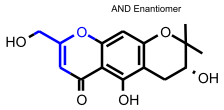
Feature Contribution

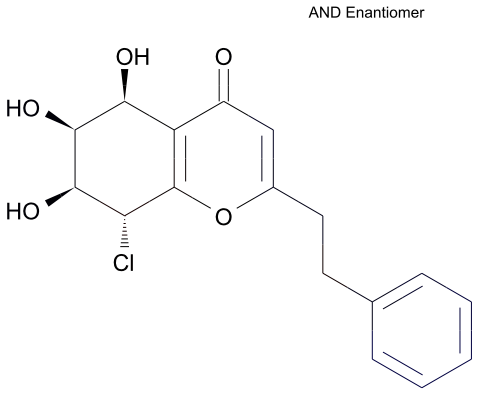
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	53207596	 [*]C([*])CC(=[*])[*]	0.459	8 out of 15

ECFP_12	-200406221	 [*]C([*])([*])O[c](:[*]):[*]	0.33	3 out of 6
ECFP_12	683445015	 [*]O[*]	0.294	28 out of 66

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1660913849	 [*][c](:[*]):[c](O):[c]([*]):[*]	-0.941	0 out of 5
ECFP_12	-1051556861	 [*]C[C@@H](O)C([*])([*])[*]	-0.661	0 out of 3
ECFP_12	1875238785	 [*]C\C(=C\[*])\O[*]	-0.272	0 out of 1



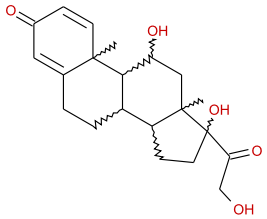
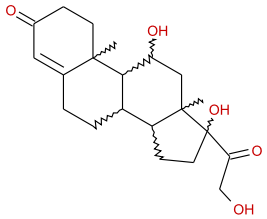
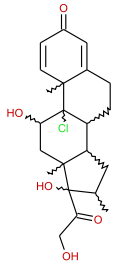
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.253
Enrichment: 0.787
Bayesian Score: -2.55
Mahalanobis Distance: 11.8
Mahalanobis Distance p-value: 0.00903

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Prednisolone	Hydrocortisone	Beclomethasone
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.551	0.555	0.633
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

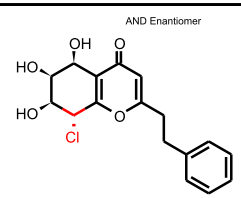
Model Applicability

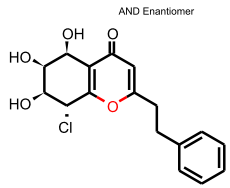
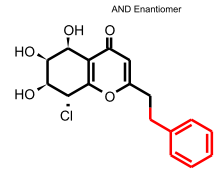
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1567907747: [*]C([*])[C@@H](O)C(=[*])[*]
3. Unknown ECFP_2 feature: -7106223: [*]C([*])[C@@H](Cl)C(=[*])[*]
4. Unknown ECFP_2 feature: 1652274794: [*]OC(=C([*])[*])C([*])[*]

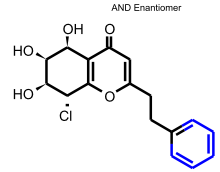
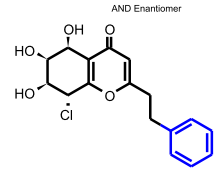
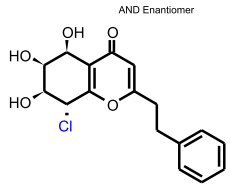
Feature Contribution

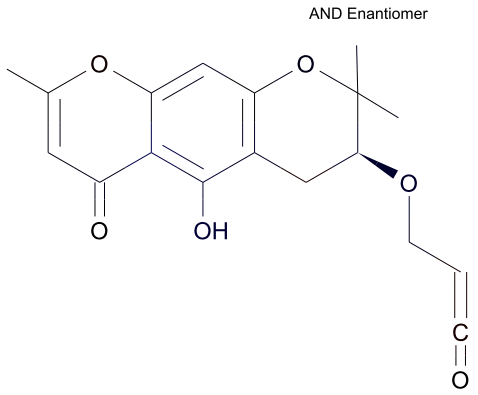
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	105634199	 [*]C([*])Cl	0.421	1 out of 1

ECFP_12	683445015	 [*]O[*]	0.294	28 out of 66
ECFP_12	-1650219925	 [*]C[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.208	6 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	-281505363	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	-817402818	 [*]Cl	-0.368	17 out of 80



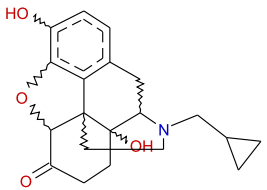
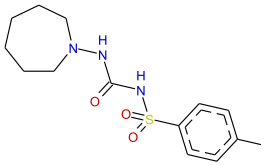
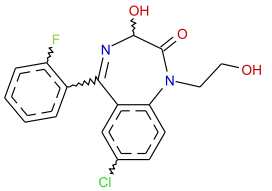
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.27
Enrichment: 0.837
Bayesian Score: -1.55
Mahalanobis Distance: 14
Mahalanobis Distance p-value: 2.45e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Naltrexone	Tolazamide	Doxefazepam
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.551	0.580	0.603
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

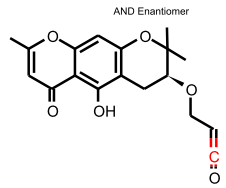
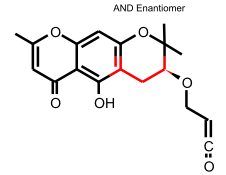
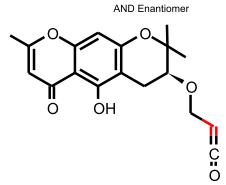
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -331149802: [*]O\C(=C/[*])\C
3. Unknown ECFP_2 feature: -1591590376: [*]C=C=O
4. Unknown ECFP_2 feature: 2106995136: [*]=C=O
5. Unknown ECFP_2 feature: -91536905: [*]CC=C=[*]
6. Unknown ECFP_2 feature: -1688150664: [*]OCC=[*]

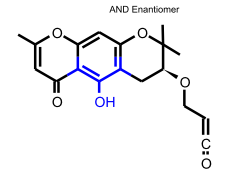
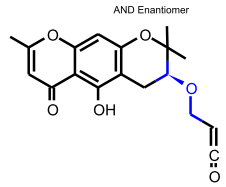
Feature Contribution

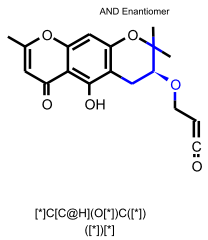
Top features for positive contribution

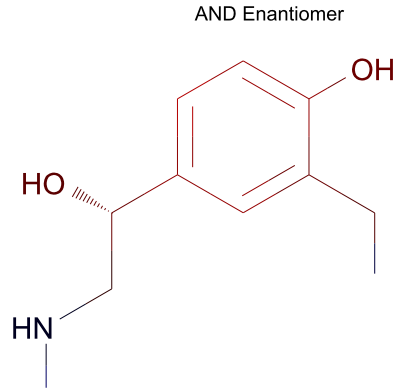
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-1114776580	 [*]=C=[*]	0.461	10 out of 19
ECFP_12	53207596	 [*]C([*])CC(=[*])[*]	0.459	8 out of 15
ECFP_12	-1925046727	 [*]C=[*]	0.407	16 out of 33

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1660913849	 [*][c](:[*]):[c](O):[c]([*]):[*]	-0.941	0 out of 5
ECFP_12	-1250019913	 [*]COC([*])[*]	-0.485	0 out of 2

ECFP_12	-2124995946	 [*]C[C@H](O[*])C([*])([*])[*]	-0.272	0 out of 1



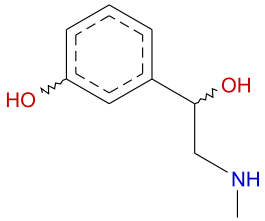
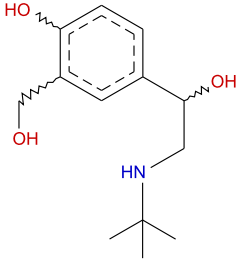
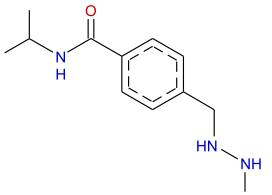
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Carcinogen
Probability: 0.272
Enrichment: 0.845
Bayesian Score: -1.41
Mahalanobis Distance: 9.23
Mahalanobis Distance p-value: 0.729

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenylephrine	Albuterol	Procarbazine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.433	0.502	0.565
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

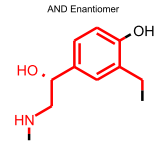
Model Applicability

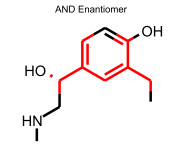
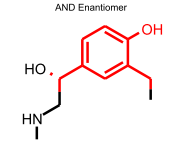
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

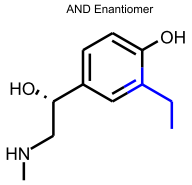
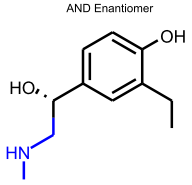
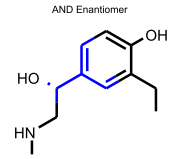
Feature Contribution

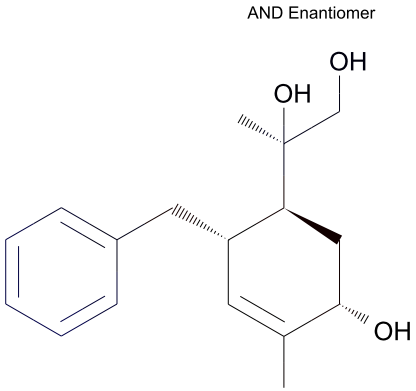
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1411846807	 [*]C[c]1:[cH]:[c](:[cH]:[cH]:[c]:1[*])[C@@H](O)CN[*]	0.613	2 out of 2

ECFP_12	-283284905	 [*]C[c]1:[cH]:[c](:[cH]:[*]:[c]:1[*])C([*])[*]	0.613	2 out of 2
ECFP_12	-188314265	 [*]C[c]1:[cH]:[c](:[cH]:[cH]:[c]:1O)C([*])[*]	0.613	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	767488533	 [*]:[c](:[*])CC	-0.941	0 out of 5
ECFP_12	493154328	 [*]CNC	-0.661	0 out of 3
ECFP_12	-176846085	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.374	7 out of 34



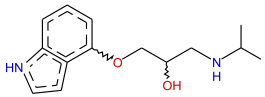
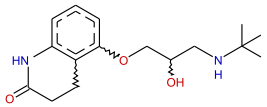
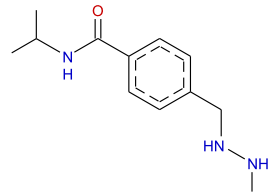
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.252
Enrichment: 0.781
Bayesian Score: -2.67
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 0.0121

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Pindolol	Carteolol	Procarbazine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.604	0.635	0.647
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

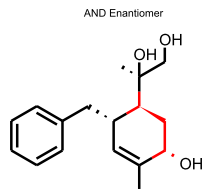
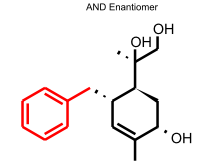
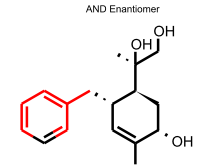
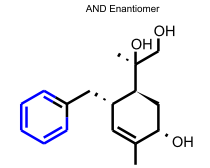
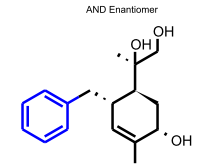
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

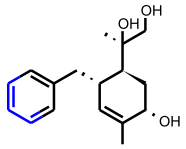
1. OPS PC20 out of range. Value: -3.2486. Training min, max, SD, explained variance: -3.0915, 4.6636, 1.259, 0.0166.
2. Unknown ECFP_2 feature: -2097867909: [*]C[C@H](O)C(=[*])[*]
3. Unknown ECFP_2 feature: -327548242: [*]C[C@@H](C([*])[*])C([*])([*])[*]
4. Unknown ECFP_2 feature: -1263967621: [*]C[C@@H](C=[*])C([*])[*]
5. Unknown ECFP_2 feature: -1042330089: [*]\C=C(\C)/C([*])[*]
6. Unknown ECFP_2 feature: 1280892564: [*]C[C@](C)(O)C([*])[*]
7. Unknown ECFP_2 feature: 771121623: [*]C([*])C[c](:[*]):[*]

Feature Contribution

Top features for positive contribution

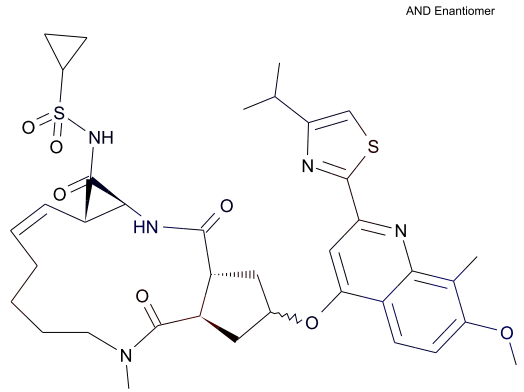
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-801490360	 [*]C([*])CC([*])[*]	0.339	14 out of 31
ECFP_12	-1650219925	 [*]C[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.208	6 out of 15
ECFP_12	1095683433	 [*]C[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0.154	6 out of 16
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_12	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	-281505363	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64

ECFP_12	1997021792	AND Enantiomer  [*]:[cH]:[cH]:[cH]:[*]	-0.296	36 out of 156

Simeprevir

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.232

Enrichment: 0.719

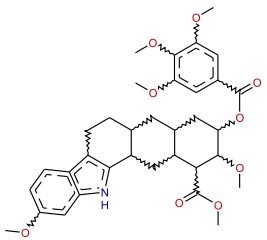
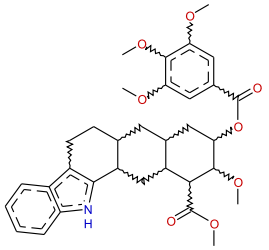
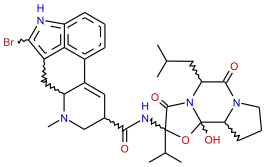
Bayesian Score: -4.13

Mahalanobis Distance: 14.6

Mahalanobis Distance p-value: 1.61e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Reserpine	Deserpidine	Bromocriptine
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.807	0.846	0.874
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

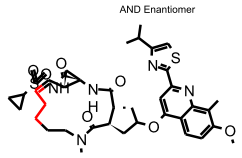
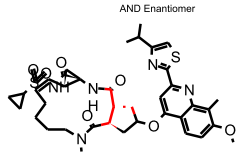
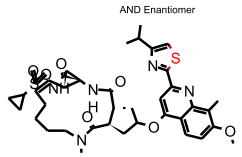
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -309676563: [*]N[C@]1(C[C@H]1[*])C(=[*])[*]
3. Unknown ECFP_2 feature: -1049290660: [*]C1([*])C[C@H]1C=[*]
4. Unknown ECFP_2 feature: 1616402542: [*]CN(C)C(=[*])[*]
5. Unknown ECFP_2 feature: -1818486371: [*]NC(=O)C1([*])[*][*]1
6. Unknown ECFP_2 feature: 1310940530: [*]C([*])O[c](:[*]):[*]
7. Unknown ECFP_2 feature: 733491677: [*]:[c](:[*])C(C)C
8. Unknown ECFP_2 feature: 2121941848: [*]NS(=O)(=O)C1[*][*]1
9. Unknown ECFP_2 feature: -622223421: [*]S(=[*])(=[*])C1CC1

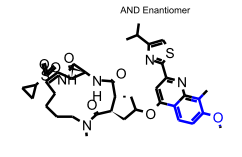
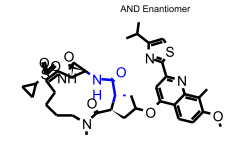
Feature Contribution

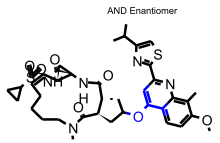
Top features for positive contribution

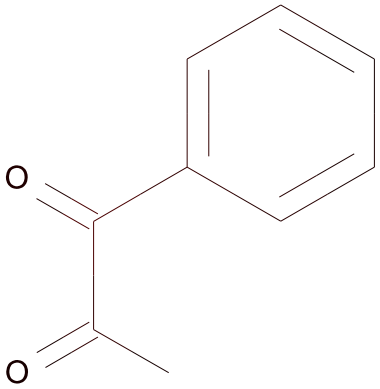
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-1331088410	 [*]CCC=[*]	0.725	3 out of 3
ECFP_12	-2095963820	 [*][C@@H]1[*][*]C[C@H]1C(=[*])[*]	0.722	12 out of 17
ECFP_12	914325265	 [*]:s:[*]	0.516	8 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	2077607946	 [*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-1.25	0 out of 8
ECFP_12	-867777309	 [*]NC(=O)C([*])[*]	-0.661	0 out of 3

ECFP_12	1408898974	AND Enantiomer  [*]O[c](:[cH]:[*]):[c](:[*]):[*]	-0.517	5 out of 29



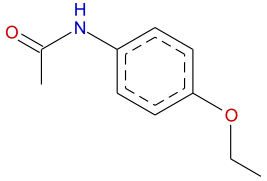
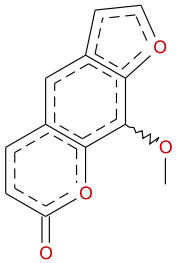
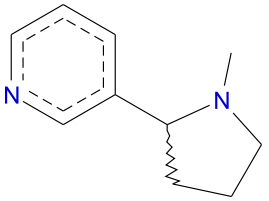
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.556
Enrichment: 1.49
Bayesian Score: 2.28
Mahalanobis Distance: 12.3
Mahalanobis Distance p-value: 0.000946

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenacetin	Methoxsalen; 8-	Nicotine
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.589	0.596	0.621
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

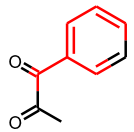
Model Applicability

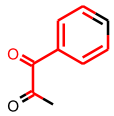
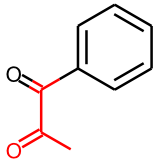
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

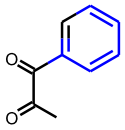
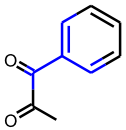
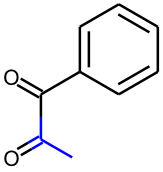
Feature Contribution

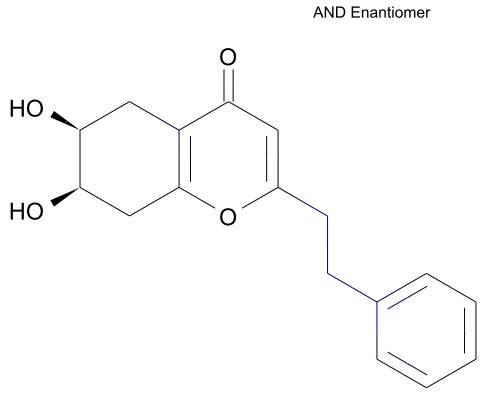
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	387787917	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0.449	6 out of 11

SCFP_4	-541146916	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.419	1 out of 1
SCFP_4	571795252	 [*]C(=[*])C(=O)C	0.283	2 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1379591900	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.0895	10 out of 35
SCFP_4	-2056718782	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0	6 out of 19
SCFP_4	136627117	 [*]C(=[*])C	0	6 out of 18



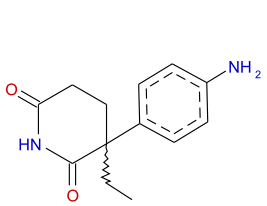
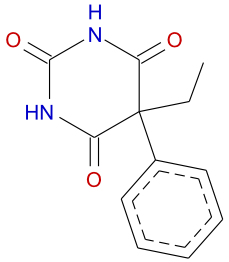
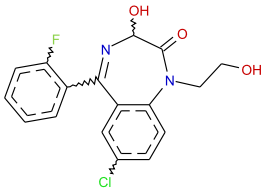
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.448
Enrichment: 1.2
Bayesian Score: -2.03
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 0.000121

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aminoglutethimide	Phenobarbital	Doxefazepam
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.574	0.580	0.582
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

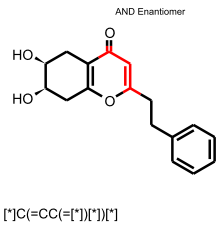
Model Applicability

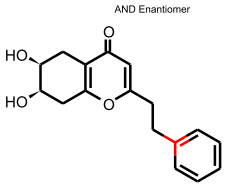
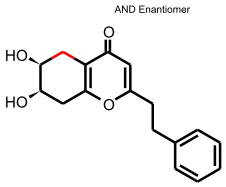
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC15 out of range. Value: -3.2503. Training min, max, SD, explained variance: -2.9572, 2.6953, 1.089, 0.0176.

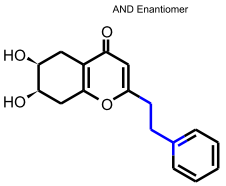
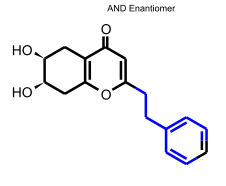
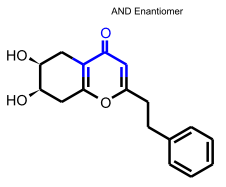
Feature Contribution

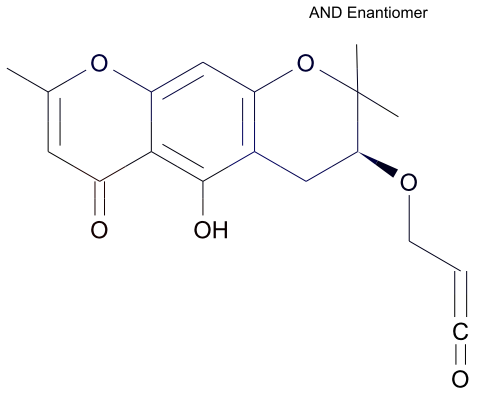
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1971196727	 [*]C(=CC(=[*])[*])[*]	0.295	5 out of 11

SCFP_4	3	 [*]:[cH]:[*]	0.199	36 out of 93
SCFP_4	0	 [*]C	0.155	41 out of 111

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1272709286	 [*]CC[c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	-1211866396	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.06	0 out of 6
SCFP_4	-1980361709	 [*]C(=[*])C(=O)C=[*]	-0.489	0 out of 2



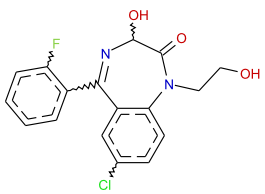
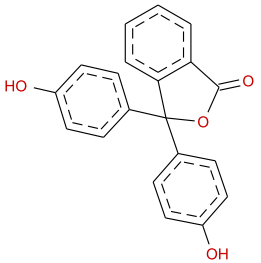
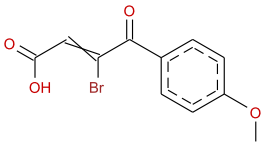
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.465
Enrichment: 1.24
Bayesian Score: -1.56
Mahalanobis Distance: 17.7
Mahalanobis Distance p-value: 8.28e-009

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Doxefazepam	Phenolphthalein	Cytembena
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.582	0.628	0.630
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

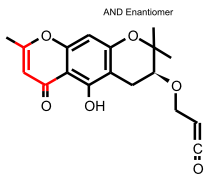
Model Applicability

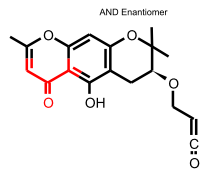
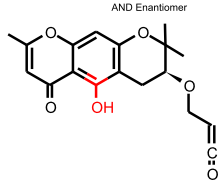
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

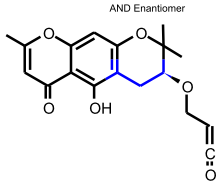
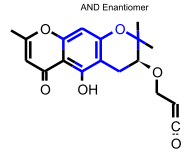
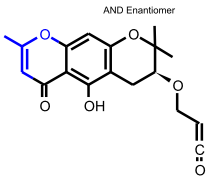
Feature Contribution

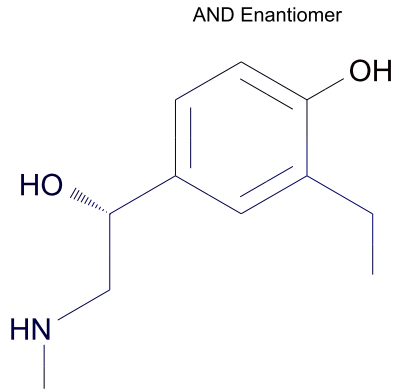
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1971196727	 [*]C(=CC(=[*])[*])[*]	0.295	5 out of 11

SCFP_4	-1980302127	 [*]C(=[*])C(=O)[c](:[*]):[*]	0.204	1 out of 2
SCFP_4	-424425761	 [*]:[c](:[*])O	0.201	6 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1272709286	 [*]CC[c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	1238198777	 [*][c]1:[*]:[c]([*]):[c]2C[*]C([*])([*])O[c]:2:[cH]:1	-0.489	0 out of 2
SCFP_4	616547045	 [*]C\C(=C\[*])\O[*]	-0.489	0 out of 2



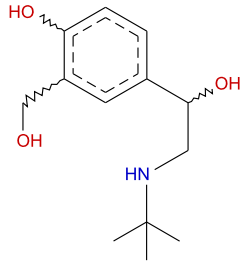
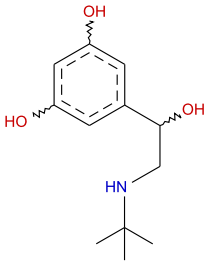
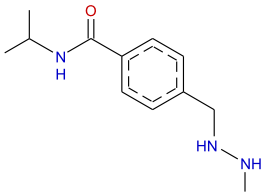
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.377
Enrichment: 1.01
Bayesian Score: -3.64
Mahalanobis Distance: 9.35
Mahalanobis Distance p-value: 0.153

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Albuterol	Terbutaline	Procarbazine
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.413	0.462	0.518
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

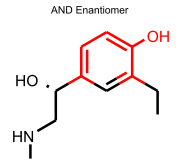
Model Applicability

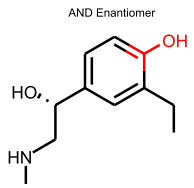
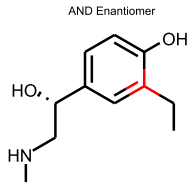
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

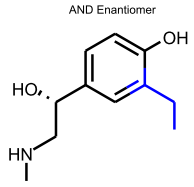
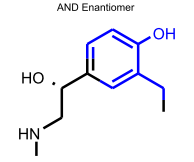
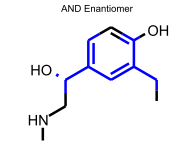
Feature Contribution

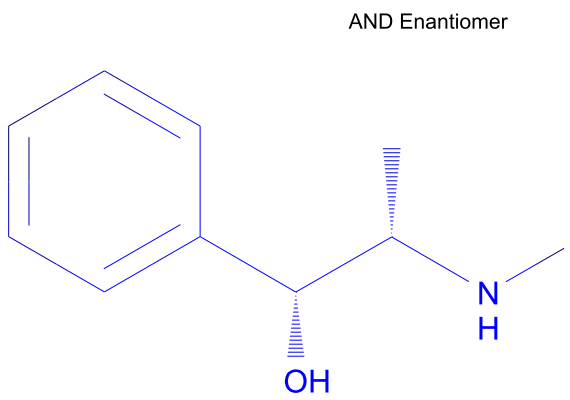
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1374800107	 [*][c]1:[*]:[c]([*]):[c](O):[cH]:[cH]:1	0.288	10 out of 23

SCFP_4	-424425761	 [*]:[c](:[*])O	0.201	6 out of 15
SCFP_4	3	 [*]:[cH]:[*]	0.199	36 out of 93

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1272709286	 [*]CC[c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	-1931277081	 [*]C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1O	-0.489	0 out of 2
SCFP_4	-350503170	 [*]C[c]1:[cH]:[c](:[cH]:[*]:[c]:1[*])C([*])[*]	-0.489	0 out of 2



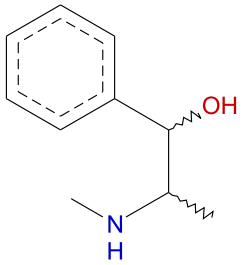
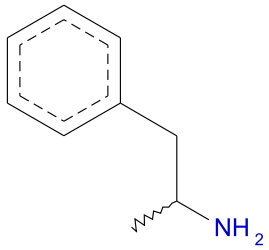
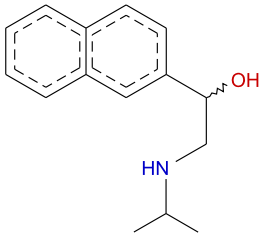
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.182
Enrichment: 0.544
Bayesian Score: -8.68
Mahalanobis Distance: 8.88
Mahalanobis Distance p-value: 0.966

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Amphetamine	Pronetalol
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.000	0.515	0.522
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

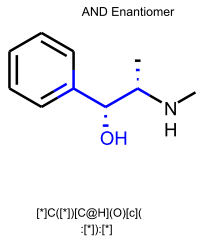
Model Applicability

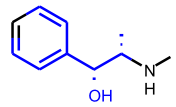
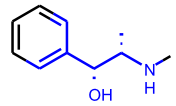
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

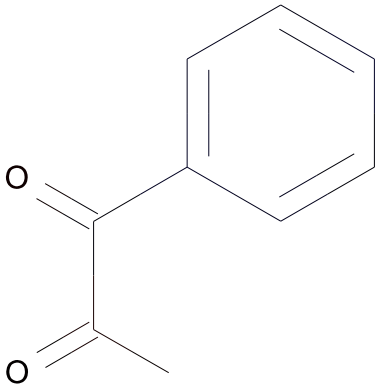
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1702724181	 [*]C([*])[C@H](O)[c](:[*]):[*]	-1.22	1 out of 18

SCFP_6	-561151481	AND Enantiomer  [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	-1.07	1 out of 15
SCFP_6	1318513260	AND Enantiomer  [*]N[C@@H](C)[C@H](O) [c](:[cH]:[*]):[cH]: [*]	-0.957	0 out of 5



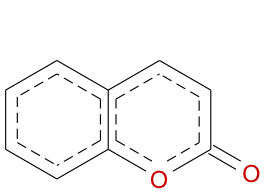
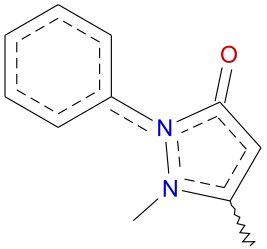
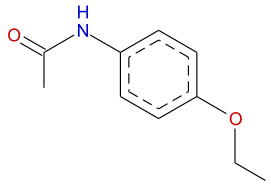
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Carcinogen
Probability: 0.379
Enrichment: 1.13
Bayesian Score: 0.756
Mahalanobis Distance: 9.11
Mahalanobis Distance p-value: 0.936

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Coumarin	Antipyrine	Phenacetin
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.525	0.547	0.583
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

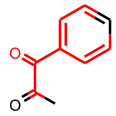
Model Applicability

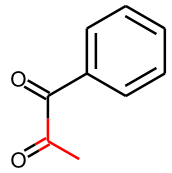
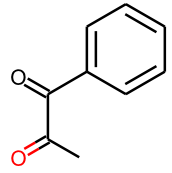
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

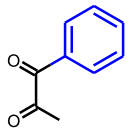
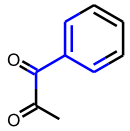
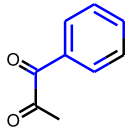
Feature Contribution

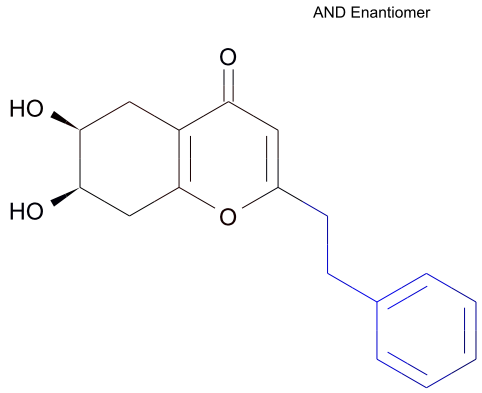
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-541146916	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.415	1 out of 1

SCFP_6	136627117	 [*]C(=[*])C	0.167	18 out of 47
SCFP_6	13	 [*]=O	0.0717	90 out of 261

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.504	12 out of 64
SCFP_6	-2056718782	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.219	23 out of 90
SCFP_6	387787917	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.164	14 out of 52



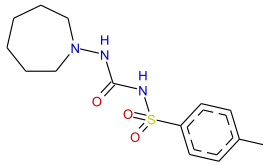
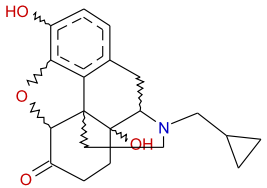
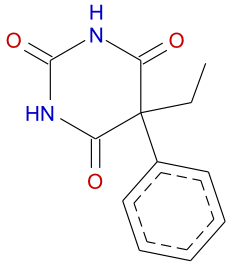
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.275
Enrichment: 0.822
Bayesian Score: -3.21
Mahalanobis Distance: 12.2
Mahalanobis Distance p-value: 0.0174

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Tolazamide	Naltrexone	Phenobarbital
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.536	0.585	0.598
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

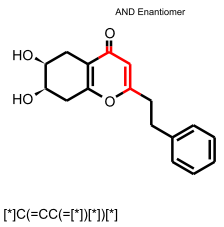
Model Applicability

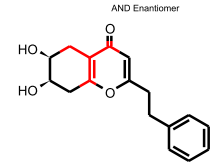
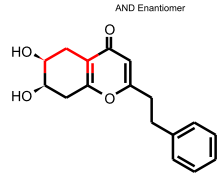
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC29 out of range. Value: -4.2515. Training min, max, SD, explained variance: -3.1746, 3.7825, 1.007, 0.0095.

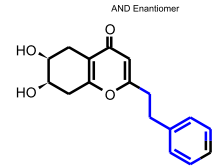
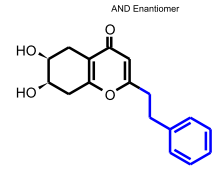
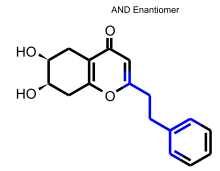
Feature Contribution

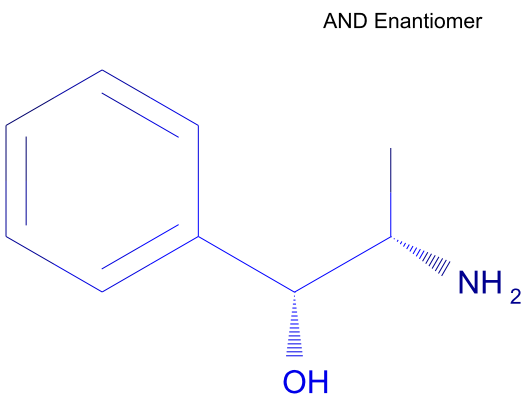
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971196727	 [*]C(=CC(=[*])[*])[*]	0.361	17 out of 36

SCFP_6	55464376	 [*]CC(=C([*])[*])C(=[*])[*]	0.345	14 out of 30
SCFP_6	-1272768868	 [*]C([*])CC(=[*])[*]	0.242	26 out of 63

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1211866396	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.1	2 out of 25
SCFP_6	-1640858361	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.817	1 out of 11
SCFP_6	-1852892018	 [*]C(=[*])CC[c](:[cH]:[*]):[cH]:[*]	-0.674	0 out of 3



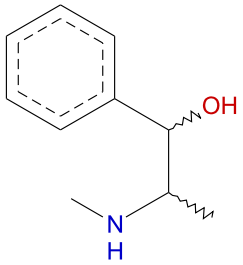
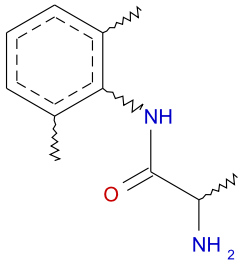
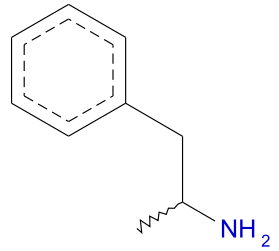
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.214
Enrichment: 0.641
Bayesian Score: -6.4
Mahalanobis Distance: 8.47
Mahalanobis Distance p-value: 0.991

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Tocainide	Amphetamine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.314	0.489	0.514
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

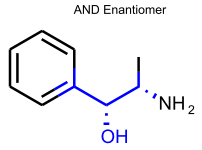
Model Applicability

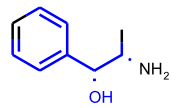
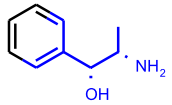
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

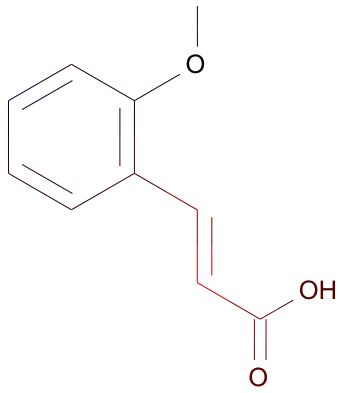
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1702724181	 [*]C([*])[C@H](O)[c](:[*]):[*]	-1.22	1 out of 18

SCFP_6	-561151481	AND Enantiomer  [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	-1.07	1 out of 15
SCFP_6	1318513260	AND Enantiomer  [*]N[C@@H](C)[C@H](O) [c](:[cH]:[*]):[cH]: [*]	-0.957	0 out of 5



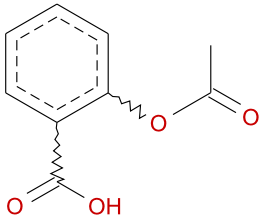
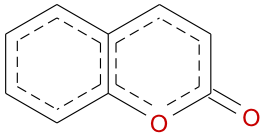
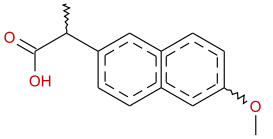
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.426
Enrichment: 1.27
Bayesian Score: 2.23
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 0.281

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aspirin	Coumarin	Naproxen
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.507	0.511	0.528
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

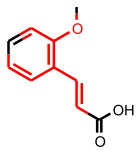
Model Applicability

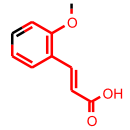
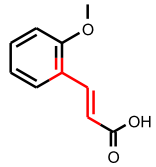
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

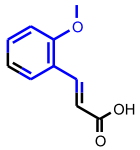
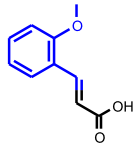
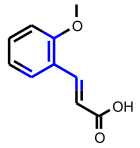
Feature Contribution

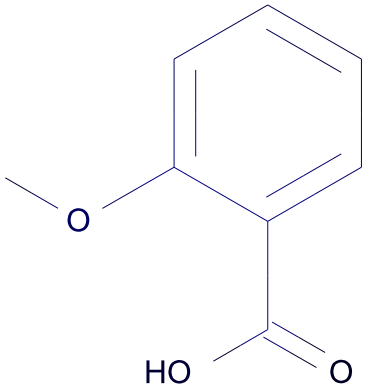
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	966282057	 [*]O[c]1:[cH]:[*]:[cH]:[cH]:[c]:1\C=C\[*]	0.603	2 out of 2

SCFP_6	2109620068	 [*]O[c]1:[cH]:[*]:[cH]:[cH]:[c]:1\C=C\C(=O)O	0.603	2 out of 2
SCFP_6	-1971137145	 [*]\C=C\[c](:[*]):[*]	0.434	5 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	503541685	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	-0.484	1 out of 7
SCFP_6	1434056623	 [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	-0.278	0 out of 1
SCFP_6	-2056718782	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.219	23 out of 90



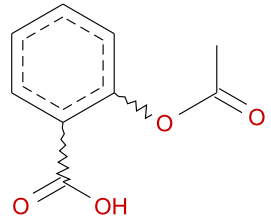
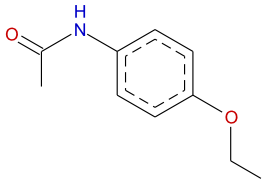
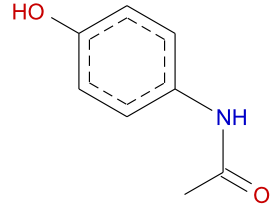
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.296
Enrichment: 0.886
Bayesian Score: -2.28
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.442

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Aspirin	Phenacetin	Acetaminophen
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.369	0.559	0.580
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

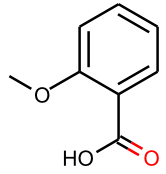
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

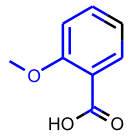
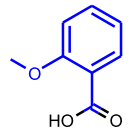
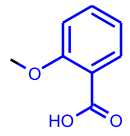
Feature Contribution

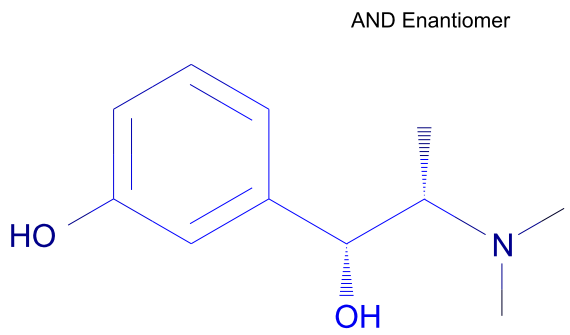
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	13	 [*]=O	0.0717	90 out of 261

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

SCFP_6	503541685	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	-0.484	1 out of 7
SCFP_6	1434056623	 [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	-0.278	0 out of 1
SCFP_6	1726855784	 [*]O[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C(=O)O	-0.278	0 out of 1



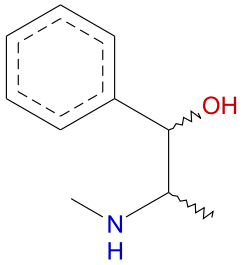
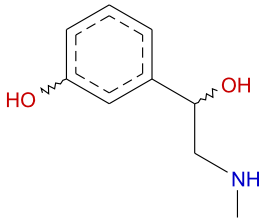
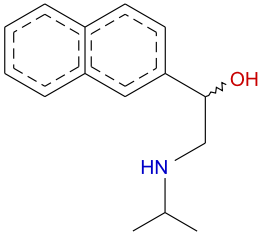
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.18
Enrichment: 0.54
Bayesian Score: -8.79
Mahalanobis Distance: 9.77
Mahalanobis Distance p-value: 0.75

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ephedrine	Phenylephrine	Pronetalol
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.399	0.431	0.543
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

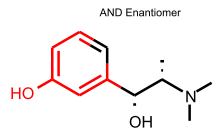
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

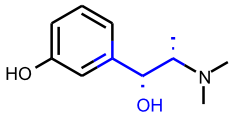
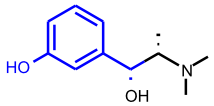
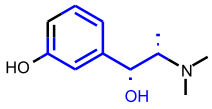
Feature Contribution

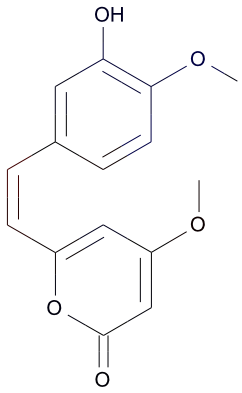
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	611156666	 [*][c]1:[*]:[cH]:[cH] :[c](O):[cH]:1	0.186	6 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

SCFP_6	1702724181	AND Enantiomer  [*]C([*])[C@H](O)[c](:[*]):[*]	-1.22	1 out of 18
SCFP_6	-1632615624	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[c](O):[cH]:1	-1.07	0 out of 6
SCFP_6	-561151481	AND Enantiomer  [*]C([*])[C@H](O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.07	1 out of 15



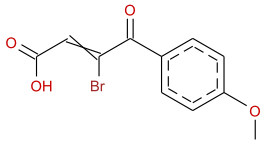
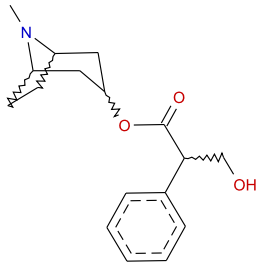
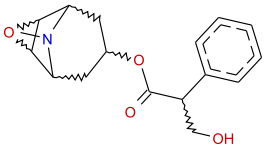
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.35
Enrichment: 1.05
Bayesian Score: -0.232
Mahalanobis Distance: 13
Mahalanobis Distance p-value: 0.00175

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Cytembena	Atropine	Scopolamine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.478	0.588	0.588
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

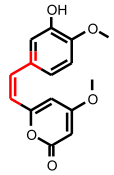
Model Applicability

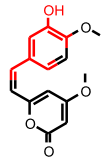
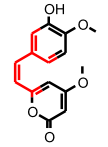
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC10 out of range. Value: -4.2526. Training min, max, SD, explained variance: -4.2502, 5.657, 1.784, 0.0297.

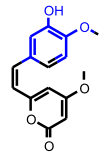
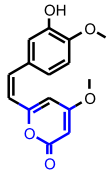
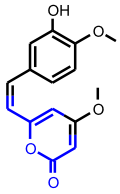
Feature Contribution

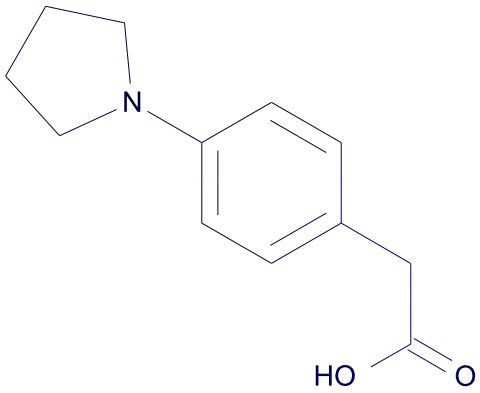
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971137145	 [*]\C=C\[c](:[*]):[*]	0.434	5 out of 9

SCFP_6	392579710	 [*][c]1:[*]:[cH]:[c](C=[*]):[cH]:[c]:1O	0.425	2 out of 3
SCFP_6	-1977229858	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	0.425	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	2116304939	 [*]O[c]1:[cH]:[*]:[c]([*]):[cH]:[c]:1O	-0.825	0 out of 4
SCFP_6	1559374977	 [*]C1=[*]C(=CC(=O)O1)[*]	-0.278	0 out of 1
SCFP_6	1616083408	 [*]=CC1=C[*]=CC(=O)O1	-0.278	0 out of 1



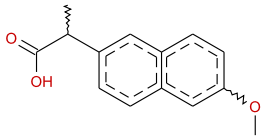
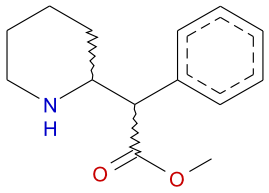
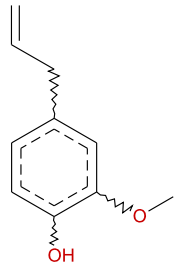
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.266
Enrichment: 0.797
Bayesian Score: -3.6
Mahalanobis Distance: 8.82
Mahalanobis Distance p-value: 0.971

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Naproxen	Methylphenidate	Eugenol
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.490	0.500	0.543
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

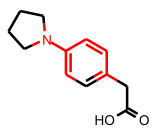
Model Applicability

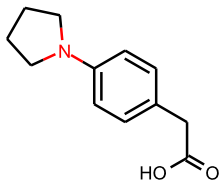
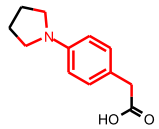
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

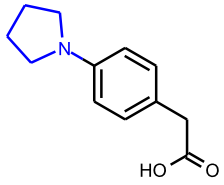
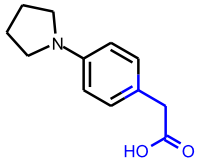
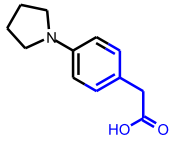
Feature Contribution

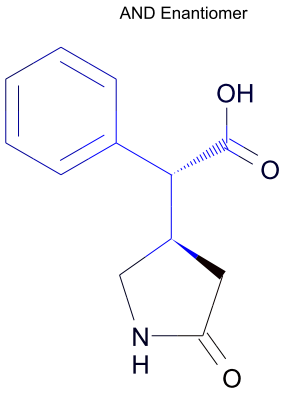
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.287	17 out of 39

SCFP_6	10	 [*]N([*])[*]	0.21	39 out of 98
SCFP_6	-1009128698	 [*]C[c]1:[cH]:[cH]:[c](:[cH]:[cH]:1)N2C[*][*]C2	0.198	1 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-2056996303	 [*]N1CCCC1	-0.825	0 out of 4
SCFP_6	-1576616475	 [*]:[c](:[*])CC(=O)O	-0.674	0 out of 3
SCFP_6	-1476721585	 [*]:[cH]:[c](CC(=O)O) :[cH]:[*]	-0.674	0 out of 3



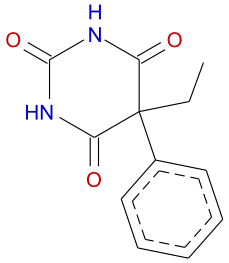
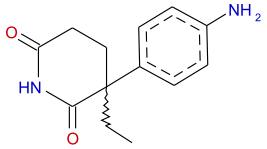
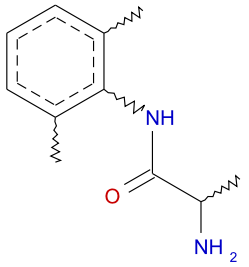
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.217
Enrichment: 0.65
Bayesian Score: -6.22
Mahalanobis Distance: 13
Mahalanobis Distance p-value: 0.00168

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenobarbital	Aminoglutethimide	Tocainide
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.470	0.502	0.559
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

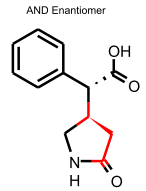
Model Applicability

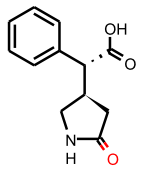
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

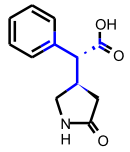
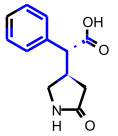
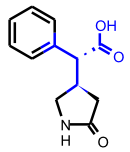
Feature Contribution

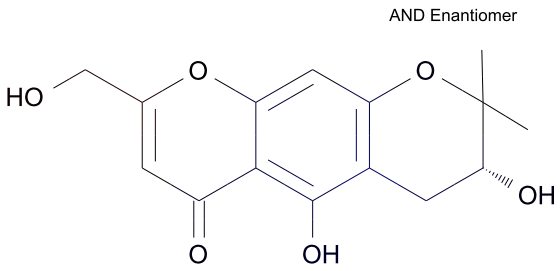
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1272768868	 [*]C([*])CC(=[*])[*]	0.242	26 out of 63

SCFP_6	13	AND Enantiomer  [*]=O	0.0717	90 out of 261

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	699559848	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]	-1.44	0 out of 10
SCFP_6	-1214451192	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.27	0 out of 8
SCFP_6	413428563	AND Enantiomer  [*]C([*])[C@@H](C(=O)O)[c](:[*]):[*]	-0.957	0 out of 5



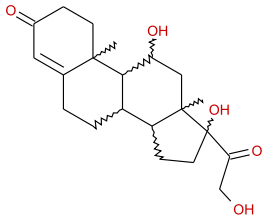
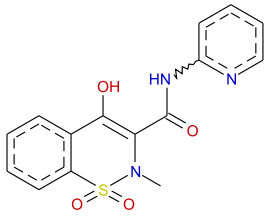
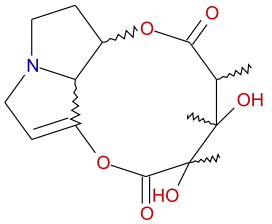
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.285
Enrichment: 0.853
Bayesian Score: -2.75
Mahalanobis Distance: 12.8
Mahalanobis Distance p-value: 0.00267

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Hydrocortisone	Piroxicam	Monocrotaline
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.645	0.652	0.659
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

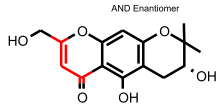
Model Applicability

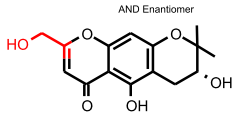
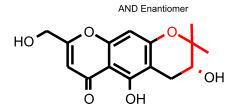
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

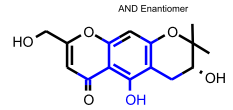
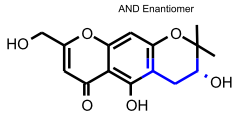
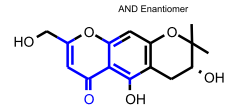
Feature Contribution

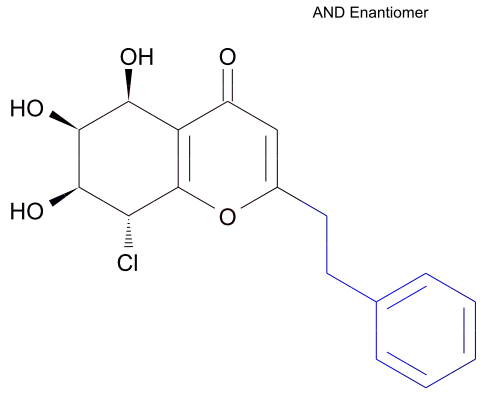
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971196727	 [*]C(=CC(=[*])[*])[*]	0.361	17 out of 36

SCFP_6	-711656408	 [*]C(=[*])CO	0.266	7 out of 16
SCFP_6	1416196903	 [*]OC(C)(C)C([*])[*]	0.213	5 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1681833008	 [*]C[c]1:[c]([*]):[*] :[c]([*]):[c](C(=[*])[*]):[c]:1O	-0.496	0 out of 2
SCFP_6	-1272709286	 [*]CC[c](:[*]):[*]	-0.459	12 out of 61
SCFP_6	-617610981	 [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*]	-0.278	0 out of 1



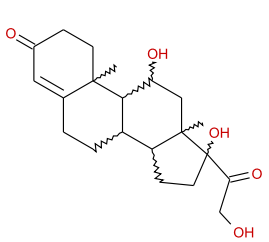
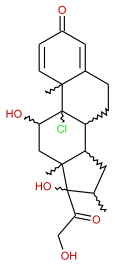
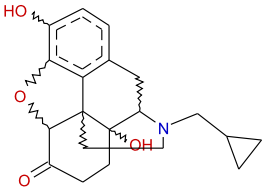
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.258
Enrichment: 0.772
Bayesian Score: -4
Mahalanobis Distance: 13.5
Mahalanobis Distance p-value: 0.000308

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Hydrocortisone	Beclomethasone	Naltrexone
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.538	0.605	0.606
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

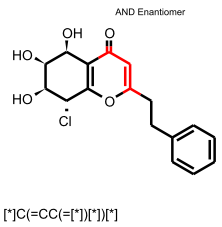
Model Applicability

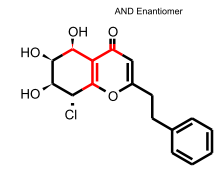
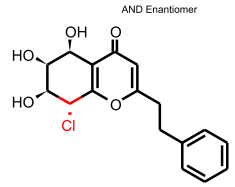
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC29 out of range. Value: -4.1695. Training min, max, SD, explained variance: -3.1746, 3.7825, 1.007, 0.0095.

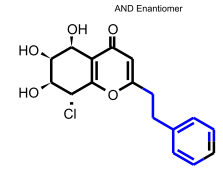
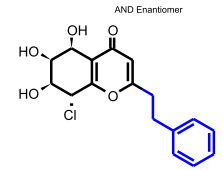
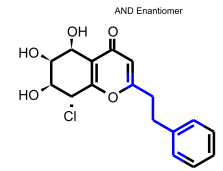
Feature Contribution

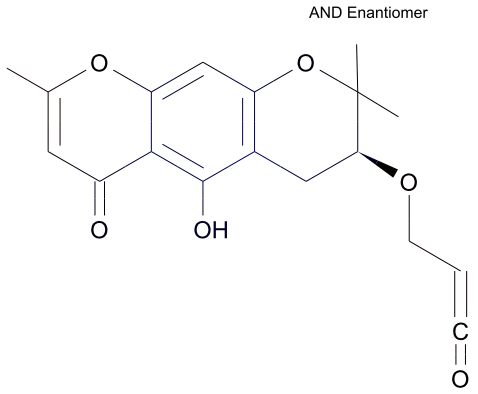
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971196727	 [*]C(=CC(=[*])[*])[*]	0.361	17 out of 36

SCFP_6	55464376	 [*]CC(=C([*])[*])C(=[*])[*]	0.345	14 out of 30
SCFP_6	-52163885	 [*]C([*])Cl	0.32	7 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1211866396	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.1	2 out of 25
SCFP_6	-1640858361	 [*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.817	1 out of 11
SCFP_6	-1852892018	 [*]C(=[*])CC[c](:[cH]:[*]):[cH]:[*]	-0.674	0 out of 3



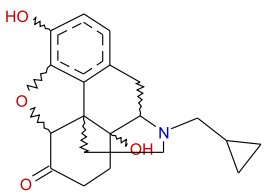
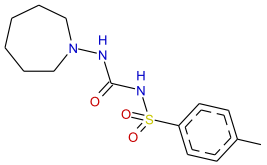
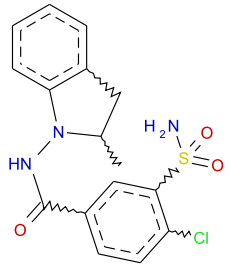
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.307
Enrichment: 0.919
Bayesian Score: -1.83
Mahalanobis Distance: 16.5
Mahalanobis Distance p-value: 4.92e-010

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Naltrexone	Tolazamide	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.526	0.562	0.585
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

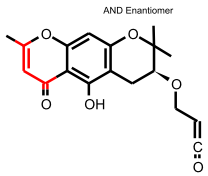
Model Applicability

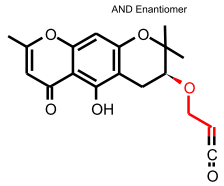
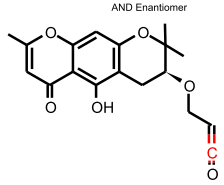
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

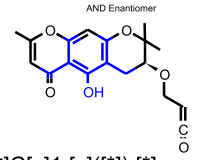
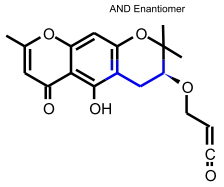
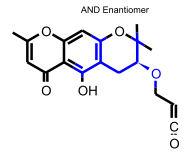
Feature Contribution

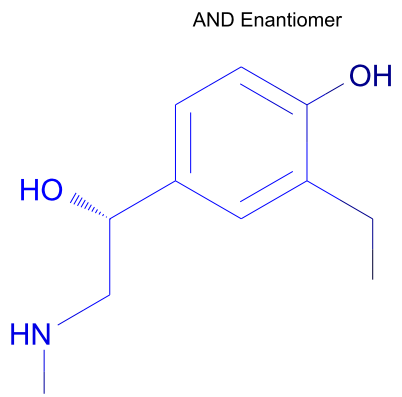
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971196727	 [*]C(=CC(=[*])[*])[*]	0.361	17 out of 36

SCFP_6	-711656408	 [*]C(=[*])CO	0.266	7 out of 16
SCFP_6	2	 [*]=C=[*]	0.238	8 out of 19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1681833008	 [*]C[c]1:[c]([*]):[*] :[c]([*]):[c](C(=[*])[*]):[c]:1O	-0.496	0 out of 2
SCFP_6	-1272709286	 [*]CC[c](:[*]):[*]	-0.459	12 out of 61
SCFP_6	-1849894309	 [*][c](:[*]):[c]1C[C@ @H](O)C([*])([*])[*] [c]:1:[*]	-0.278	0 out of 1



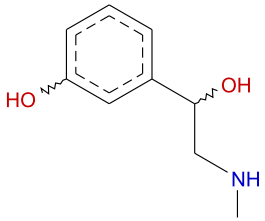
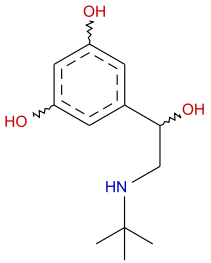
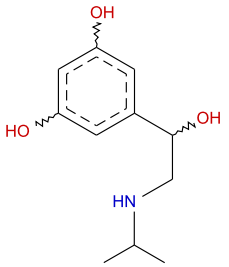
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.145
Enrichment: 0.434
Bayesian Score: -11.5
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 0.515

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenylephrine	Terbutaline	Metaproterenol
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.377	0.495	0.510
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

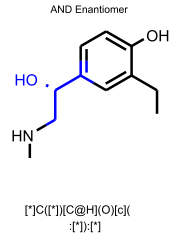
Model Applicability

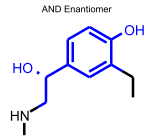
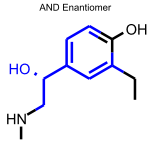
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

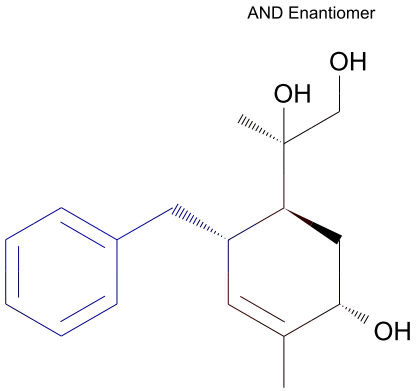
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1702724181	 [*]C([*])[C@H](O)[c](:[*]):[*]	-1.22	1 out of 18

SCFP_6	1578545183	 [*]C[C@H](O)[c]1:[cH] :[cH]:[c](O):[c]([*]):[cH]:1	-1.07	0 out of 6
SCFP_6	-561151481	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]:[cH]: [cH]:1	-1.07	1 out of 15



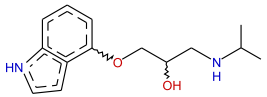
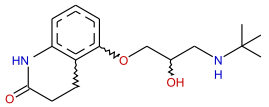
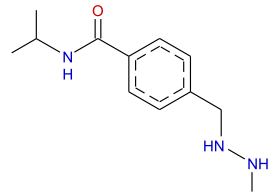
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.291
Enrichment: 0.872
Bayesian Score: -2.48
Mahalanobis Distance: 15.3
Mahalanobis Distance p-value: 1.51e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Pindolol	Carteolol	Procarbazine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.585	0.611	0.627
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

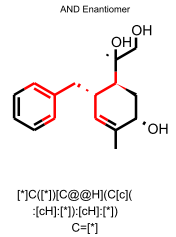
Model Applicability

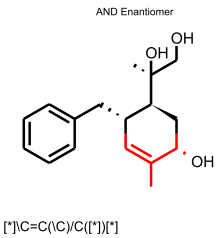
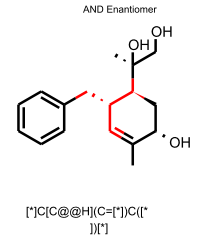
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC29 out of range. Value: -3.8031. Training min, max, SD, explained variance: -3.1746, 3.7825, 1.007, 0.0095.

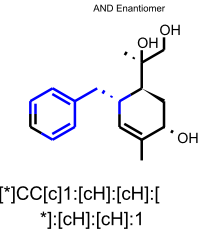
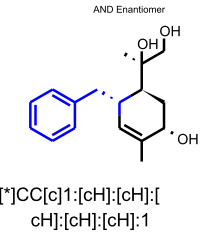
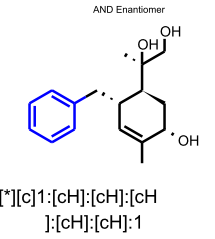
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1848330701	 [*]C([*])[C@@H](C[c](:[cH]:[*]):[cH]:[*])C=[*]	0.415	1 out of 1

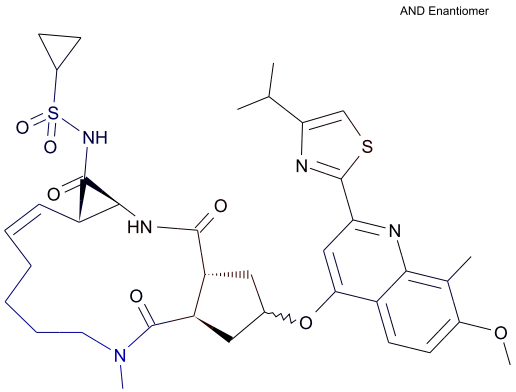
SCFP_6	55434585		0.331	12 out of 26
SCFP_6	-1043310069		0.296	15 out of 34

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1211866396		-1.1	2 out of 25
SCFP_6	-1640858361		-0.817	1 out of 11
SCFP_6	1653911926		-0.504	12 out of 64

Simeprevir

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.231

Enrichment: 0.69

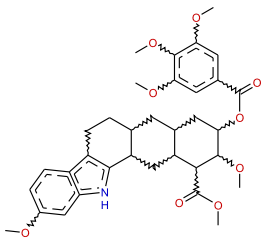
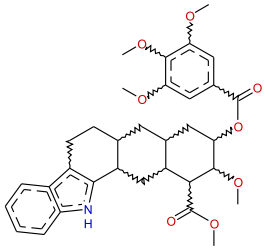
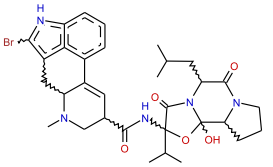
Bayesian Score: -5.43

Mahalanobis Distance: 20.3

Mahalanobis Distance p-value: 1.2e-019

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Reserpine	Deserpidine	Bromocriptine
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.771	0.812	0.842
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

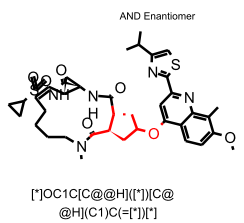
Model Applicability

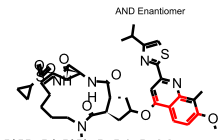
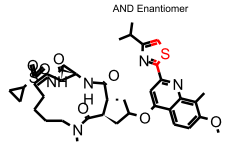
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 7.7766. Training min, max, SD, explained variance: -4.0702, 6.5927, 2.146, 0.0429.
2. OPS PC22 out of range. Value: 3.5288. Training min, max, SD, explained variance: -3.2994, 3.2573, 1.197, 0.0134.

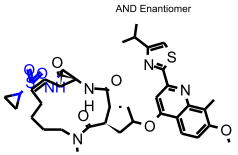
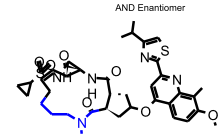
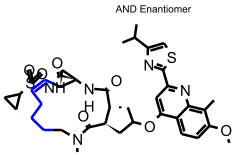
Feature Contribution

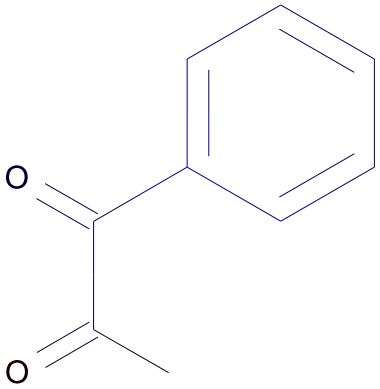
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1903488337	 [*]OC1C[C@@H]([*])[C@@H](C1)C(=[*])[*]	0.603	2 out of 2

SCFP_6	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1:[*]	0.526	11 out of 19
SCFP_6	1310748454	 [*][c]1:[*]:[*]:[cH]:s:1	0.437	7 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	182902497	 [*]NS(=O)(=O)C1[*][*]1	-0.825	0 out of 4
SCFP_6	306578635	 [*]CCCN(C)C(=[*])[*]	-0.825	0 out of 4
SCFP_6	1260369147	 [*]CCC\C=C/[*]	-0.674	0 out of 3



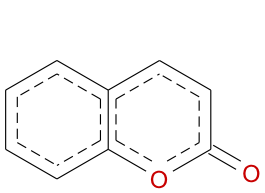
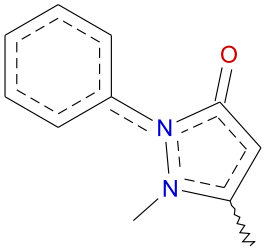
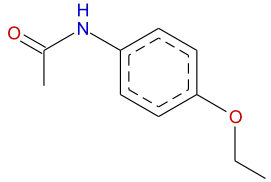
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.537
Enrichment: 1.3
Bayesian Score: -1.89
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 0.00179

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Coumarin	Antipyrine	Phenacetin
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.534	0.548	0.631
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

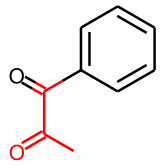
Model Applicability

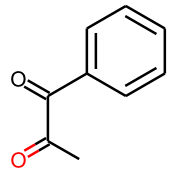
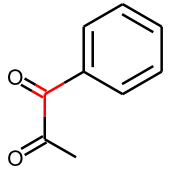
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

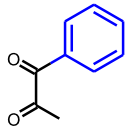
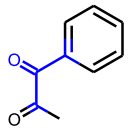
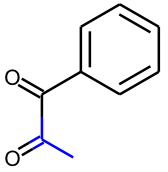
Feature Contribution

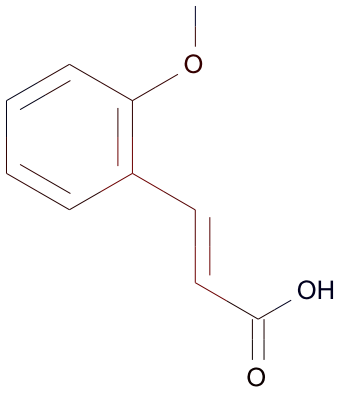
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	571795252	 [*]C(=[*])C(=O)C	0.201	2 out of 4

SCFP_8	13	 [*]=O	0.172	39 out of 90
SCFP_8	1	 [*]C(=[*])[*]	0.146	38 out of 90

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.985	1 out of 12
SCFP_8	-1980302127	 [*]C(=[*])C(=O)[c](:[*]):[*]	-0.737	0 out of 3
SCFP_8	136627117	 [*]C(=[*])C	-0.41	4 out of 18



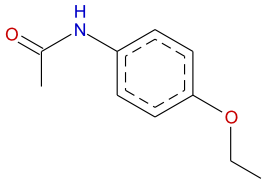
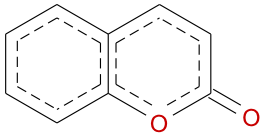
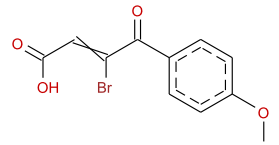
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.567
Enrichment: 1.37
Bayesian Score: 0.517
Mahalanobis Distance: 15.9
Mahalanobis Distance p-value: 1.61e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenacetin	Coumarin	Cytembena
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.542	0.553	0.564
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

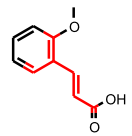
Model Applicability

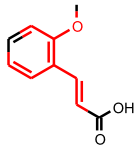
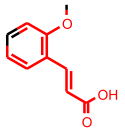
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

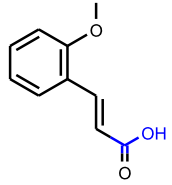
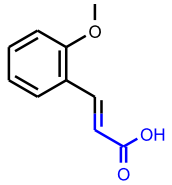
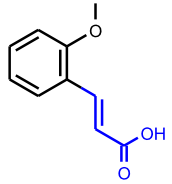
Feature Contribution

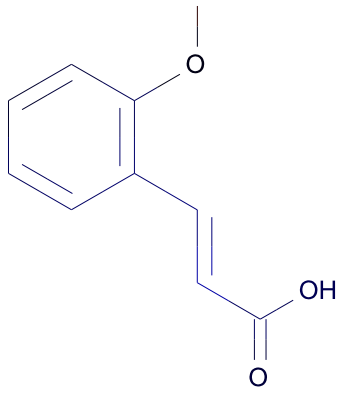
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-1977229858	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c](/[*]):[*]	0.553	2 out of 2

SCFP_8	966282057	 [*]O[c]1:[cH]:[*]:[cH]:[cH]:[c]:1\C=C\[*]	0.553	2 out of 2
SCFP_8	2109620068	 [*]O[c]1:[cH]:[*]:[cH]:[cH]:[c]:1\C=C\C(=O)O	0.553	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-424485343	 [*]C(=[*])O	-0.584	3 out of 17
SCFP_8	1132907712	 [*]=CC(=O)O	-0.58	2 out of 12
SCFP_8	2034654200	 [*]\C=C\C(=O)O	-0.546	0 out of 2



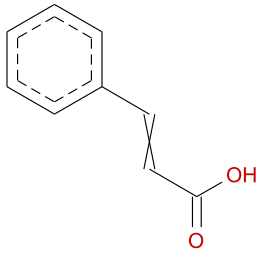
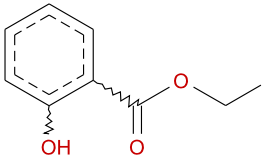
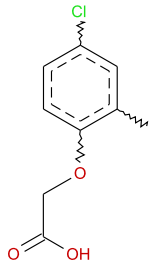
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.255
Enrichment: 0.693
Bayesian Score: -3.76
Mahalanobis Distance: 7.79
Mahalanobis Distance p-value: 0.859

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Cinnamic acid	Salicylic acid, ethyl ester	Acetic acid, ((4-chloro-o-tolyl)oxy)-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.390	0.438	0.447
Reference	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 16,687,1978	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 16,751,1978	ATDAEI Acute Toxicity Data. Journal of the American College of Toxicology, Pa rt B. (Mary Ann Liebert, Inc., 1651 Third Ave., New York, NY 10128) V.1- 1990- Volume(issue)/page/year: 1,85,1990

Model Applicability

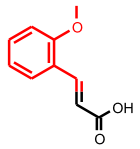
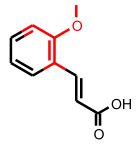
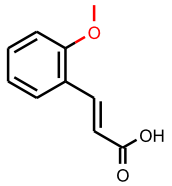
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

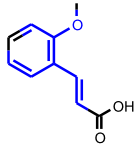
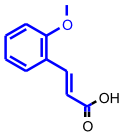
Feature Contribution

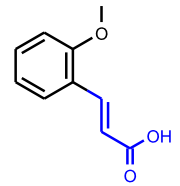
Top features for positive contribution

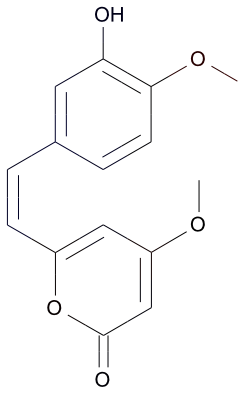
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set

FCFP_12	1458856986	 [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	0.493	3 out of 4
FCFP_12	-1977641857	 [*][c](:[*]):[c](OC):[cH]:[*]	0.416	18 out of 32
FCFP_12	136627117	 [*]OC	0.361	47 out of 90

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	-395337223	 [*]O[c]1:[cH]:[*]:[cH]:[cH]:[c]:1\C=C\[*]	-0.733	0 out of 3
FCFP_12	-1996489595	 [*]C(=[*])\C=C\[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	-0.543	0 out of 2

FCFP_12	-2107131107	 [*]\C=C\C(=O)O	-0.458	1 out of 6



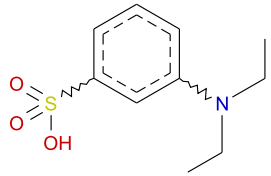
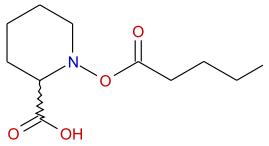
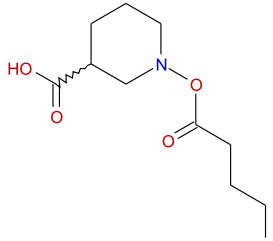
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.305
Enrichment: 0.828
Bayesian Score: -2.58
Mahalanobis Distance: 8.36
Mahalanobis Distance p-value: 0.612

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzenesulfonic acid, 3-(diethylamino)-, sodium salt	1-Valeryl-2-pipecoline	1-Valeryl-3-pipecoline
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.587	0.613	0.622
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1056,1986	US ARMY	US ARMY

Model Applicability

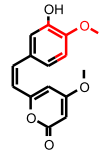
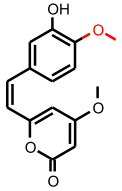
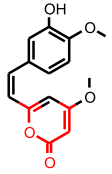
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

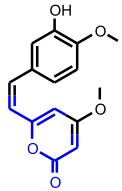
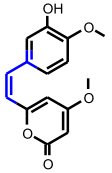
Feature Contribution

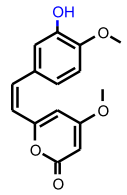
Top features for positive contribution

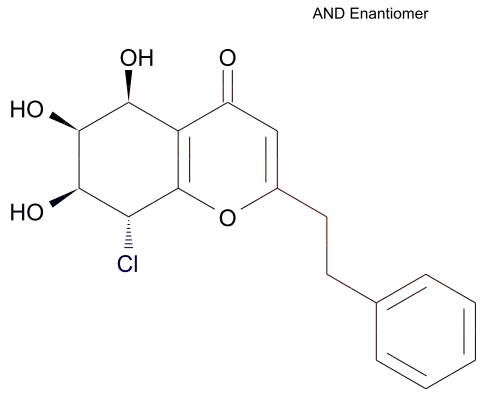
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set

FCFP_12	-1977641857	 [*][c](:[*]):[c](OC):[cH]:[*]	0.416	18 out of 32
FCFP_12	136627117	 [*]OC	0.361	47 out of 90
FCFP_12	-1251367201	 [*]C1=[*]C(=CC(=O)O1)[*]	0.295	11 out of 22

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	-166452859	 [*]=CC1=C[*]=CC(=O)O1	-0.543	0 out of 2
FCFP_12	451371068	 [*]\C=C\[c](:[*]):[*]	-0.226	9 out of 32

FCFP_12	7	 [*]O	-0.223	20 out of 70



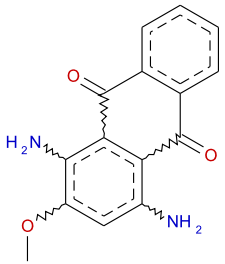
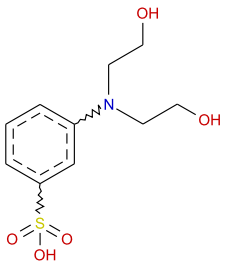
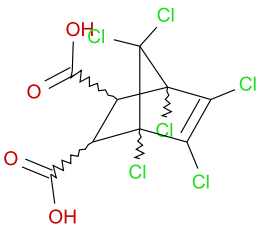
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Mild
Probability: 0.27
Enrichment: 0.733
Bayesian Score: -3.4
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 4.28e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Anthraquinone, 1,4-diamino-2-methoxy-	C.I. Fluorescent Brightening Agent 24	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.743	0.757	0.758
Reference	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: AD-A172-758	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986

Model Applicability

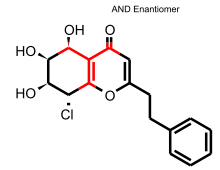
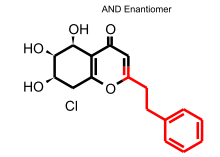
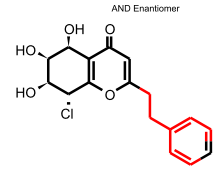
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC16 out of range. Value: 3.6979. Training min, max, SD, explained variance: -4.234, 3.1326, 1.134, 0.0174.

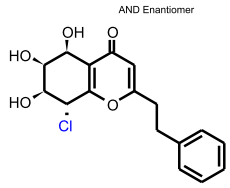
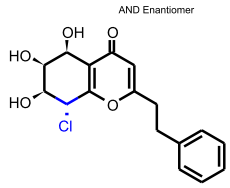
Feature Contribution

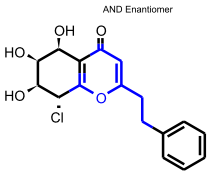
Top features for positive contribution

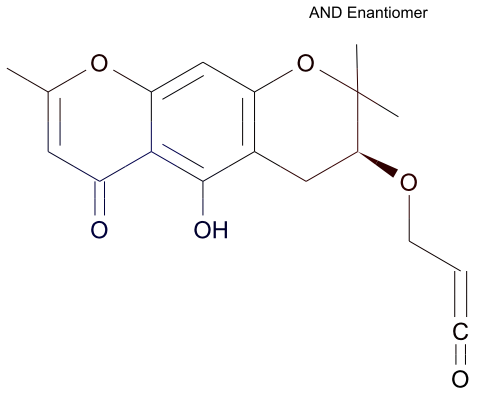
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set

FCFP_12	436886043	 [*]C([*])C(=C([*])[*])C(=[*])[*]	0.503	68 out of 113
FCFP_12	1388176727	 [*]C(=[*])CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.458	4 out of 6
FCFP_12	-497728148	 [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.332	21 out of 41

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	32	 [*]Cl	-0.789	16 out of 101
FCFP_12	71953198	 [*]C([*])Cl	-0.548	10 out of 50

FCFP_12	-1305493555	 [*]CCC1=CC(=[*])[*]=C([*])O1	-0.308	0 out of 1



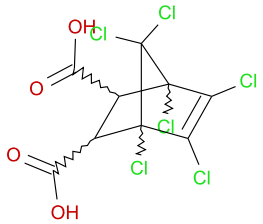
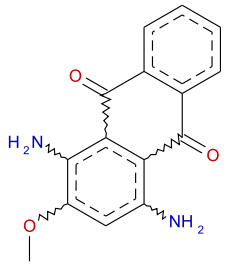
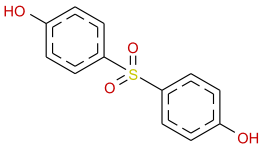
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild
Probability: 0.258
Enrichment: 0.701
Bayesian Score: -3.69
Mahalanobis Distance: 12.7
Mahalanobis Distance p-value: 1.23e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	Anthraquinone, 1,4-diamino-2-methoxy-	Phenol, 4,4'-sulfonyldi-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.642	0.691	0.692
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: AD-A172-758	BIOFX* BIOFAX Industrial Bio-Test Laboratories, Inc., Data Sheets. (1810 Fro ntage Rd., Northbrook, IL 60062) Volume(issue)/page/year: 601-05501,1974

Model Applicability

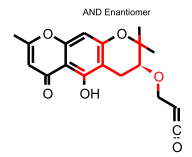
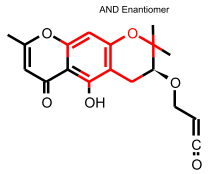
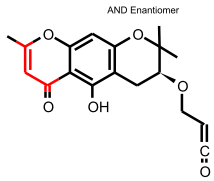
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 444624378: [*]C=C=O

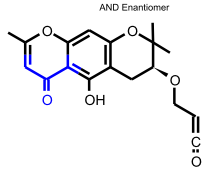
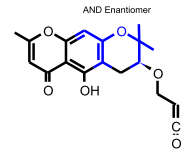
Feature Contribution

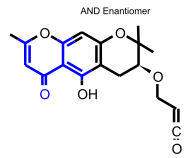
Top features for positive contribution

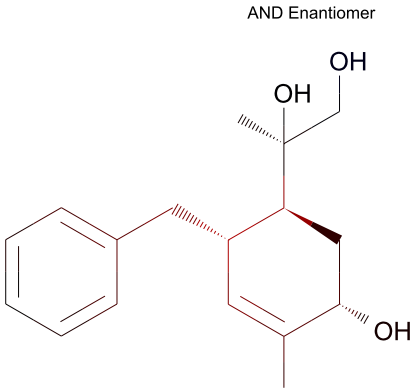
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set

FCFP_12	-2000646149	 [*]O[C@H]1C[c](:[c]([*]):[*]):[c](:[*])[*]C1([*])[*]	0.365	2 out of 3
FCFP_12	-1099193755	 [*]=C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1OC	0.343	6 out of 11
FCFP_12	451847724	 [*]\C=C\C(=[*])[*]	0.257	101 out of 216

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	-1549192822	 [*]=CC(=O)[c](:[*]):[*]	-0.909	1 out of 11
FCFP_12	-1979984779	 [*][C@H]1[*][c](:[*]):[c](OC1(C)C):[cH]:[*]	-0.308	0 out of 1

FCFP_12	1244036906	 [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*]	-0.308	0 out of 1



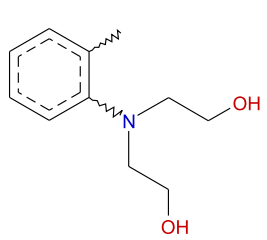
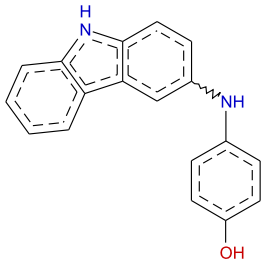
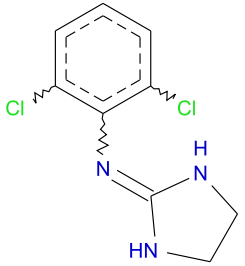
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Mild
Probability: 0.452
Enrichment: 1.23
Bayesian Score: 0.935
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.00124

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	o-Toluidine, N,N-bis(2-hydroxyethyl)-	Phenol, 4-(3-carbazolylamino)-	2-Imidazoline, 2-(2,6-dichloroanilino)-
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.714	0.747	0.754
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,697,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,825,1986	OYYAA2 Oyo Yakuri. Pharmacometrics. (Oyo Yakuri Kenkyukai, CPO Box 180, Sen dai 980-91, Japan) V.1- 1967- Volume(issue)/page/year: 45,257,1993

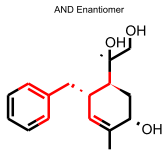
Model Applicability

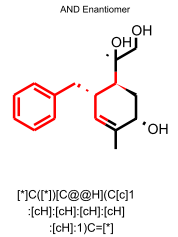
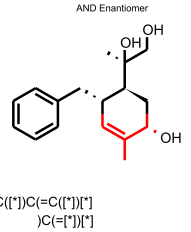
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

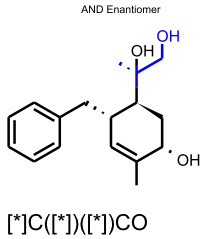
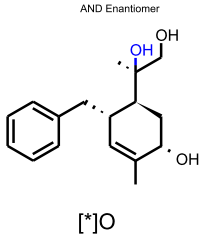
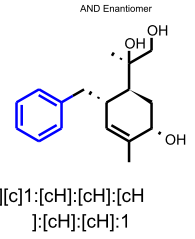
Feature Contribution

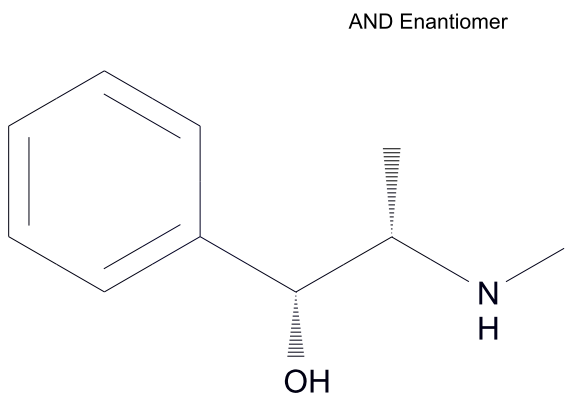
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	-2005486458	 [*]C([*])[C@@H](C[c](:[cH]:[*]):[cH]:[*])C=[*]	0.717	4 out of 4

FCFP_12	1384456758		0.653	3 out of 3
FCFP_12	436886043		0.503	68 out of 113

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	-1272709286		-0.475	18 out of 82
FCFP_12	3		-0.234	58 out of 204
FCFP_12	-2093839777		-0.207	30 out of 103



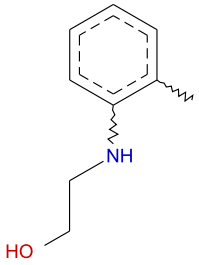
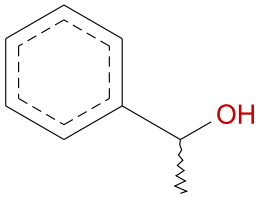
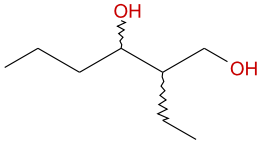
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.958
Enrichment: 1.04
Bayesian Score: -1.63
Mahalanobis Distance: 8.62
Mahalanobis Distance p-value: 0.624

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ethanol, 2-toluidino-	Benzyl alcohol, alpha-methyl-	1,3-Hexanediol, 2-ethyl-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.439	0.519	0.557
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,693,1986	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 12,995,1974	34ZIAG "Toxicology of Drugs and Chemicals," Deichmann, W.B., New York, Academ ic Press, Inc., 1969 Volume(issue)/page/year: -,731,1969

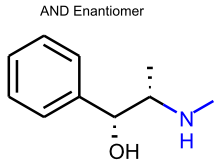
Model Applicability

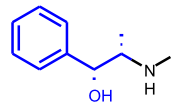
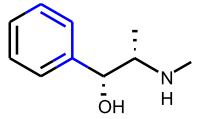
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

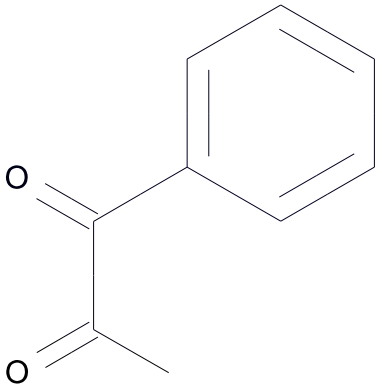
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	136686699	 [*]NC	-0.484	3 out of 6

FCFP_12	-1083860676	AND Enantiomer  [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[cH]:[cH] :[cH]:1	-0.153	3 out of 4
FCFP_12	1618154665	AND Enantiomer  [*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490



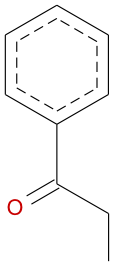
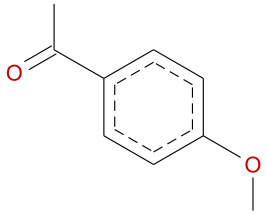
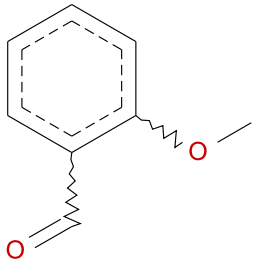
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: Non-Irritant
Probability: 0.973
Enrichment: 1.06
Bayesian Score: -0.795
Mahalanobis Distance: 4.84
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Propiophenone	Acetophenone, 4'-methoxy-	o-Anisaldehyde
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.411	0.427	0.430
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,291,1986	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 12,927,1974	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 17,855,1979

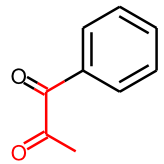
Model Applicability

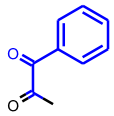
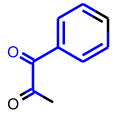
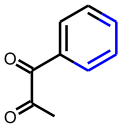
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

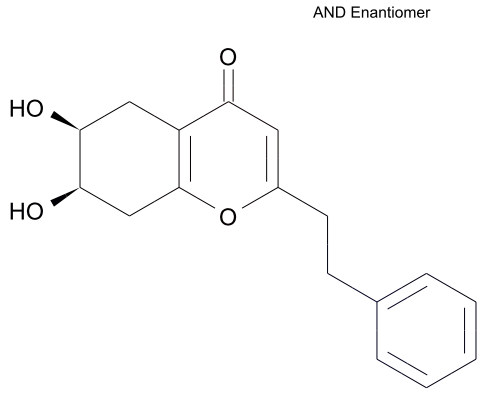
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79

Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Irritant in training set**
FCFP_12	975909016	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.261	4 out of 6
FCFP_12	-581162801	 [*]C(=[*])C(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.101	9 out of 11
FCFP_12	1618154665	 [*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490



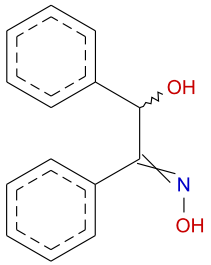
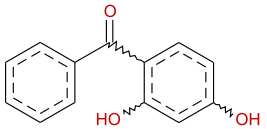
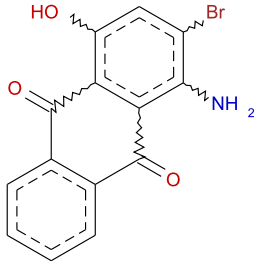
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.97
Enrichment: 1.05
Bayesian Score: -0.999
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.000773

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzoin, oxime	Benzophenone, 2,4-dihydroxy-	1-Amino-2-bromo-4-hydroxyanthraquinone
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.639	0.668	0.669
Reference	28ZPAK -,111,72	28ZPAK -,101,72	28ZPAK -,83,72

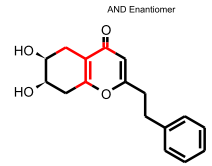
Model Applicability

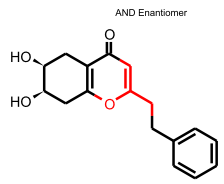
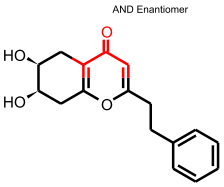
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

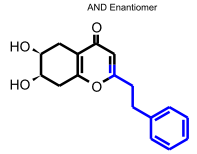
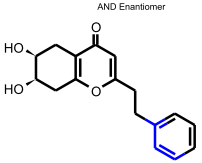
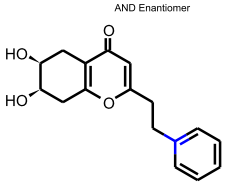
Feature Contribution

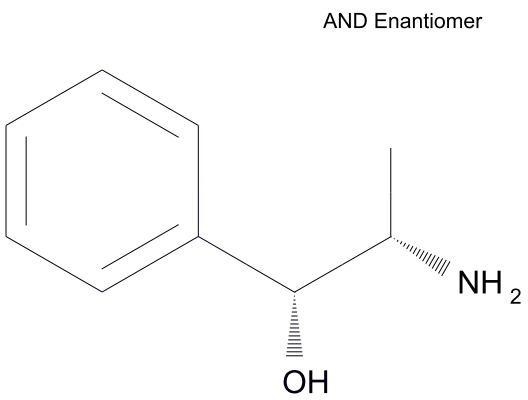
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	436886043	 [*]CC(=C([*])[*])C(=[*])[*]	0.0804	129 out of 130

FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.0756	6 out of 6
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1388176727	 [*]C(=[*])CC[c]1:[cH] :[cH]:[cH]:[cH]:[cH] :1	-0.12	8 out of 10
FCFP_12	1618154665	 [*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490
FCFP_12	16	 [*][c](:[*]):[*]	-0.0843	423 out of 503



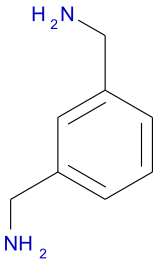
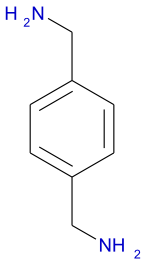
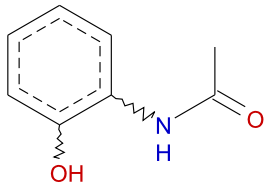
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.968
Enrichment: 1.05
Bayesian Score: -1.16
Mahalanobis Distance: 5.58
Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Methylamine, m-phenylenebis-	Methylamine, p-phenylenebis-	Acetanilide, 2'-hydroxy-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.460	0.468	0.492
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,447,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,448,1986	28ZPAK -,106,72

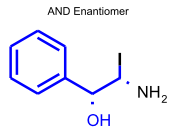
Model Applicability

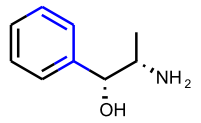
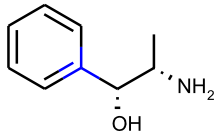
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

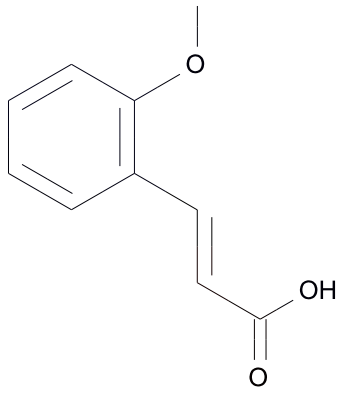
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1083860676	 [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[cH]:[cH] :[cH]:1	-0.153	3 out of 4

FCFP_12	1618154665	AND Enantiomer  [*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490
FCFP_12	16	AND Enantiomer  [*][c](:[*]):[*]	-0.0843	423 out of 503



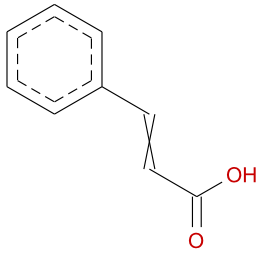
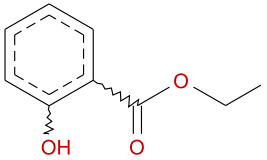
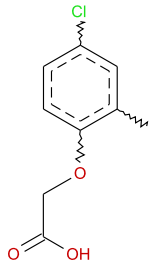
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 0.973
Enrichment: 1.06
Bayesian Score: -0.728
Mahalanobis Distance: 7.74
Mahalanobis Distance p-value: 0.941

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Cinnamic acid	Salicylic acid, ethyl ester	Acetic acid, ((4-chloro-o-tolyl)oxy)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.397	0.438	0.449
Reference	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 16,687,1978	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 16,751,1978	ATDAEI Acute Toxicity Data. Journal of the American College of Toxicology, Pa rt B. (Mary Ann Liebert, Inc., 1651 Third Ave., New York, NY 10128) V.1- 1990- Volume(issue)/page/year: 1,85,1990

Model Applicability

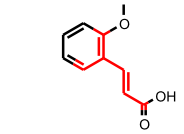
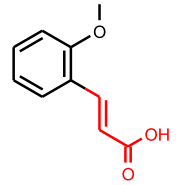
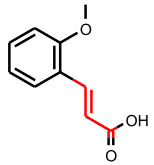
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

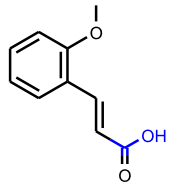
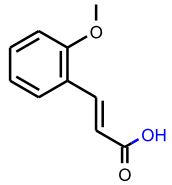
Feature Contribution

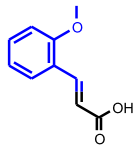
Top features for positive contribution

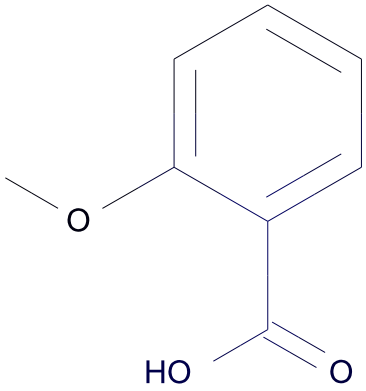
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	0.085	24 out of 24
FCFP_12	-2107131107	 [*]\C=C\C(=O)O	0.0785	8 out of 8
FCFP_12	451847724	 [*]C(=CC(=[*])[*])[*]	0.0737	270 out of 274

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-548632217	 [*]C(=[*])O	-0.128	49 out of 61
FCFP_12	7	 [*]O	-0.118	104 out of 128

FCFP_12	1458856986	 [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1OC	-0.109	4 out of 5



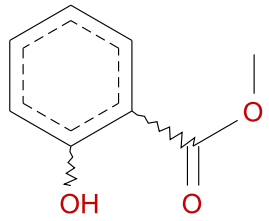
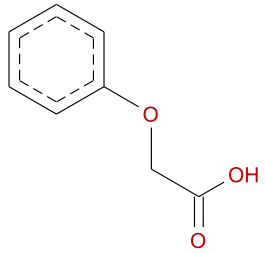
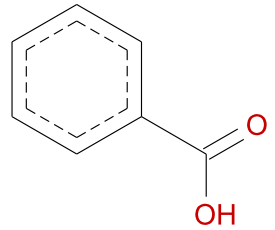
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 0.931
Enrichment: 1.01
Bayesian Score: -2.31
Mahalanobis Distance: 10.7
Mahalanobis Distance p-value: 0.00653

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Salicylic acid, methyl ester	Acetic acid, phenoxy-	Benzoic acid
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.355	0.381	0.393
Reference	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 16,821,1978	FCTXAV Food and Cosmetics Toxicology. (London, UK) V.1-19, 1963-81. For pu blisher information, see FCTOD7. Volume(issue)/page/year: 17,887,1979	BIOFX* BIOFAX Industrial Bio-Test Laboratories, Inc., Data Sheets. (1810 Fro ntage Rd., Northbrook, IL 60062) Volume(issue)/page/year: 28-4/1973

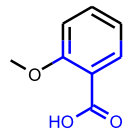
Model Applicability

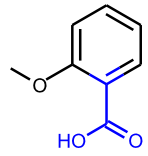
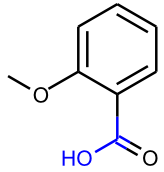
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

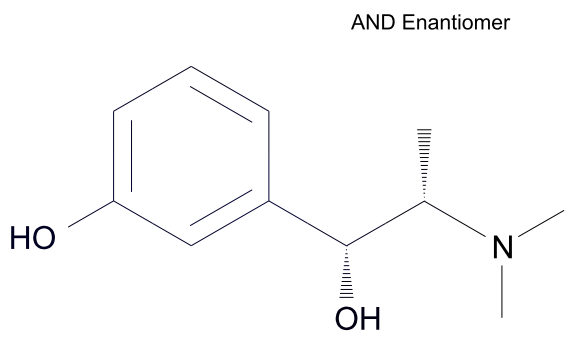
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	937923569	 [*][c](:[*]):[c](:[cH]:[*])C(=O)O	-0.612	5 out of 11

FCFP_12	-1549222613	 [*]:[c](:[*])C(=O)O	-0.612	5 out of 11
FCFP_12	-548632217	 [*]C(=[*])O	-0.128	49 out of 61



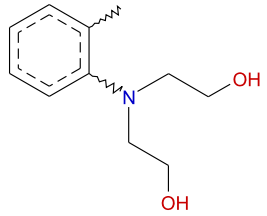
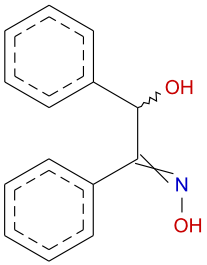
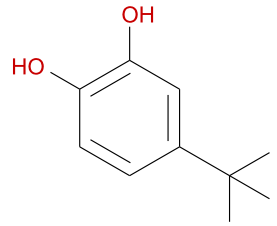
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.951
Enrichment: 1.03
Bayesian Score: -1.85
Mahalanobis Distance: 16.2
Mahalanobis Distance p-value: 4.69e-019

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	o-Toluidine, N,N-bis(2-hydroxyethyl)-	Benzoin, oxime	Pyrocatechol, 4-tert-butyl-
Structure			
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.513	0.529	0.551
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,697,1986	28ZPAK -,111,72	AMIHBC 10,61,54

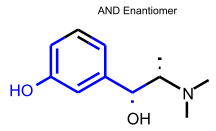
Model Applicability

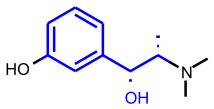
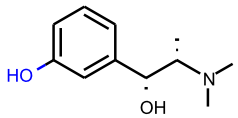
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

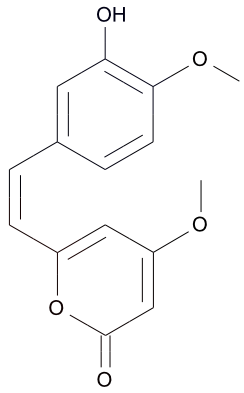
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	949015626	 [*]C([*])[c]1:[cH]:[*]:[cH]:[c](O):[cH]:1	-0.222	2 out of 3

FCFP_12	-1083860676	AND Enantiomer  [*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[cH]:[cH] :[cH]:1	-0.153	3 out of 4
FCFP_12	7	AND Enantiomer  [*]O	-0.118	104 out of 128



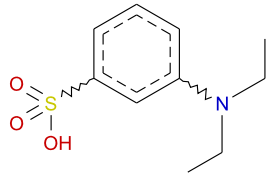
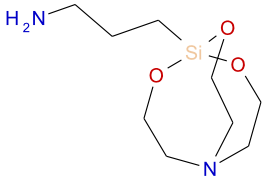
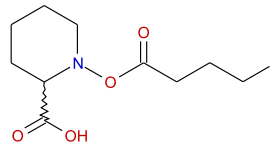
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 0.976
Enrichment: 1.06
Bayesian Score: -0.515
Mahalanobis Distance: 7.73
Mahalanobis Distance p-value: 0.943

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzenesulfonic acid, 3-(diethylamino)-, sodium salt	2,8,9-Trioxa-5-aza-1-silabicyclo(3.3.3)undecane, 1-(3-aminopropyl)-	1-Valeryl-2-pipecoline
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.592	0.607	0.616
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1056,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1233,1986	US ARMY

Model Applicability

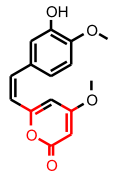
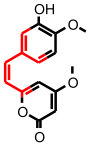
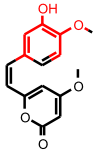
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

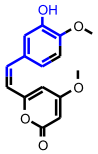
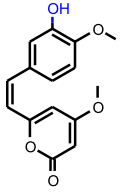
Feature Contribution

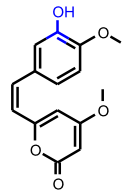
Top features for positive contribution

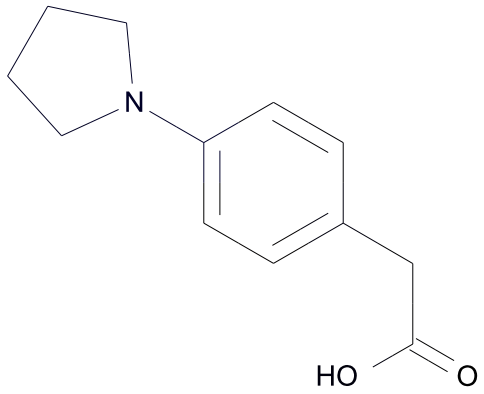
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1251367201	 [*]C1=[*]C(=CC(=O)O1) [*]	0.0858	31 out of 31
FCFP_12	-146015125	 [*]C(=[*])\C=C\[c](:[cH]:[*]):[c]([*]):[*]	0.085	24 out of 24
FCFP_12	523826990	 [*]O[c]1:[cH]:[*]:[c] ([*]):[cH]:[c]:1O	0.0756	6 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	949015626	 [*]C([*])[c]1:[cH]:[*]:[cH]:[c](O):[cH]:1	-0.222	2 out of 3
FCFP_12	7	 [*]O	-0.118	104 out of 128

FCFP_12	-549108873	 [*]:[c](:[*])O	-0.11	54 out of 66



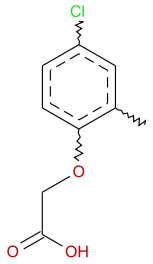
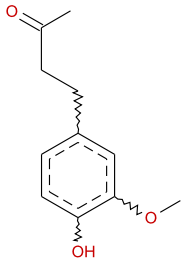
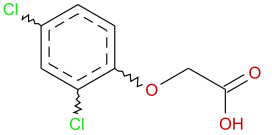
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 0.964
Enrichment: 1.05
Bayesian Score: -1.37
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.00314

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Acetic acid, ((4-chloro-o-tolyl)oxy)-	2-Butanone, 4-(4-hydroxy-3-methoxyphenyl)-	Acetic acid, (2,4-dichlorophenoxy)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.476	0.481	0.498
Reference	ATDAEI Acute Toxicity Data. Journal of the American College of Toxicology, Pa rt B. (Mary Ann Liebert, Inc., 1651 Third Ave., New York, NY 10128) V.1- 1990- Volume(issue)/page/year: 1,85,1990	FCTOD7 Food and Chemical Toxicology. (Pergamon Press Inc., Maxwell House, Fa irview Park, Elmsford, NY 10523) V.20- 1982- Volume(issue)/page/year: 20,851 ,1982	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,279,1

Model Applicability

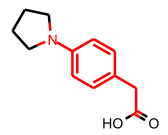
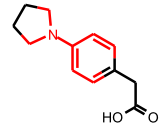
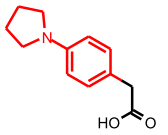
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

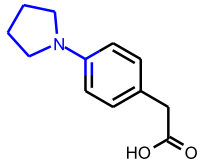
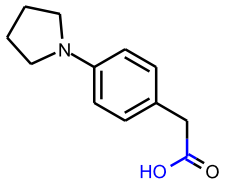
Feature Contribution

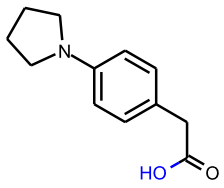
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

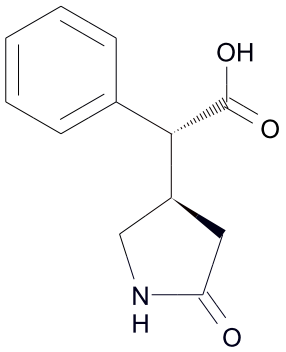
FCFP_12	1985089045	 [*]N([*])[c]1:[cH]:[cH]:[c](CC(=[*])[*]):[cH]:[cH]:1	0.0785	8 out of 8
FCFP_12	1604677718	 [*]1[*]CN(C1)[c]2:[cH]:[cH]:[*]:[cH]:[cH]:2	0.0785	8 out of 8
FCFP_12	202105689	 [*][c]1:[cH]:[cH]:[c](:[cH]:[cH]:1)N2CCCC2	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-822674211	 [*]:[c](:[*])N1CCCC1	-0.65	0 out of 1
FCFP_12	-548632217	 [*]C(=[*])O	-0.128	49 out of 61

FCFP_12	7	 [*]O	-0.118	104 out of 128

AND Enantiomer



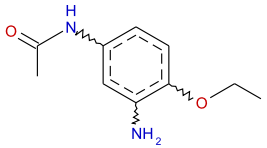
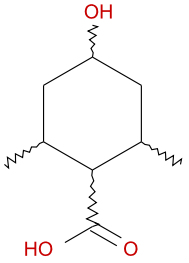
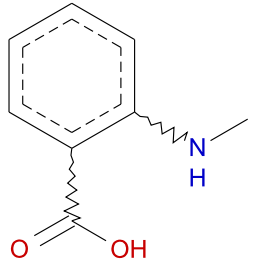
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.961
Enrichment: 1.04
Bayesian Score: -1.5
Mahalanobis Distance: 9.76
Mahalanobis Distance p-value: 0.111

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Acetanilide, 3'-amino-4'-ethoxy-	Ethyl 4-hydroxy-2,6-dimethylcyclohexane carboxylate	Anthranilic acid, N-methyl-
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.504	0.560	0.564
Reference	28ZPAK -,115,72	US ARMY	28ZPAK -,441,72

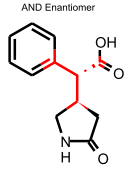
Model Applicability

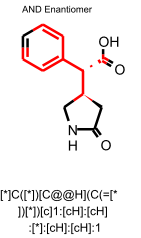
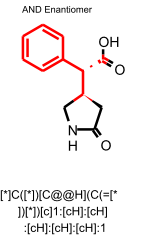
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

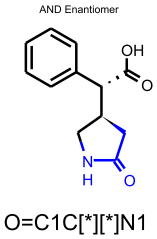
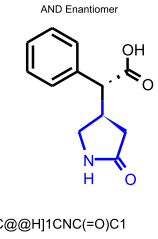
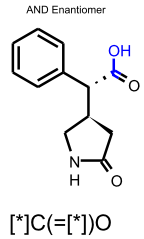
Feature Contribution

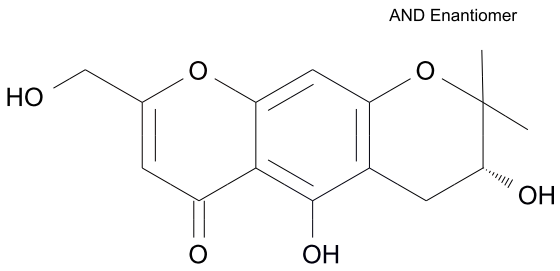
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1186303932	AND Enantiomer  [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]	0.0838	18 out of 18

FCFP_12	-1133295320	 [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.0816	12 out of 12
FCFP_12	-1716639150	 [*]C([*])[C@@H](C(=[*])[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.0658	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	566058135	 O=C1C[*][*]N1	-0.367	13 out of 21
FCFP_12	-547731249	 [*][C@@H]1CNC(=O)C1	-0.222	2 out of 3
FCFP_12	-548632217	 [*]C(=[*])O	-0.128	49 out of 61



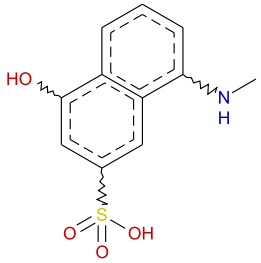
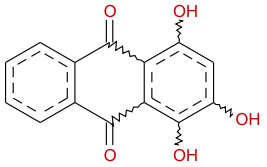
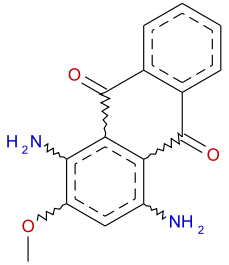
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: Non-Irritant
Probability: 0.97
Enrichment: 1.05
Bayesian Score: -0.998
Mahalanobis Distance: 12.3
Mahalanobis Distance p-value: 4.91e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	8-Methylamino-4-hydroxy-2-naphthalene sulfonic acid	Anthraquinone, 1,2,4-trihydroxy-	Anthraquinone, 1,4-diamino-2-methoxy-
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.619	0.650	0.692
Reference	28ZPAK -,190,72	28ZPAK -,103,72	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: AD-A172-758

Model Applicability

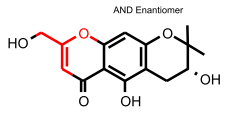
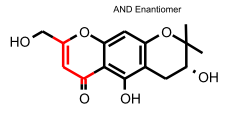
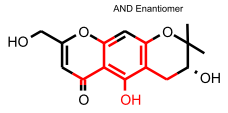
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

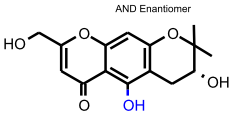
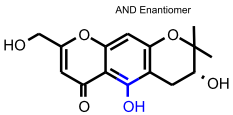
Feature Contribution

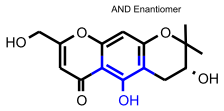
Top features for positive contribution

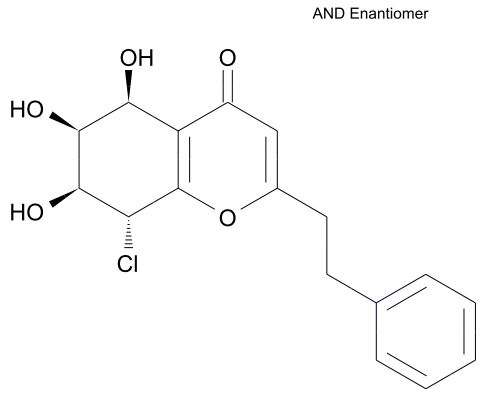
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.0756	6 out of 6
FCFP_12	451847724	 [*]C(=CC(=[*])[*])[*]	0.0737	270 out of 274
FCFP_12	-1601875224	 [*]C[c]1:[c]([*]):[*] :[c]([*]):[c](C(=[*])[*]):[c]:1O	0.0734	5 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	7	 [*]O	-0.118	104 out of 128
FCFP_12	-549108873	 [*]:[c](:[*])O	-0.11	54 out of 66

FCFP_12	74595001	 [*]:[cH]:[c](O):[cH]: [*]	-0.11	54 out of 66



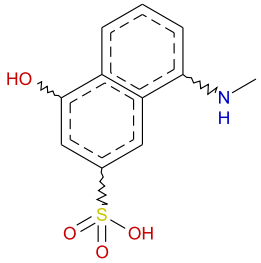
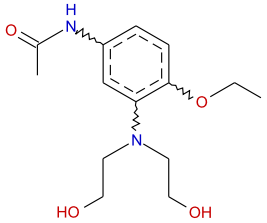
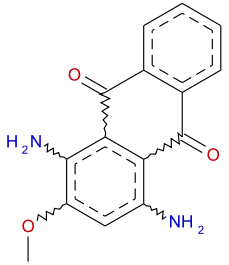
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: Irritant
Probability: 0.975
Enrichment: 1.06
Bayesian Score: -0.625
Mahalanobis Distance: 12.4
Mahalanobis Distance p-value: 2.86e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	8-Methylamino-4-hydroxy-2-naphthalene sulfonic acid	p-Acetophenetidide, 3'-(bis(2-hydroxyethyl)amino)-	Anthraquinone, 1,4-diamino-2-methoxy-
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.703	0.738	0.739
Reference	28ZPAK -,190,72	28ZPAK -,100,72	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: AD-A172-758

Model Applicability

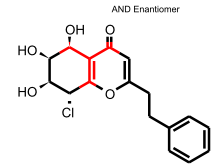
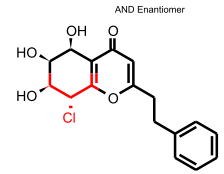
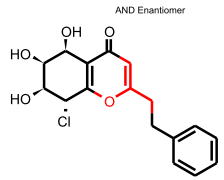
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

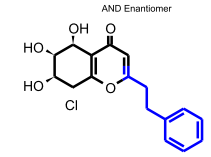
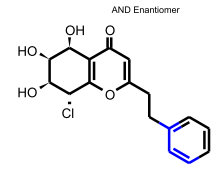
Feature Contribution

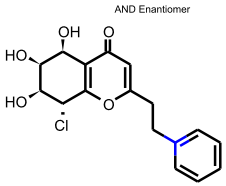
Top features for positive contribution

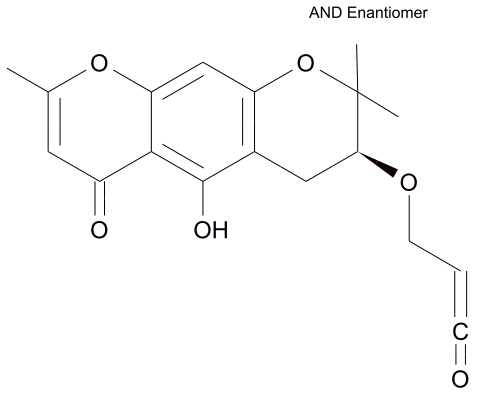
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	436886043	 [*]CC(=C([*])[*])C(=[*])[*]	0.0804	129 out of 130
FCFP_12	-879019572	 [*]C([*])[C@@H](Cl)C(=[*])[*]	0.0772	7 out of 7
FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.0756	6 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1388176727	 [*]C(=[*])CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.12	8 out of 10
FCFP_12	1618154665	 [*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490

FCFP_12	16	 [*][c](:[*]):[*]	-0.0843	423 out of 503



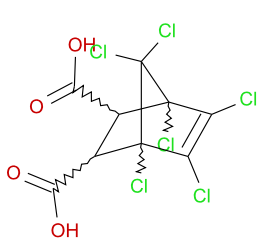
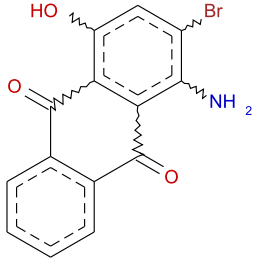
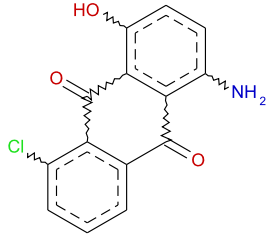
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 0.975
Enrichment: 1.06
Bayesian Score: -0.544
Mahalanobis Distance: 13
Mahalanobis Distance p-value: 6.69e-008

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-2-bromo-4-hydroxyanthraquinone	1-Amino-4-hydroxy-5-chloroanthraquinone
Structure			
Actual Endpoint	Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.649	0.655	0.692
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72	28ZPAK -,83,72

Model Applicability

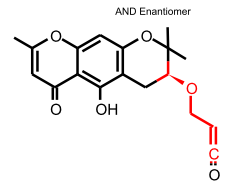
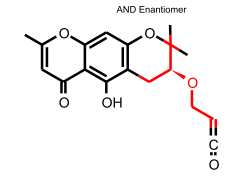
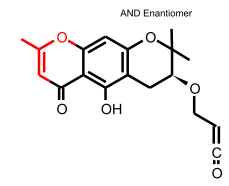
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 444624378: [*]C=C=O

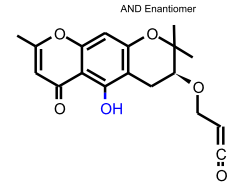
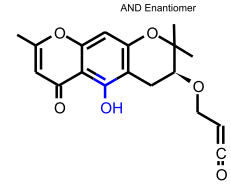
Feature Contribution

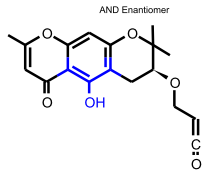
Top features for positive contribution

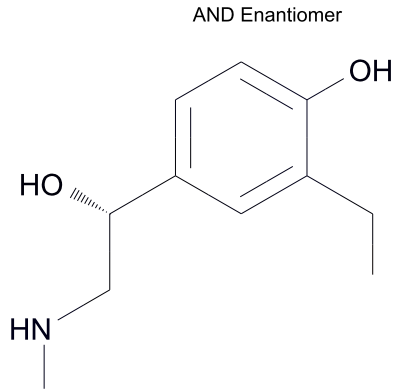
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	353886735	 [*]C([*])OCC=C=[*]	0.0863	37 out of 37
FCFP_12	-1687409379	 [*]C[C@H](OCC=[*])C([*])([*])[*]	0.0795	9 out of 9
FCFP_12	436915834	 [*]C\C(=C\[*])\O[*]	0.0756	6 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	7	 [*]O	-0.118	104 out of 128
FCFP_12	-549108873	 [*]:[c](:[*])O	-0.11	54 out of 66

FCFP_12	74595001	 [*]:[cH]:[c](O):[cH]:[*]	-0.11	54 out of 66



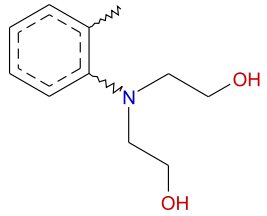
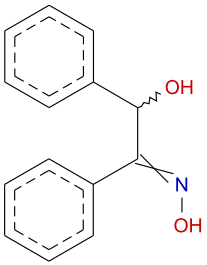
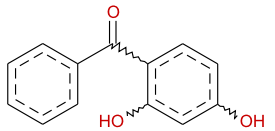
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Non-Irritant
Probability: 0.955
Enrichment: 1.04
Bayesian Score: -1.71
Mahalanobis Distance: 8.74
Mahalanobis Distance p-value: 0.559

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	o-Toluidine, N,N-bis(2-hydroxyethyl)-	Benzoin, oxime	Benzophenone, 2,4-dihydroxy-
Structure			
Actual Endpoint	Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.650	0.657	0.675
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,697,1986	28ZPAK -,111,72	28ZPAK -,101,72

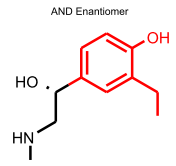
Model Applicability

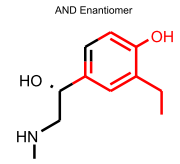
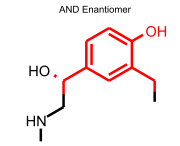
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

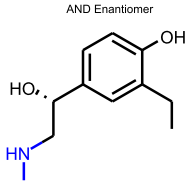
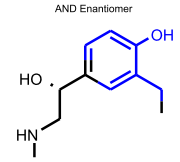
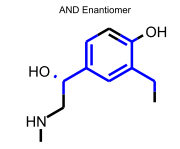
Feature Contribution

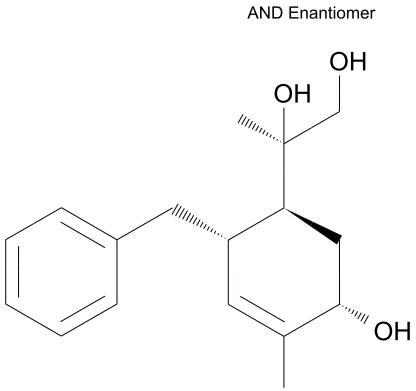
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1710155973	 [*][c]1:[cH]:[cH]:[c](O):[c](CC):[cH]:1	0.0658	3 out of 3

FCFP_12	-542873837	 [*][c]1:[*]:[cH]:[c](O):[c](CC):[cH]:1	0.0658	3 out of 3
FCFP_12	-306804326	 [*]C[c]1:[cH]:[c](:[cH]:[cH]:[c]:1O)C([*])[*]	0.0658	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	136686699	 [*]NC	-0.484	3 out of 6
FCFP_12	-1604301295	 [*]C[c]1:[cH]:[*]:[cH]:[cH]:[c]:1O	-0.18	22 out of 29
FCFP_12	-451251206	 [*]C[c]1:[cH]:[c](:[cH]:[*]:[c]:1[*])C([*])[*]	-0.132	44 out of 55



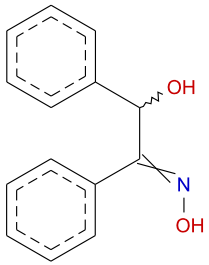
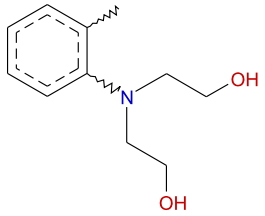
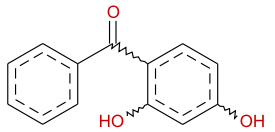
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: Irritant
Probability: 0.977
Enrichment: 1.06
Bayesian Score: -0.373
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.00105

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzoin, oxime	o-Toluidine, N,N-bis(2-hydroxyethyl)-	Benzophenone, 2,4-dihydroxy-
Structure			
Actual Endpoint	Non-Irritant	Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.669	0.713	0.724
Reference	28ZPAK -,111,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,697,1986	28ZPAK -,101,72

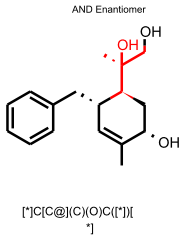
Model Applicability

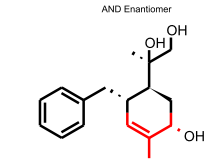
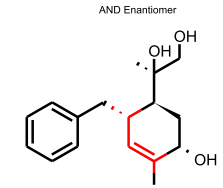
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

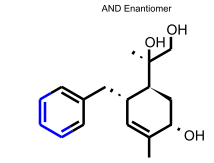
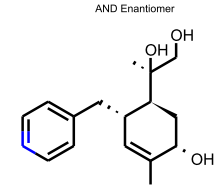
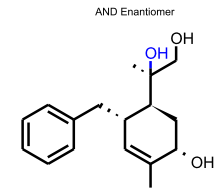
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-415156552	 [*]C[C@](C)(O)C([*])[*]	0.0854	27 out of 27

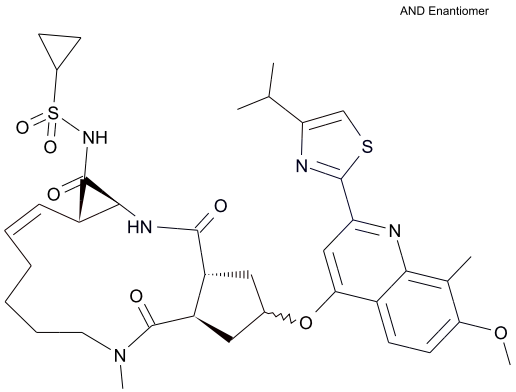
FCFP_12	436886043	 [*]CC(=C([*])[*])C(=[*])[*]	0.0804	129 out of 130
FCFP_12	451847724	 [*]C(=CC(=[*])[*])[*]	0.0737	270 out of 274

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1618154665	 [*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490
FCFP_12	16	 [*][c](:[*]):[*]	-0.0843	423 out of 503
FCFP_12	3	 [*]O	-0.0812	291 out of 345

Simeprevir

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.969

Enrichment: 1.05

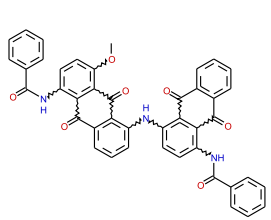
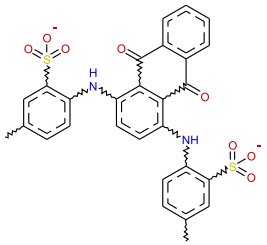
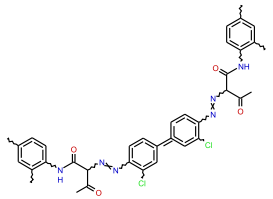
Bayesian Score: -1.05

Mahalanobis Distance: 16.8

Mahalanobis Distance p-value: 1.98e-021

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Anthraquinone, 3-methoxy-5,4'-iminobis(1-benzamido-	Benzenesulfonic acid, 2,2'-(1,4-anthraquinonylenediimino)bis(5-meth yl-, disodium salt	Butanamide, 2,2'-((3,3'-dichloro(1,1'-biphenyl)-4,4'-diyl)bis(azo)) bis(N-(2,4- dimethylphenyl)-3-oxo-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.883	1.016	1.027
Reference	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,114,1	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1330,1986	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: OTS0555058

Model Applicability

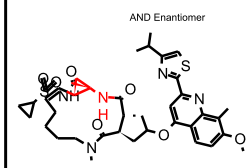
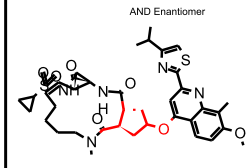
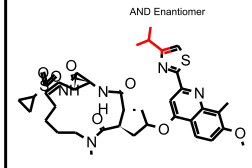
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC6 out of range. Value: 7.6964. Training min, max, SD, explained variance: -5.7225, 6.867, 1.786, 0.0403.
2. OPS PC22 out of range. Value: 5.13. Training min, max, SD, explained variance: -2.9568, 3.7845, 1.016, 0.0131.

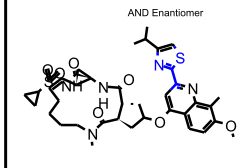
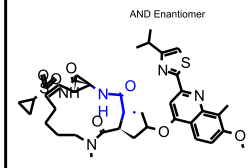
Feature Contribution

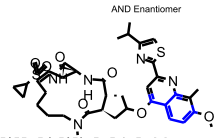
Top features for positive contribution

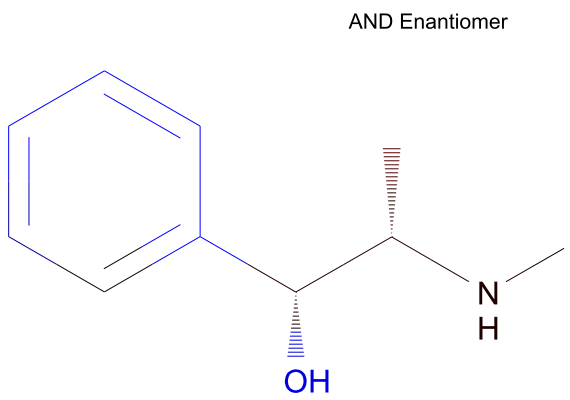
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-415156552	 [*]C[C@](C)(O)C([*])[*]	0.0854	27 out of 27
FCFP_12	-53728878	 [*]OC1C[C@@H]([*])[C@ @H](C1)C(=[*])[*]	0.0845	21 out of 21
FCFP_12	1186303932	 [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]	0.0838	18 out of 18

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	690481386	 [*]:[c](:[*])[c]1:n:[*]:[*]:s:1	-0.65	0 out of 1
FCFP_12	566058135	 O=C1C[*][*]N1	-0.367	13 out of 21

FCFP_12	-1320007763	 [*][c](:[*]):[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1:[*]	-0.0893	20 out of 24



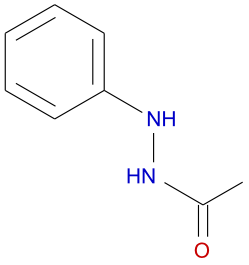
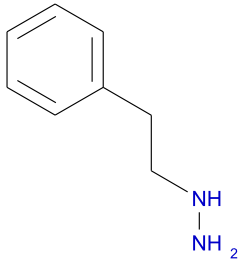
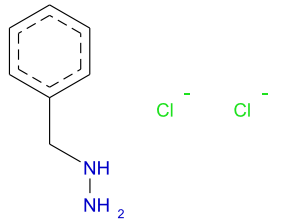
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 307
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.27
Mahalanobis Distance p-value: 0.185

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	1-Acetyl-2-phenylhydrazine	529	485
Structure			
Actual Endpoint (-log C)	3.46734	3.96981	3.15601
Predicted Endpoint (-log C)	2.52008	3.43178	3.43754
Distance	0.464	0.469	0.497
Reference	CPDB	CPDB	CPDB

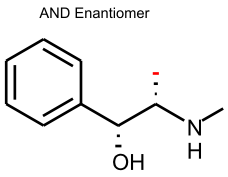
Model Applicability

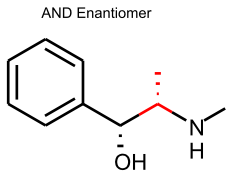
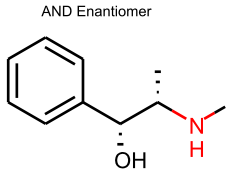
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -222940837: [*]C([*])[C@H](O)[c](:[*]):[*]
3. Unknown ECFP_2 feature: -871548073: [*]N[C@@H](C)C([*])[*]
4. Unknown ECFP_2 feature: 1336634178: [*]C([*])NC

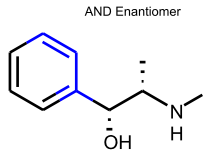
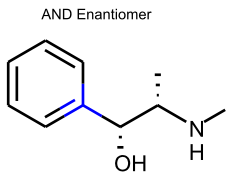
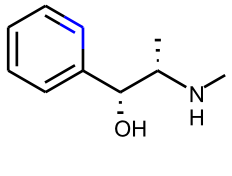
Feature Contribution

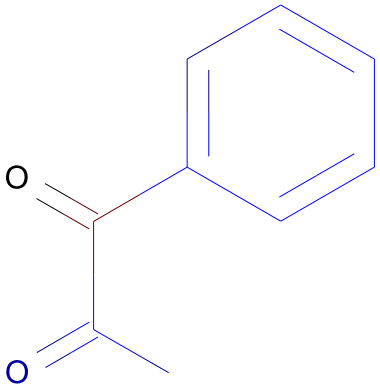
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	734603939	 [*]C	0.0424

ECFP_6	864518973	 [*]C([*])C	0.0366
ECFP_6	-1897341097	 [*]N[*]	0.0284

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



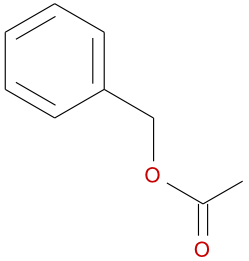
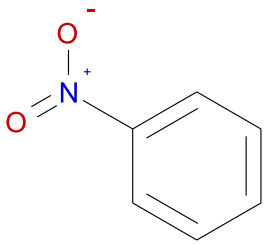
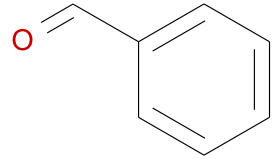
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: 734
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.35
Mahalanobis Distance p-value: 0.159

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzyl acetate	Nitrobenzene	Benzaldehyde
Structure			
Actual Endpoint (-log C)	2.01823	2.619	1.85262
Predicted Endpoint (-log C)	2.37939	2.53727	2.81264
Distance	0.417	0.464	0.500
Reference	CPDB	CPDB	CPDB

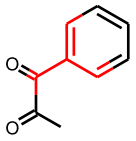
Model Applicability

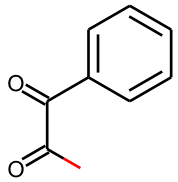
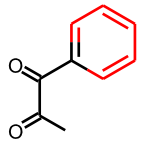
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1432101658: [*]C(=[*])C(=O)[c](:[*]):[*]
3. Unknown ECFP_2 feature: -472613004: [*]C(=[*])C(=O)C

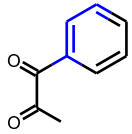
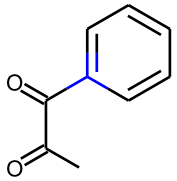
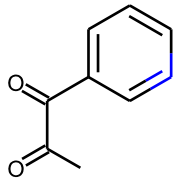
Feature Contribution

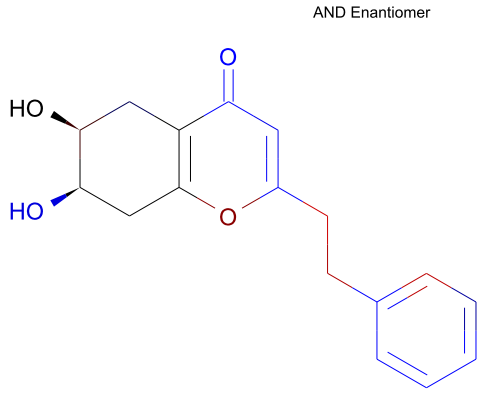
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

ECFP_6	734603939	 [*]C	0.0424
ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.0145

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



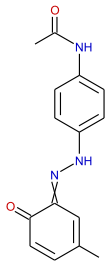
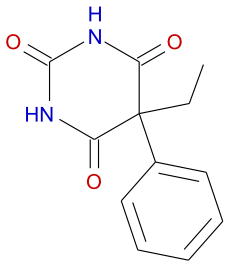
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: 71.2
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.68
Mahalanobis Distance p-value: 0.0794

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	C.I. disperse yellow 3	Phenobarbital s	581
Structure			
Actual Endpoint (-log C)	2.42163	4.49846	3.89128
Predicted Endpoint (-log C)	3.11116	3.14828	3.31105
Distance	0.586	0.592	0.633
Reference	CPDB	CPDB	CPDB

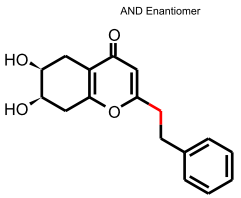
Model Applicability

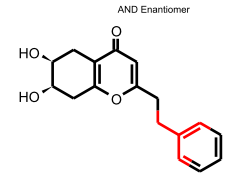
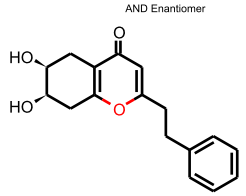
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1650944136: [*]CC(=C([*])[*])O[*]

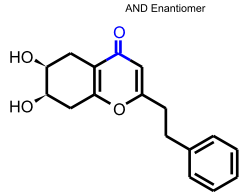
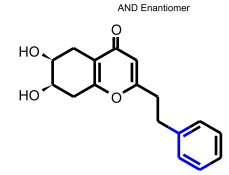
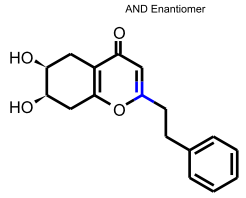
Feature Contribution

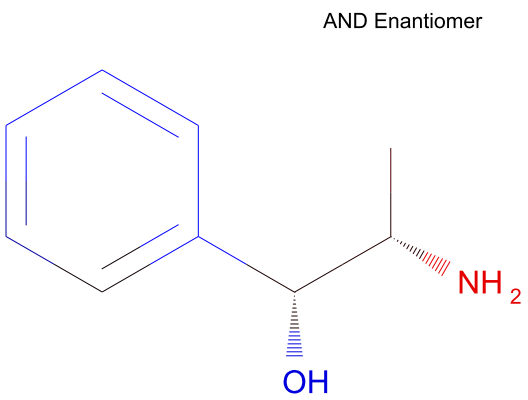
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

ECFP_6	-2024255407	 [*]C[c](:[cH]:[*]):[cH]:[*]	0.172
ECFP_6	683445015	 [*]O[*]	0.136

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247



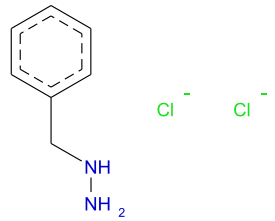
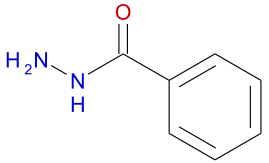
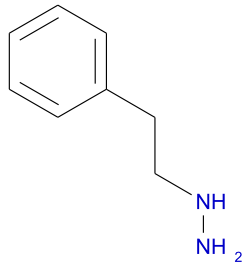
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 209
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 10.1
Mahalanobis Distance p-value: 0.0262

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	485	Benzoyl hydrazine	529
Structure			
Actual Endpoint (-log C)	3.15601	4.1522	3.96981
Predicted Endpoint (-log C)	3.43754	3.02635	3.43178
Distance	0.461	0.470	0.477
Reference	CPDB	CPDB	CPDB

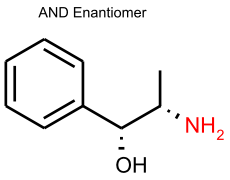
Model Applicability

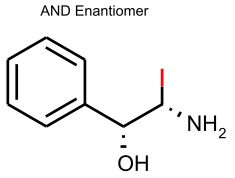
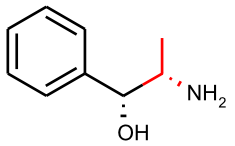
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -222940837: [*]C([*])[C@H](O)[c](:[*]):[*]

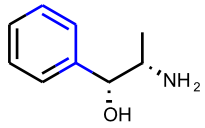
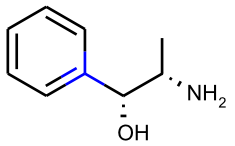
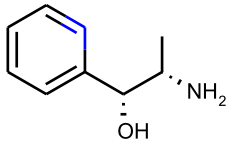
Feature Contribution

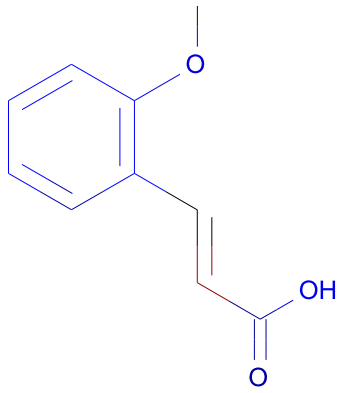
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1572579716	 [*]N	0.225

ECFP_6	734603939	AND Enantiomer  [*]C	0.0424
ECFP_6	864518973	AND Enantiomer  [*]C([*])C	0.0366

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	AND Enantiomer  [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	AND Enantiomer  [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	AND Enantiomer  [*]:[cH]:[*]	-0.232



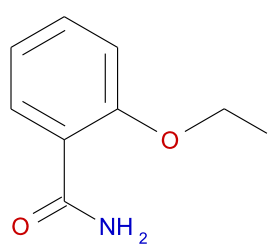
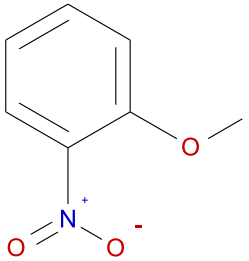
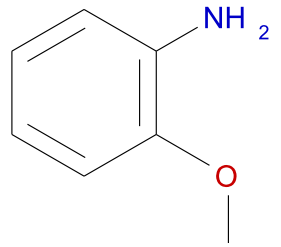
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 896
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.81
Mahalanobis Distance p-value: 0.0587

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	o-Ethoxybenzamide	o-Nitroanisole	153
Structure			
Actual Endpoint (-log C)	2.50786	2.93466	2.10547
Predicted Endpoint (-log C)	2.72346	2.36925	2.33861
Distance	0.479	0.499	0.507
Reference	CPDB	CPDB	CPDB

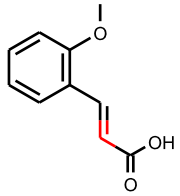
Model Applicability

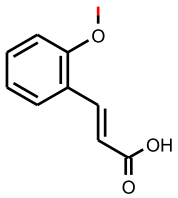
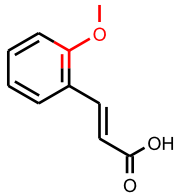
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

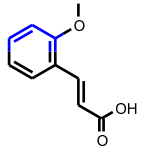
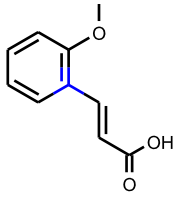
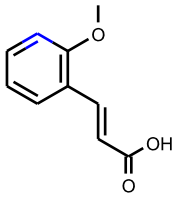
1. All properties and OPS components are within expected ranges.

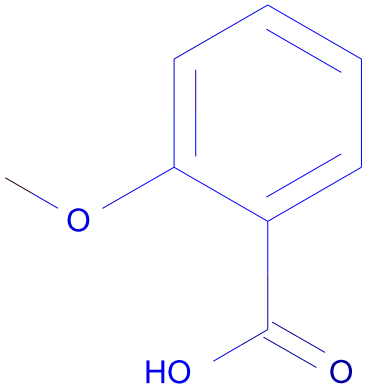
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-1925046727	 [*]C=[*]	0.145

ECFP_6	734603939	 [*]C	0.0424
ECFP_6	1307307440	 [*]:[c](:[*])OC	0.0156

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



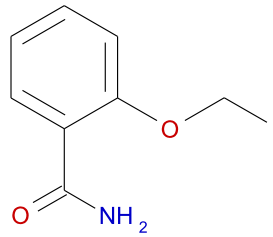
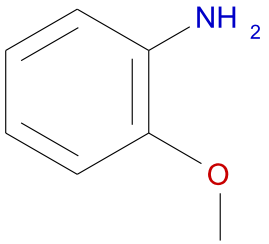
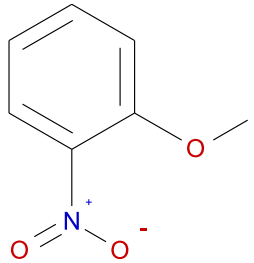
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 1.15e+003
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.04
Mahalanobis Distance p-value: 0.273

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	o-Ethoxybenzamide	153	o-Nitroanisole
Structure			
Actual Endpoint (-log C)	2.50786	2.10547	2.93466
Predicted Endpoint (-log C)	2.72346	2.33861	2.36925
Distance	0.417	0.426	0.461
Reference	CPDB	CPDB	CPDB

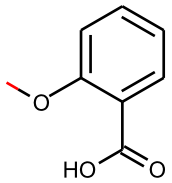
Model Applicability

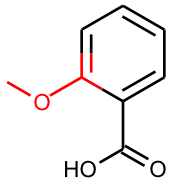
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

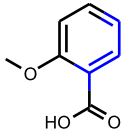
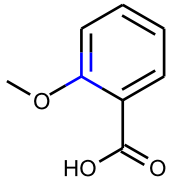
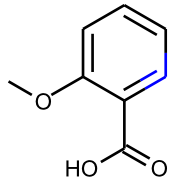
Feature Contribution

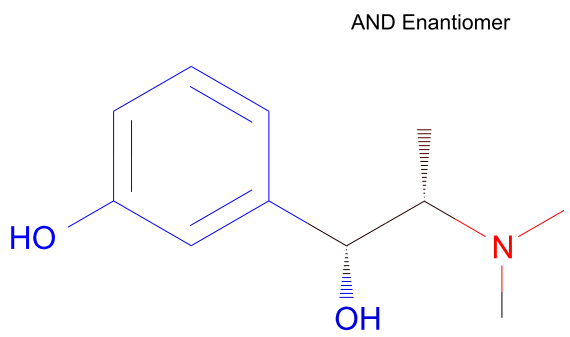
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	734603939	 [*]C	0.0424

ECFP_6	1307307440	 [*]:[c](:[*])OC	0.0156

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



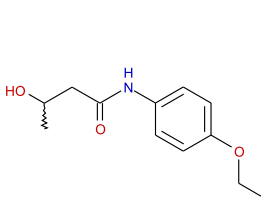
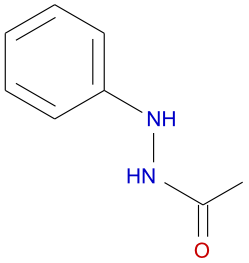
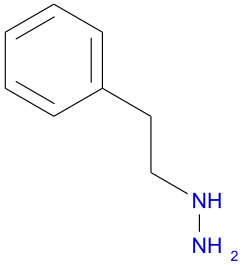
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 145
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.5
Mahalanobis Distance p-value: 0.000229

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3-Hydroxy-p-butyrophenetidide	1-Acetyl-2-phenylhydrazine	529
Structure			
Actual Endpoint (-log C)	1.6061	3.46734	3.96981
Predicted Endpoint (-log C)	2.97088	2.52008	3.43178
Distance	0.559	0.562	0.571
Reference	CPDB	CPDB	CPDB

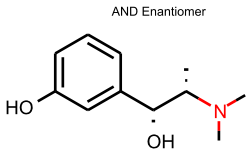
Model Applicability

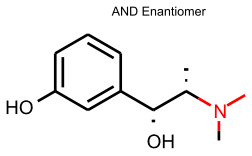
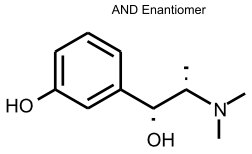
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -222940837: [*]C([*])[C@H](O)[c](:[*]):[*]
3. Unknown ECFP_2 feature: -1280034494: [*]C([*])[C@H](C)N([*])[*]
4. Unknown ECFP_2 feature: 1064495017: [*]C([*])N(C)C

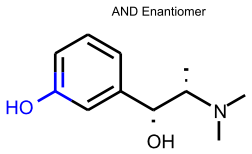
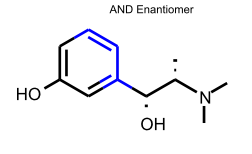
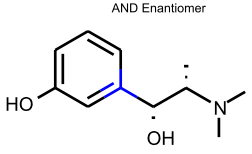
Feature Contribution

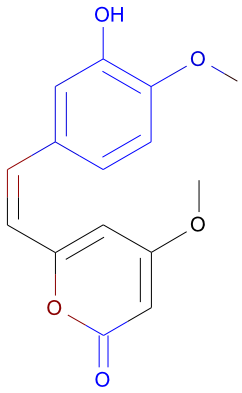
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-1072294614	 [*]N([*])[*]	0.428

ECFP_6	865379614	 [*]N([*])C	0.219
ECFP_6	734603939	 [*]C	0.0424

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2019062761	 [*]:[c](:[*])O	-0.258
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247



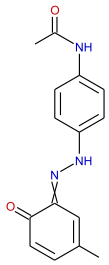
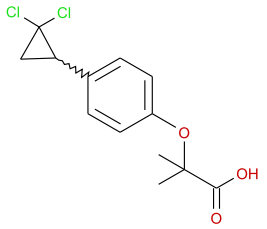
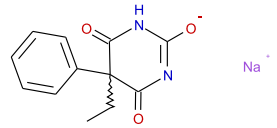
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: 588
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.03
Mahalanobis Distance p-value: 0.276

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	C.I. disperse yellow 3	Ciprofibrate	581
Structure			
Actual Endpoint (-log C)	2.42163	4.66874	3.89128
Predicted Endpoint (-log C)	3.11116	3.1937	3.31105
Distance	0.613	0.623	0.627
Reference	CPDB	CPDB	CPDB

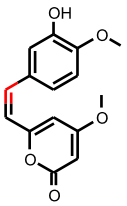
Model Applicability

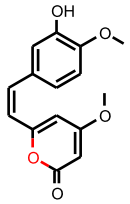
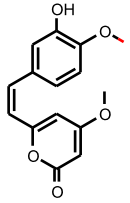
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

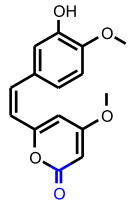
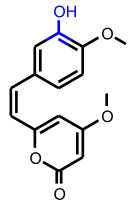
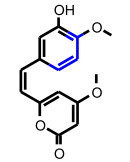
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1053980253: [*]O\C(=C/[*])\C=[*]
3. Unknown ECFP_2 feature: -444332269: [*]O\C(=C\[*])\C=[*]

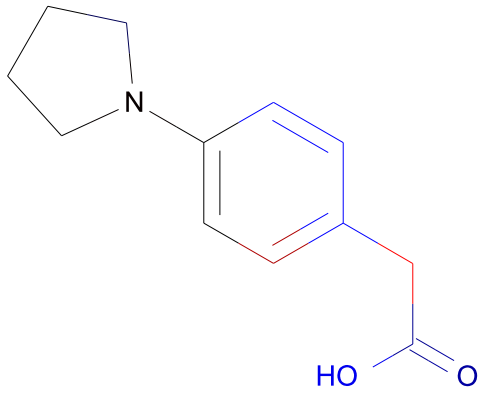
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-1925046727	 [*]C=[*]	0.145

ECFP_6	683445015	 [*]O[*]	0.136
ECFP_6	734603939	 [*]C	0.0424

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	2019062761	 [*]:[c](:[*])O	-0.258
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251



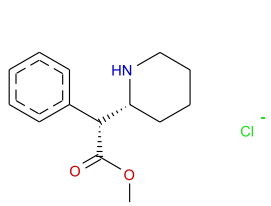
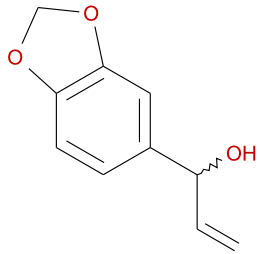
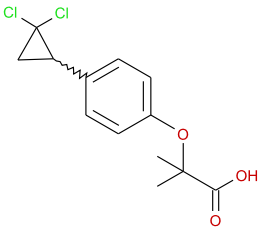
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 86.3
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.43
Mahalanobis Distance p-value: 0.135

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	666	1´-Hydroxysafrole	Ciprofibrate
Structure			
Actual Endpoint (-log C)	3.36793	3.39839	4.66874
Predicted Endpoint (-log C)	2.71924	3.24585	3.1937
Distance	0.507	0.533	0.561
Reference	CPDB	CPDB	CPDB

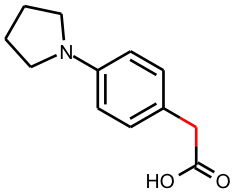
Model Applicability

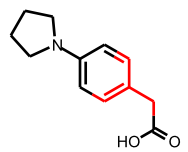
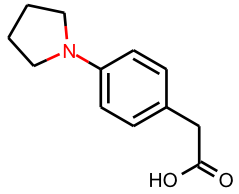
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC23 out of range. Value: 3.4508. Training min, max, SD, explained variance: -2.6901, 3.3252, 1.05, 0.0138.
2. Unknown ECFP_2 feature: 1951894094: [*]:[c](:[*])N1C[*][*]C1

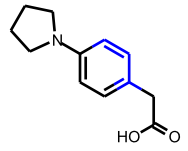
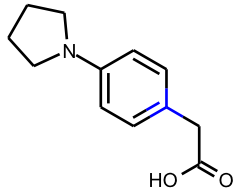
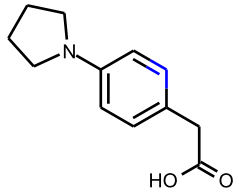
Feature Contribution

Top features for positive contribution

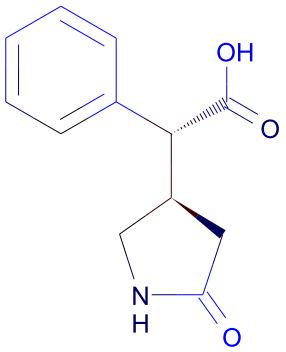
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

ECFP_6	-2024255407	 [*]C[c](:[cH]:[*]):[cH]:[*]	0.172
ECFP_6	670515721	 [*]N([*])[*]	0.00735

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232

AND Enantiomer



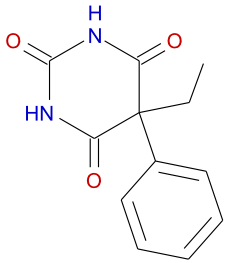
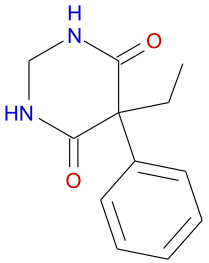
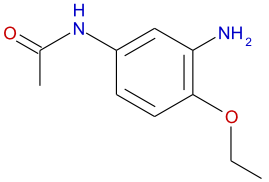
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 1.02e+003
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.84
Mahalanobis Distance p-value: 0.0543

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenobarbital s	Primidone	3-Amino-4-ethoxyacetanilide
Structure			
Actual Endpoint (-log C)	4.49846	3.92232	1.97235
Predicted Endpoint (-log C)	3.14828	3.00875	2.44449
Distance	0.499	0.505	0.576
Reference	CPDB	CPDB	CPDB

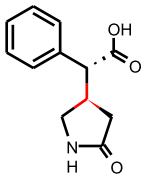
Model Applicability

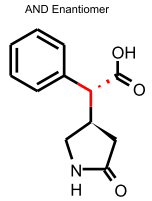
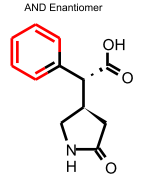
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -858846751: [*]C([*])C1C[*][*]C1
3. Unknown ECFP_2 feature: -1457159889: [*][C@H]1[*][*]NC1

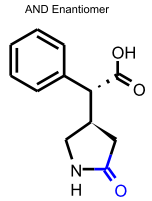
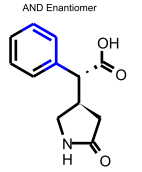
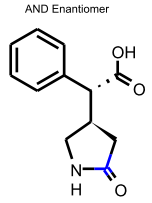
Feature Contribution

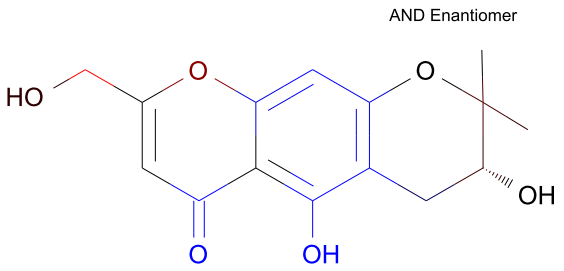
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-167460056	AND Enantiomer  [*]C([*])[*]	0.0596

ECFP_6	-1910270391	AND Enantiomer  [*]C([*])[*]	0.0279
ECFP_6	1571214559	AND Enantiomer  [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.0145

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer  [*]C(=O)[*]	-0.275
ECFP_6	1996767644	AND Enantiomer  [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	AND Enantiomer  [*][c](:[*]):[*]	-0.247



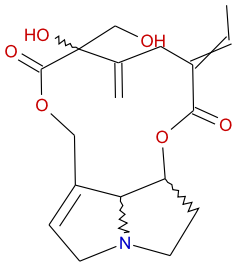
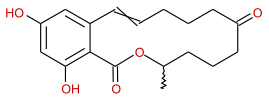
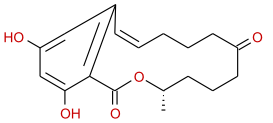
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: 31
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.95
Mahalanobis Distance p-value: 0.0405

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Riddelliine	Zearalenone	71
Structure			
Actual Endpoint (-log C)	5.24883	3.91186	3.91186
Predicted Endpoint (-log C)	4.57392	3.54452	3.54452
Distance	0.672	0.706	0.706
Reference	CPDB	CPDB	CPDB

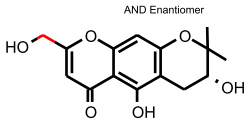
Model Applicability

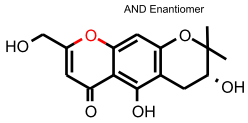
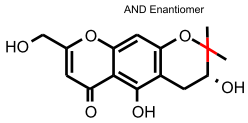
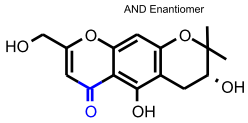
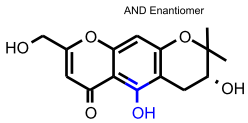
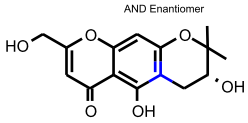
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

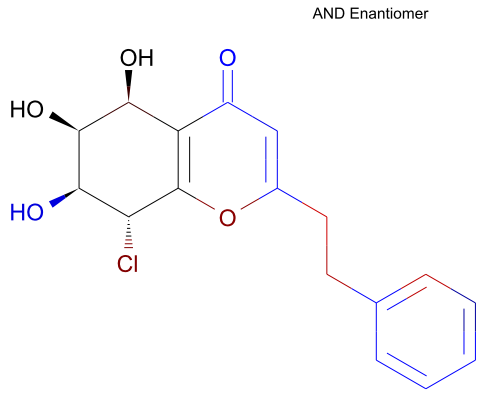
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1778376725: [*]OC(C)(C)C([*])[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

ECFP_6	683445015	 [*]O[*]	0.136
ECFP_6	657586427	 [*]C([*])([*])[*]	0.0789
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	2019062761	 [*]:[c](:[*])O	-0.258
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247



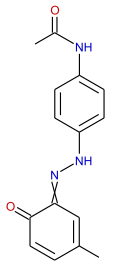
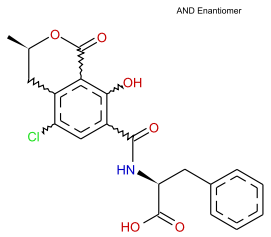
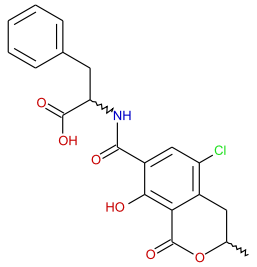
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: 30.3
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.000788

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	C.I. disperse yellow 3	542	Ochratoxin A
Structure			
Actual Endpoint (-log C)	2.42163	4.79932	4.79932
Predicted Endpoint (-log C)	3.11116	3.6353	3.6353
Distance	0.695	0.698	0.698
Reference	CPDB	CPDB	CPDB

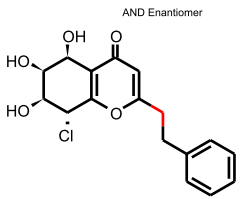
Model Applicability

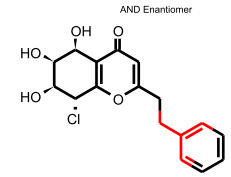
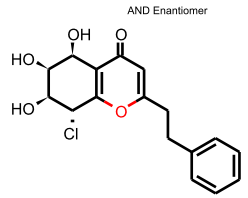
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1795792463: [*]C([*])C(=C([*])[*])C(=[*])[*]
3. Unknown ECFP_2 feature: -1567907747: [*]C([*])[C@@H](O)C(=[*])[*]
4. Unknown ECFP_2 feature: -7106223: [*]C([*])[C@@H](Cl)C(=[*])[*]
5. Unknown ECFP_2 feature: 1652274794: [*]OC(=C([*])[*])C([*])[*]

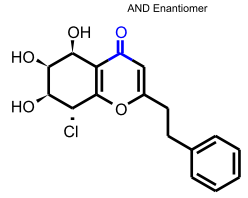
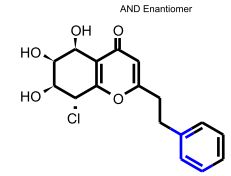
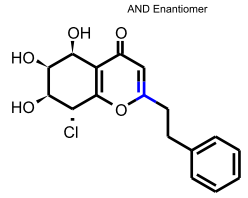
Feature Contribution

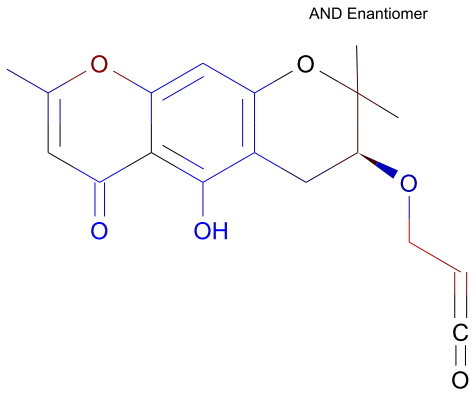
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

ECFP_6	-2024255407	 [*]C[c](:[cH]:[*]):[cH]:[*]	0.172
ECFP_6	683445015	 [*]O[*]	0.136

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247



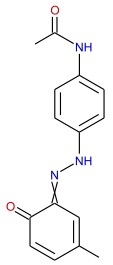
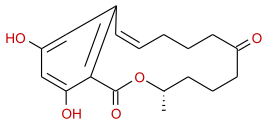
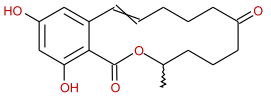
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 77.4
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.4
Mahalanobis Distance p-value: 0.000344

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	C.I. disperse yellow 3	71	Zearalenone
Structure			
Actual Endpoint (-log C)	2.42163	3.91186	3.91186
Predicted Endpoint (-log C)	3.11116	3.54452	3.54452
Distance	0.622	0.663	0.663
Reference	CPDB	CPDB	CPDB

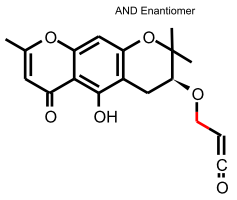
Model Applicability

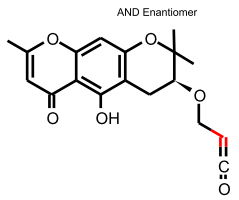
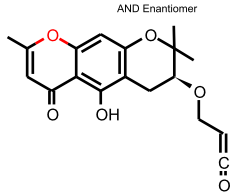
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -331149802: [*]O\C(=C/[*])\C
3. Unknown ECFP_2 feature: 1778376725: [*]OC(C)(C)C([*])[*]
4. Unknown ECFP_2 feature: -1591590376: [*]C=C=O
5. Unknown ECFP_2 feature: -91536905: [*]CC=C=[*]
6. Unknown ECFP_2 feature: -1250019913: [*]COC([*])[*]

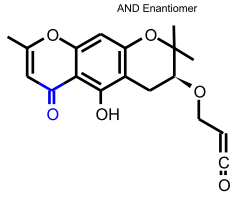
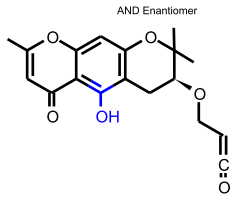
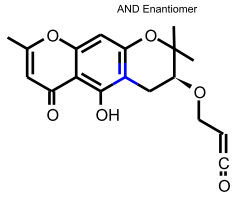
Feature Contribution

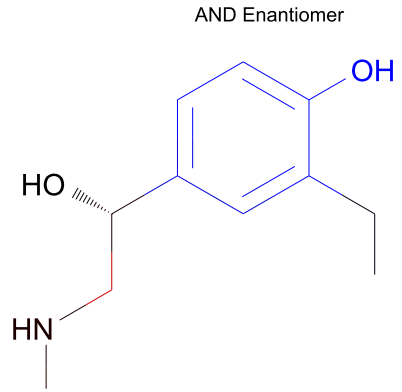
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

ECFP_6	-1925046727	 [*]C=[*]	0.145
ECFP_6	683445015	 [*]O[*]	0.136

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	2019062761	 [*]:[c](:[*])O	-0.258
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247



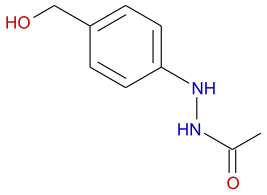
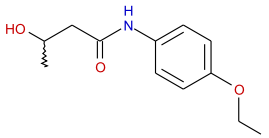
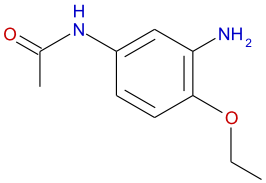
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 228
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 8.84
Mahalanobis Distance p-value: 0.361

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	N´-Acetyl-4-(hydroxy-methyl) phenyl-hydrazine	3-Hydroxy-p-butyrophenetidide	3-Amino-4-ethoxyacetanilide
Structure			
Actual Endpoint (-log C)	2.87375	1.6061	1.97235
Predicted Endpoint (-log C)	3.15825	2.97088	2.44449
Distance	0.546	0.565	0.598
Reference	CPDB	CPDB	CPDB

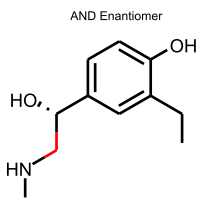
Model Applicability

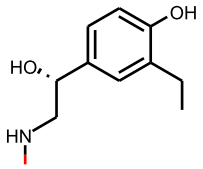
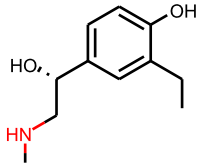
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1537923023: [*]C[C@H](O)[c](:[*]):[*]
3. Unknown ECFP_2 feature: -652986225: [*]NCC([*])[*]
4. Unknown ECFP_2 feature: 493154328: [*]CNC

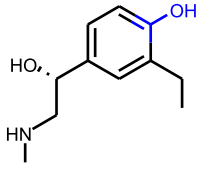
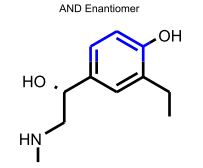
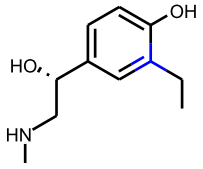
Feature Contribution

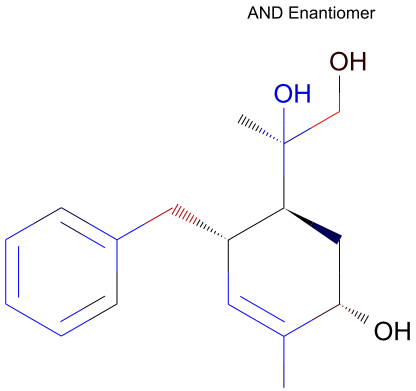
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

ECFP_6	734603939	AND Enantiomer  [*]C	0.0424
ECFP_6	-1897341097	AND Enantiomer  [*]N[*]	0.0284

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2019062761	AND Enantiomer  [*]:[c](:[*])O	-0.258
ECFP_6	1996767644	AND Enantiomer  [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	AND Enantiomer  [*][c](:[*]):[*]	-0.247



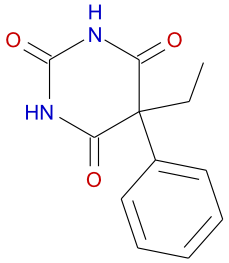
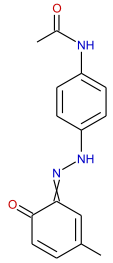
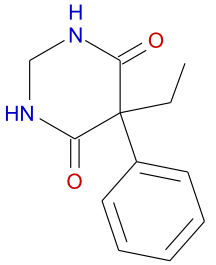
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 129
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.00175

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenobarbital s	C.I. disperse yellow 3	Primidone
Structure			
Actual Endpoint (-log C)	4.49846	2.42163	3.92232
Predicted Endpoint (-log C)	3.14828	3.11116	3.00875
Distance	0.619	0.626	0.640
Reference	CPDB	CPDB	CPDB

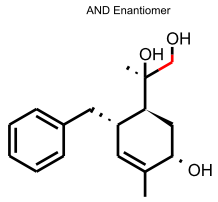
Model Applicability

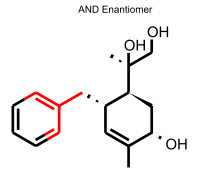
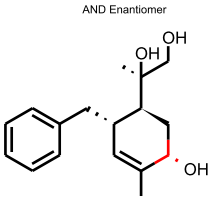
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC16 out of range. Value: -3.1092. Training min, max, SD, explained variance: -3.1026, 4.016, 1.245, 0.0193.
2. Unknown ECFP_2 feature: -2097867909: [*]C[C@H](O)C(=[*])[*]
3. Unknown ECFP_2 feature: -327548242: [*]C[C@@H](C([*])[*])C([*])([*])[*]
4. Unknown ECFP_2 feature: -1263967621: [*]C[C@@H](C=[*])C([*])[*]
5. Unknown ECFP_2 feature: 1280892564: [*]C[C@](C)(O)C([*])[*]
6. Unknown ECFP_2 feature: 771121623: [*]C([*])C[c](:[*]):[*]

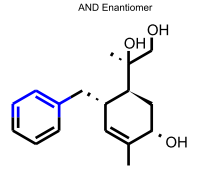
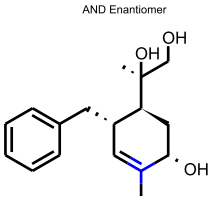
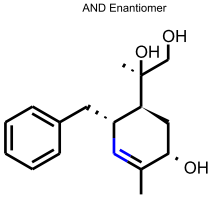
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

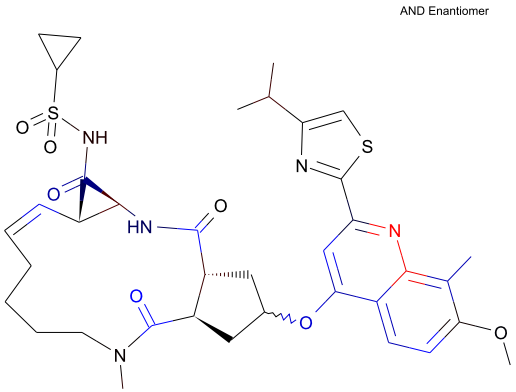
ECFP_6	-2024255407	 [*]C[c](:[cH]:[*]):[cH]:[*]	0.172
ECFP_6	-167460056	 [*]C([*])[*]	0.0596

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232

Simeprevir

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: 2.01

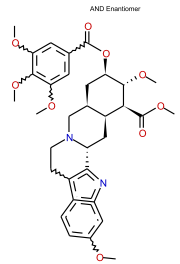
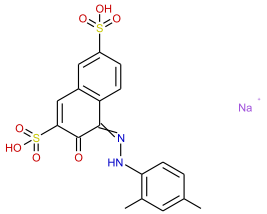
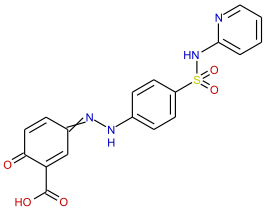
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 16.8

Mahalanobis Distance p-value: 5.81e-019

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	223	D & C red no. 5	Salicylazosulfapyridine
Structure			
Actual Endpoint (-log C)	5.08368	2.80732	2.5034
Predicted Endpoint (-log C)	5.08273	3.78615	3.54214
Distance	0.926	1.173	1.177
Reference	CPDB	CPDB	CPDB

Model Applicability

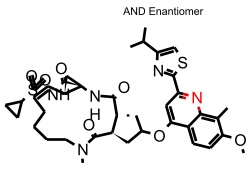
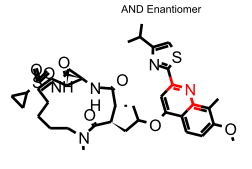
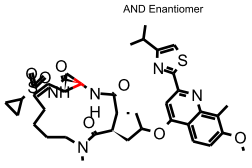
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: -4.1139. Training min, max, SD, explained variance: -3.6133, 3.7483, 1.312, 0.0215.
2. OPS PC16 out of range. Value: 4.4468. Training min, max, SD, explained variance: -3.1026, 4.016, 1.245, 0.0193.
3. Unknown ECFP_2 feature: -309676563: [*]N[C@]1(C[C@H]1[*])C(=[*])[*]
4. Unknown ECFP_2 feature: -1049290660: [*]C1([*])C[C@H]1C=[*]
5. Unknown ECFP_2 feature: -1818486371: [*]NC(=O)C1([*])[*][*]1
6. Unknown ECFP_2 feature: 946167604: [*]C(=[*])NS(=[*])(=[*])[*]
7. Unknown ECFP_2 feature: 1310940530: [*]C([*])O[c](:[*]):[*]
8. Unknown ECFP_2 feature: 1411720546: [*]C([*])[c]1:[cH]:[*]:[*]:n:1
9. Unknown ECFP_2 feature: 733491677: [*]:[c](:[*])C(C)C
10. Unknown ECFP_2 feature: 2121941848: [*]NS(=O)(=O)C1[*][*]1
11. Unknown ECFP_2 feature: -622223421: [*]S(=[*])(=[*])C1CC1

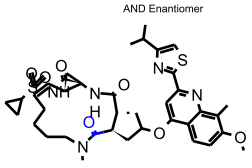
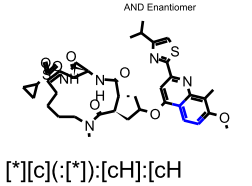
Feature Contribution

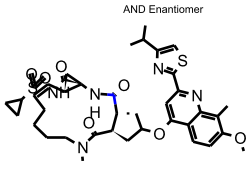
Top features for positive contribution

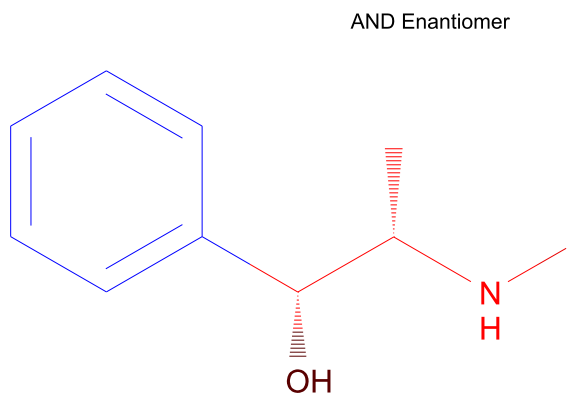
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]:n:[*]	0.229
ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	657586427	 [*]C([*])([*])[*]	0.0789

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251

ECFP_6	642810091	AND Enantiomer  [*][c](:[*]):[*]	-0.247



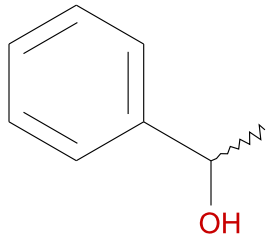
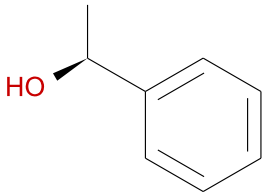
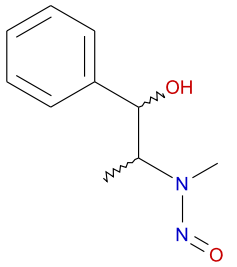
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 10.3
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.3
Mahalanobis Distance p-value: 0.524

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	a-Methylbenzyl alcohol	665	N-Nitrosoephedrine
Structure			
Actual Endpoint (-log C)	2.42608	2.42608	3.30968
Predicted Endpoint (-log C)	2.72332	2.72332	4.9805
Distance	0.419	0.419	0.490
Reference	CPDB	CPDB	CPDB

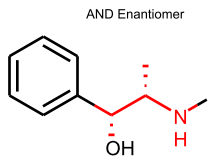
Model Applicability

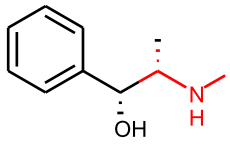
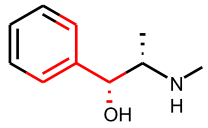
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

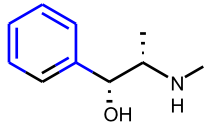
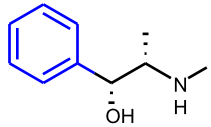
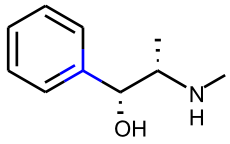
Feature Contribution

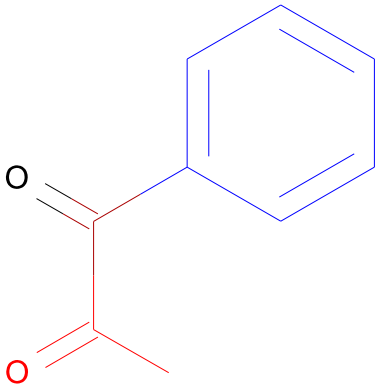
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1043250487	 [*]N[C@@H](C)C([*])[*]	1.15

FCFP_6	-885550502	AND Enantiomer  [*]C([*])NC	0.229
FCFP_6	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer  [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	-2093839777	AND Enantiomer  [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.378
FCFP_6	16	AND Enantiomer  [*][c](:[*]):[*]	-0.354



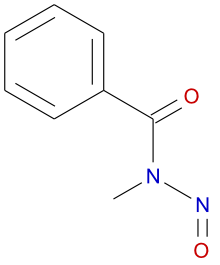
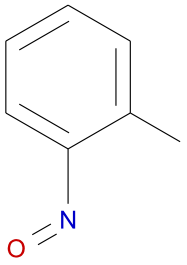
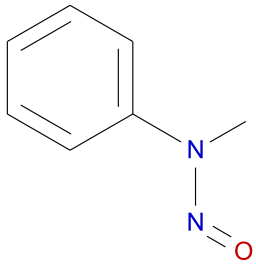
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: 174
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 7.84
Mahalanobis Distance p-value: 0.981

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	N-Methyl-N-nitrosobenzamide	o-Nitrosotoluene	Nitrosomethylaniline
Structure			
Actual Endpoint (-log C)	4.70607	3.37827	5.98173
Predicted Endpoint (-log C)	4.64533	3.14846	4.43574
Distance	0.457	0.466	0.471
Reference	CPDB	CPDB	CPDB

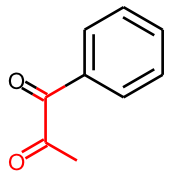
Model Applicability

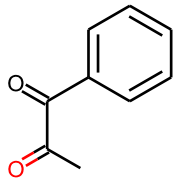
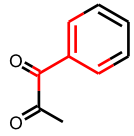
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

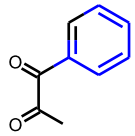
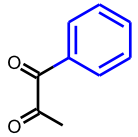
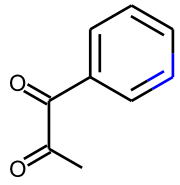
Feature Contribution

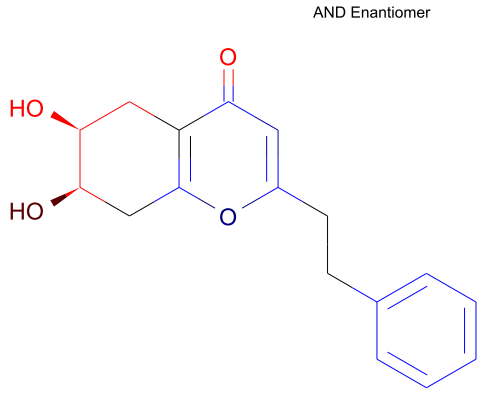
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266

FCFP_6	1	 [*]=O	0.234
FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	-2093839777	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.378
FCFP_6	16	 [*][c](:[*]):[*]	-0.354



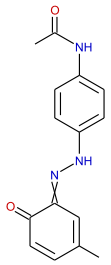
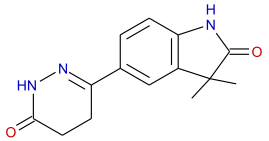
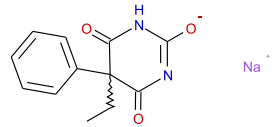
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: 16
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.01

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	C.I. disperse yellow 3	Indolidan	581
Structure			
Actual Endpoint (-log C)	2.85045	5.10722	3.42954
Predicted Endpoint (-log C)	3.2505	4.15845	3.78228
Distance	0.562	0.571	0.606
Reference	CPDB	CPDB	CPDB

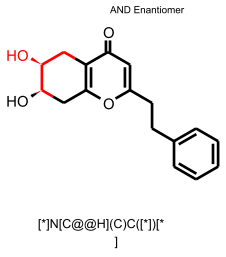
Model Applicability

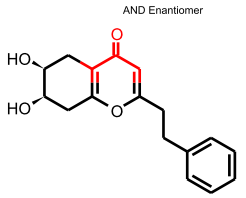
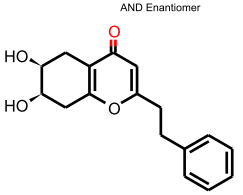
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

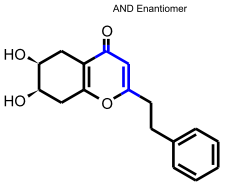
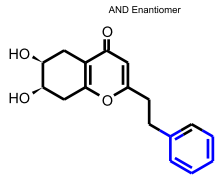
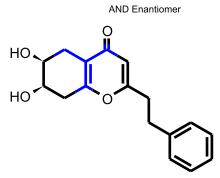
Feature Contribution

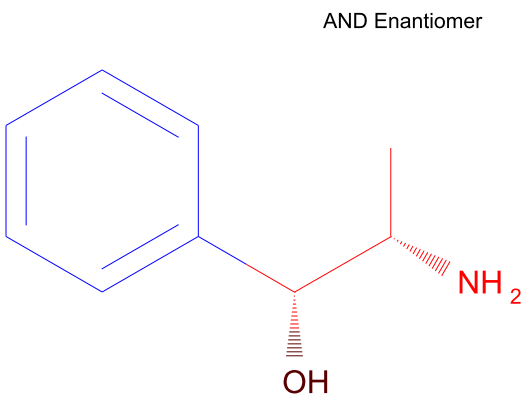
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1043250487	 [*]N[C@@H](C)C([*])[*]	1.15

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	436886043	 [*]CC(=C([*])[*])C(=[*])[*]	-0.383



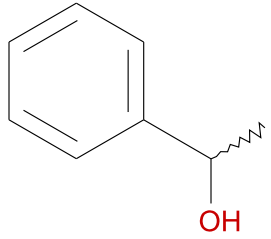
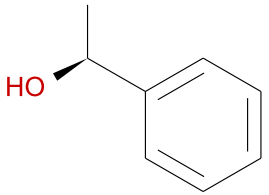
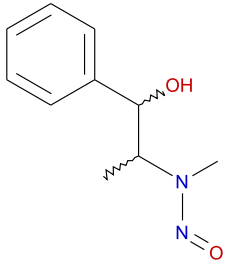
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 23.9
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 7.59
Mahalanobis Distance p-value: 0.993

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	a-Methylbenzyl alcohol	665	N-Nitrosoephedrine
Structure			
Actual Endpoint (-log C)	2.42608	2.42608	3.30968
Predicted Endpoint (-log C)	2.72332	2.72332	4.9805
Distance	0.384	0.384	0.488
Reference	CPDB	CPDB	CPDB

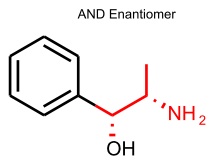
Model Applicability

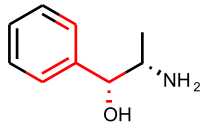
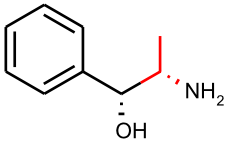
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

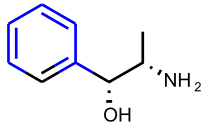
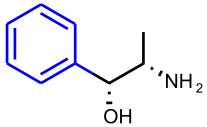
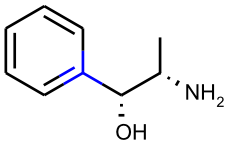
Feature Contribution

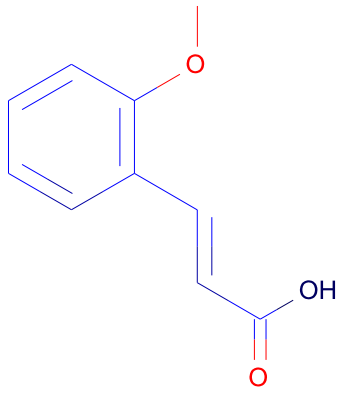
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1043250487	 [*]N[C@@H](C)C([*])[*]	1.15

FCFP_6	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137
FCFP_6	136597326	AND Enantiomer  [*]C([*])C	0.0695

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer  [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	-2093839777	AND Enantiomer  [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.378
FCFP_6	16	AND Enantiomer  [*][c](:[*]):[*]	-0.354



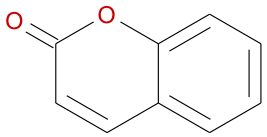
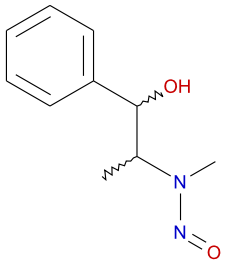
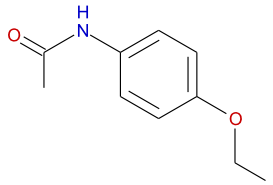
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 99.5
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 0.0352

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Coumarin s	N-Nitrosoephedrine	Phenacetin
Structure			
Actual Endpoint (-log C)	3.57149	3.30968	2.15647
Predicted Endpoint (-log C)	3.23061	4.9805	2.91803
Distance	0.468	0.482	0.497
Reference	CPDB	CPDB	CPDB

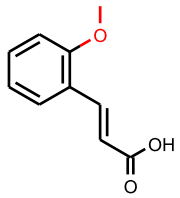
Model Applicability

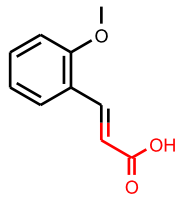
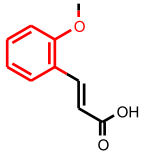
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

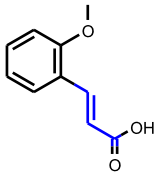
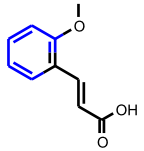
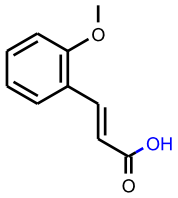
Feature Contribution

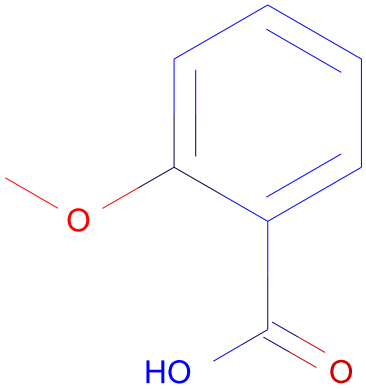
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	-1176841573	 [*]=CC(=O)O	0.277
FCFP_6	-2090462286	 [*]O[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1[*]	0.245

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	7	 [*]O	-0.372



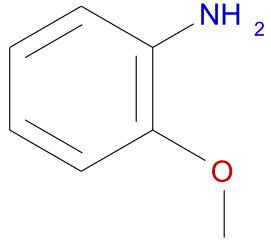
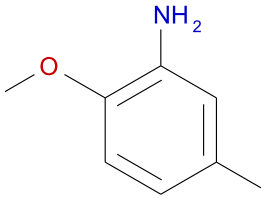
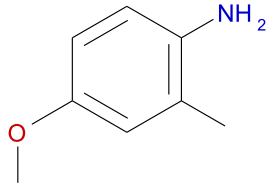
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 61.8
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12
Mahalanobis Distance p-value: 0.000451

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	153	p-Cresidine	m-Cresidine
Structure			
Actual Endpoint (-log C)	3.61769	3.14606	2.46519
Predicted Endpoint (-log C)	3.58569	3.83612	3.87814
Distance	0.433	0.441	0.458
Reference	CPDB	CPDB	CPDB

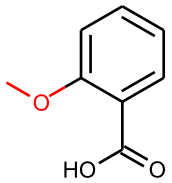
Model Applicability

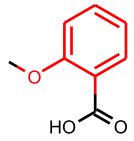
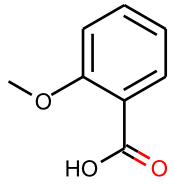
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

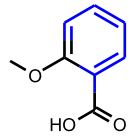
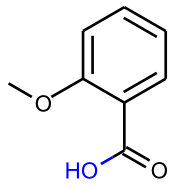
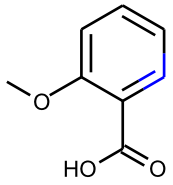
Feature Contribution

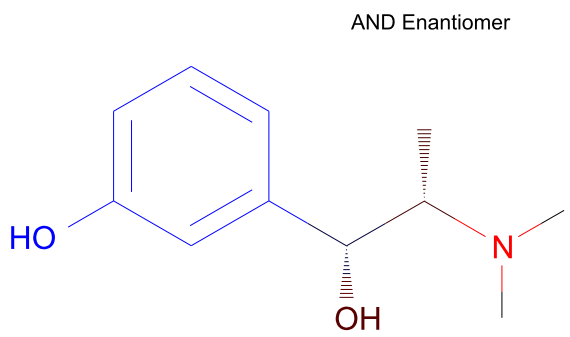
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	-2090462286	 [*]O[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1[*]	0.245
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	7	 [*]O	-0.372
FCFP_6	16	 [*][c](:[*]):[*]	-0.354



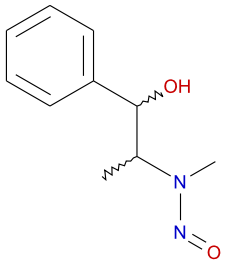
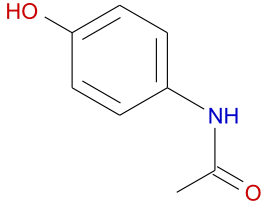
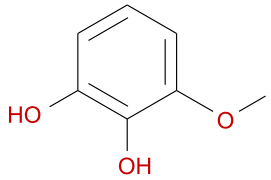
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 254
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.8
Mahalanobis Distance p-value: 1.19e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	N-Nitrosoephedrine	Acetaminophen	3-Methoxycatechol
Structure			
Actual Endpoint (-log C)	3.30968	2.48484	3.45902
Predicted Endpoint (-log C)	4.9805	2.46116	2.88664
Distance	0.508	0.521	0.533
Reference	CPDB	CPDB	CPDB

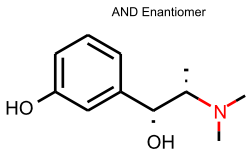
Model Applicability

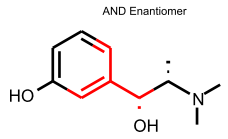
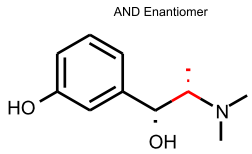
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

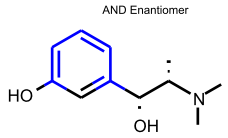
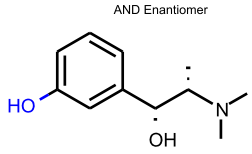
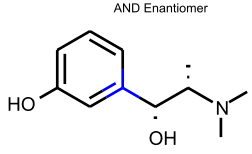
Feature Contribution

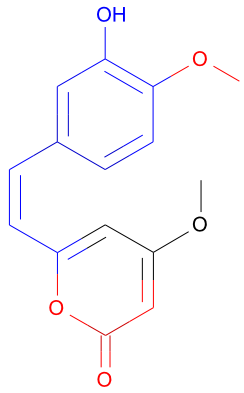
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	9	 [*]N([*])[*]	0.385

FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137
FCFP_6	136597326	 [*]C([*])C	0.0695

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	7	 [*]O	-0.372
FCFP_6	16	 [*][c](:[*]):[*]	-0.354



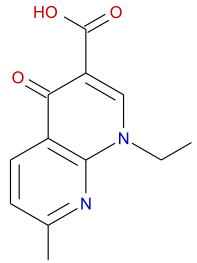
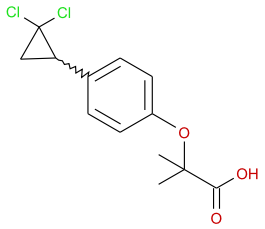
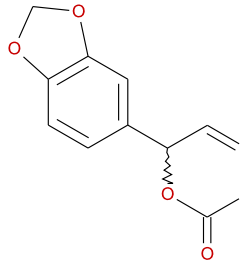
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: 65.1
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.7
Mahalanobis Distance p-value: 2.95e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Nalidixic acid	Ciprofibrate	1´-Acetoxysafrole
Structure			
Actual Endpoint (-log C)	3.06273	5.16666	3.94492
Predicted Endpoint (-log C)	3.02209	3.87886	3.37654
Distance	0.572	0.582	0.583
Reference	CPDB	CPDB	CPDB

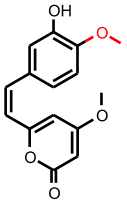
Model Applicability

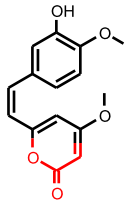
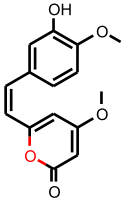
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

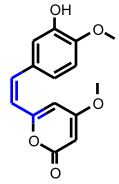
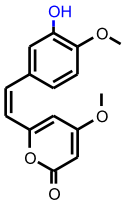
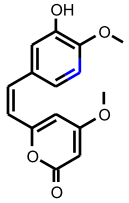
Feature Contribution

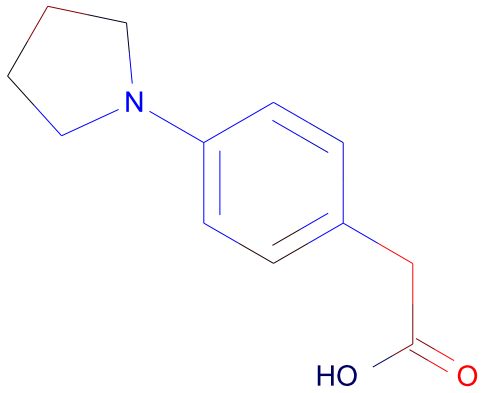
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	565998553	 [*]OC(=O)C=[*]	0.357
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	7	 [*]O	-0.372
FCFP_6	16	 [*][c](:[*]):[*]	-0.354



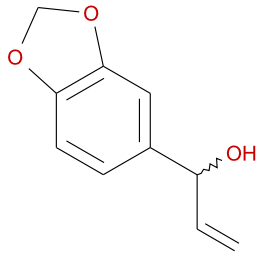
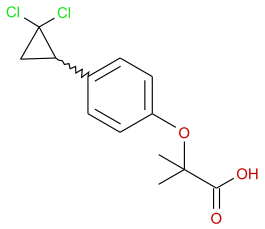
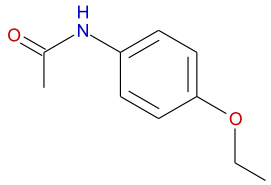
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 54.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 0.00177

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	1´-Hydroxysafrole	Ciprofibrate	Phenacetin
Structure			
Actual Endpoint (-log C)	3.98605	5.16666	2.15647
Predicted Endpoint (-log C)	3.25547	3.87886	2.91803
Distance	0.495	0.496	0.552
Reference	CPDB	CPDB	CPDB

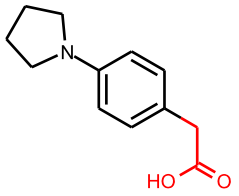
Model Applicability

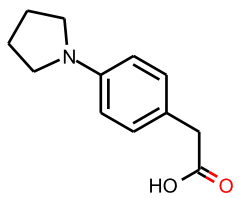
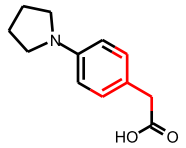
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

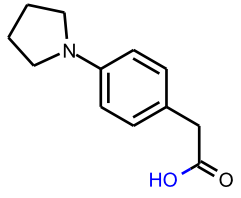
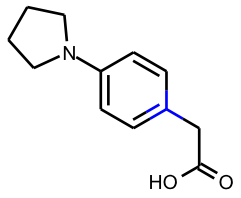
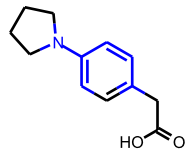
Feature Contribution

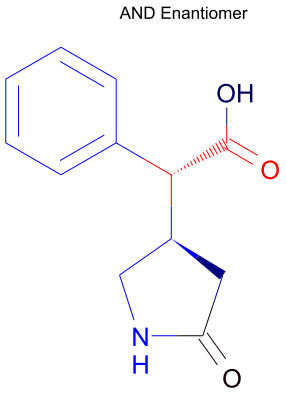
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1176841573	 [*]=CC(=O)O	0.277

FCFP_6	1	 [*]=O	0.234
FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	7	 [*]O	-0.372
FCFP_6	16	 [*][c](:[*]):[*]	-0.354
FCFP_6	1674451008	 [*]O[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-0.233



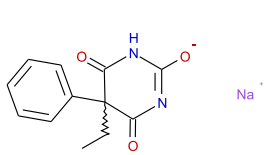
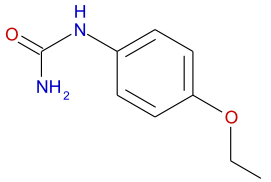
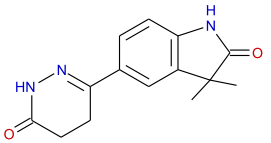
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 570
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 0.00208

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	581	4-Ethoxy-phenylurea	Indolidan
Structure			
Actual Endpoint (-log C)	3.42954	2.52579	5.10722
Predicted Endpoint (-log C)	3.78228	3.02654	4.15845
Distance	0.531	0.578	0.580
Reference	CPDB	CPDB	CPDB

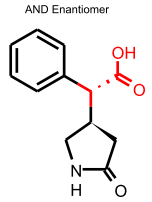
Model Applicability

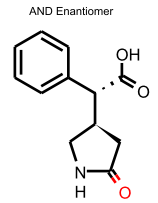
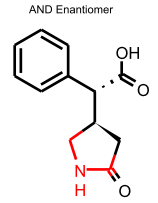
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC27 out of range. Value: -2.9933. Training min, max, SD, explained variance: -2.8642, 4.2058, 1.002, 0.0096.

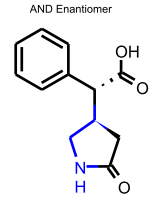
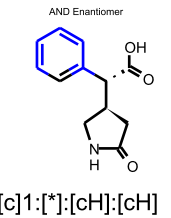
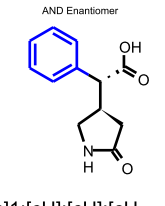
Feature Contribution

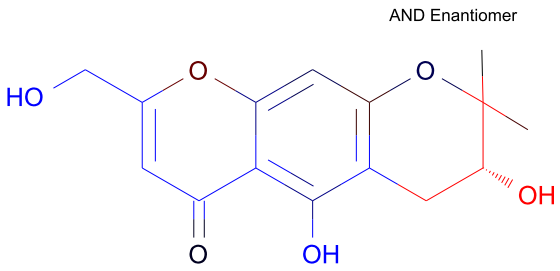
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1176841573	 [*]=CC(=O)O	0.277

FCFP_6	1	 [*]=O	0.234
FCFP_6	-885550502	 [*]C([*])NC	0.229

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1272709286	 [*][C@H]1[*][*]NC1	-0.526
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	-2093839777	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.378



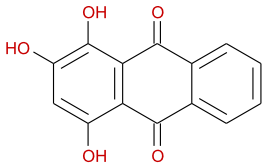
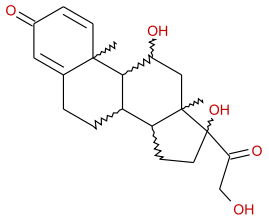
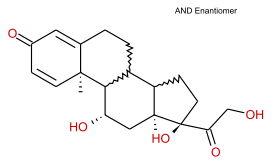
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: 20.3
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15.2
Mahalanobis Distance p-value: 1.18e-011

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Purpurin	Prednisolone	1085
Structure			
Actual Endpoint (-log C)	2.57737	5.37215	5.37215
Predicted Endpoint (-log C)	3.49183	5.32932	5.32932
Distance	0.613	0.662	0.662
Reference	CPDB	CPDB	CPDB

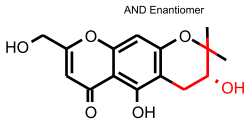
Model Applicability

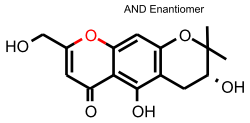
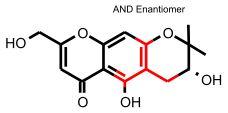
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

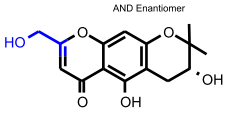
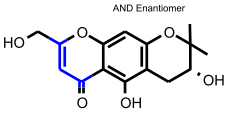
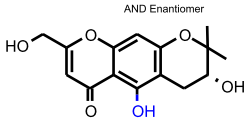
Feature Contribution

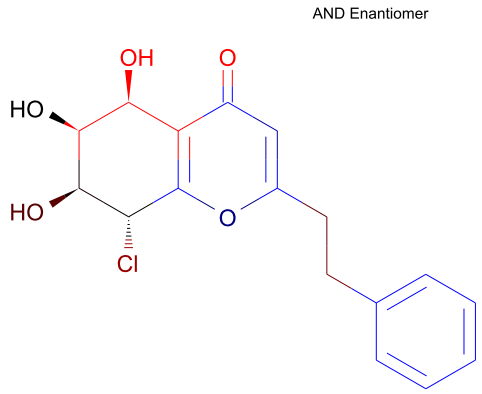
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1043250487	 [*]N[C@@H](C)C([*])[*]	1.15

FCFP_6	1	 [*]=O	0.234
FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1272709286	 [*][C@H]1[*][*]NC1	-0.526
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	7	 [*]O	-0.372



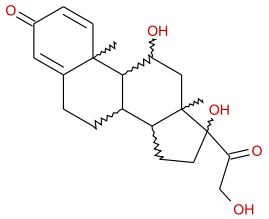
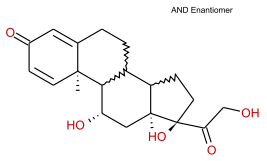
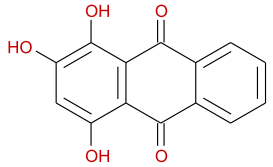
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: 10.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.8
Mahalanobis Distance p-value: 1.47e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Prednisolone	1085	Purpurin
Structure			
Actual Endpoint (-log C)	5.37215	5.37215	2.57737
Predicted Endpoint (-log C)	5.32932	5.32932	3.49183
Distance	0.625	0.625	0.652
Reference	CPDB	CPDB	CPDB

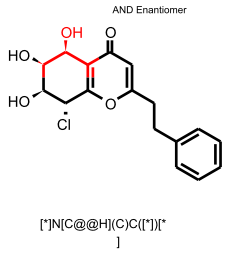
Model Applicability

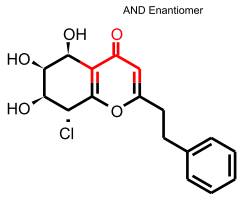
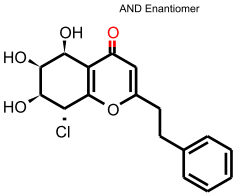
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

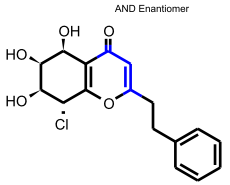
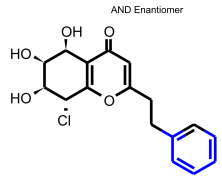
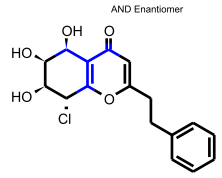
Feature Contribution

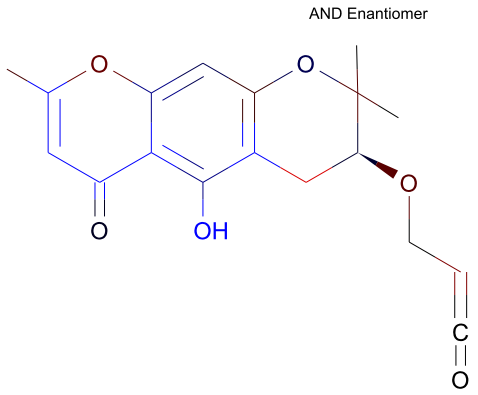
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1043250487	 [*]N[C@@H](C)C([*])[*]	1.15

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	436886043	 [*]CC(=C([*])[*])C(=[*])[*]	-0.383



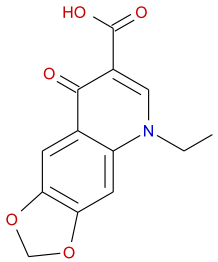
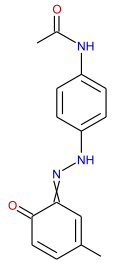
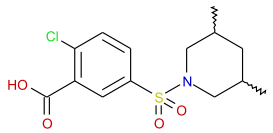
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 103
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.5
Mahalanobis Distance p-value: 9.32e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Oxolinic acid	C.I. disperse yellow 3	2-Chloro-5-(3,5-dimethylpiperi-di-nosulphonyl)benzoic acid
Structure			
Actual Endpoint (-log C)	3.19431	2.85045	4.83515
Predicted Endpoint (-log C)	3.75332	3.2505	4.5657
Distance	0.604	0.616	0.640
Reference	CPDB	CPDB	CPDB

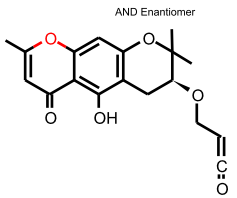
Model Applicability

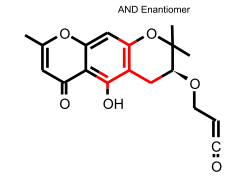
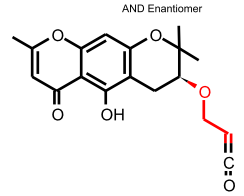
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 444624378: [*]C=C=O

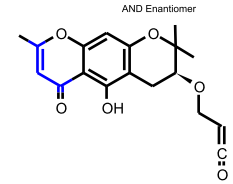
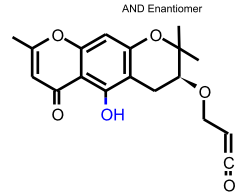
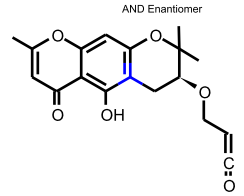
Feature Contribution

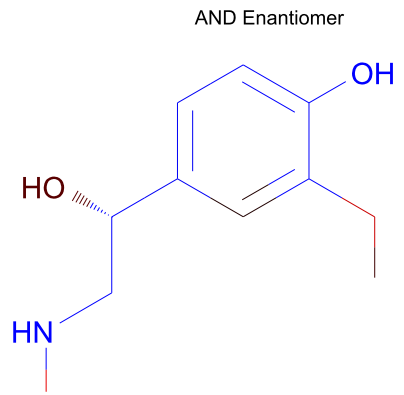
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	0.234

FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*])):[cH]:[*]	0.137
FCFP_6	-1272768868	 [*]N1[*][*]CC1	0.127

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	7	 [*]O	-0.372
FCFP_6	16	 [*][c](:[*]):[*]	-0.354



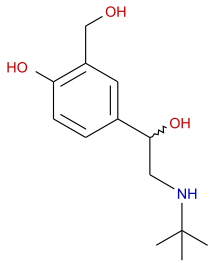
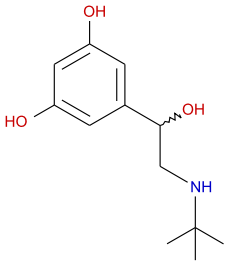
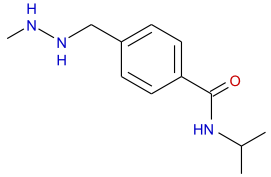
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 436
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 0.031

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Salbutamol	Terbutaline	Procarbazine
Structure			
Actual Endpoint (-log C)	3.7769	2.73995	4.74183
Predicted Endpoint (-log C)	2.66345	2.96169	5.26514
Distance	0.408	0.476	0.513
Reference	CPDB	CPDB	CPDB

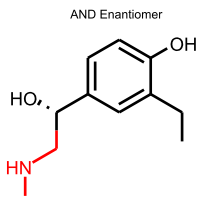
Model Applicability

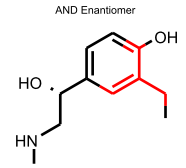
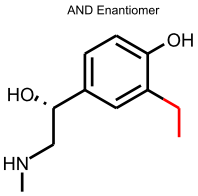
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

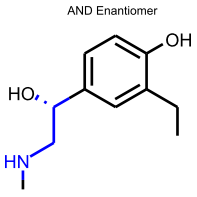
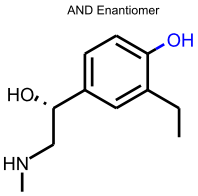
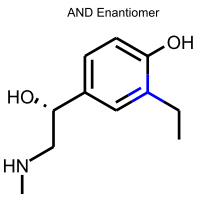
Feature Contribution

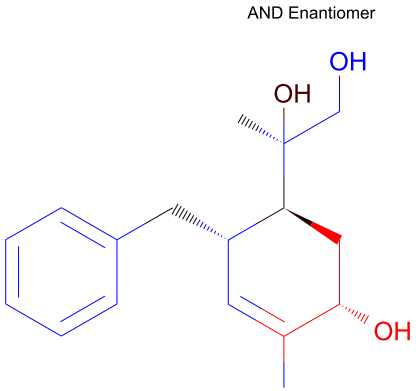
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-885550502	 [*]C([*])NC	0.229

FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137
FCFP_6	136597326	 [*]C([*])C	0.0695

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1272709286	 [*][C@H]1[*][*]NC1	-0.526
FCFP_6	7	 [*]O	-0.372
FCFP_6	16	 [*][c](:[*]):[*]	-0.354



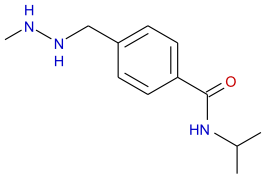
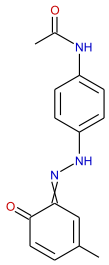
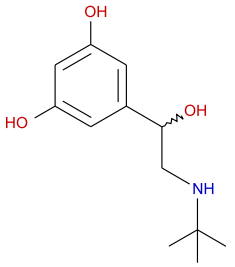
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 220
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.8
Mahalanobis Distance p-value: 1.69e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Procarbazine	C.I. disperse yellow 3	Terbutaline
Structure			
Actual Endpoint (-log C)	4.74183	2.85045	2.73995
Predicted Endpoint (-log C)	5.26514	3.2505	2.96169
Distance	0.618	0.619	0.623
Reference	CPDB	CPDB	CPDB

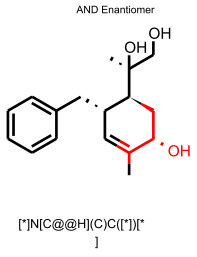
Model Applicability

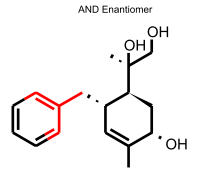
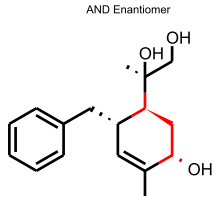
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

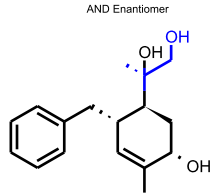
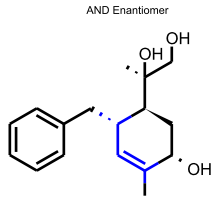
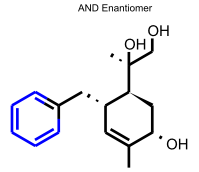
Feature Contribution

Top features for positive contribution

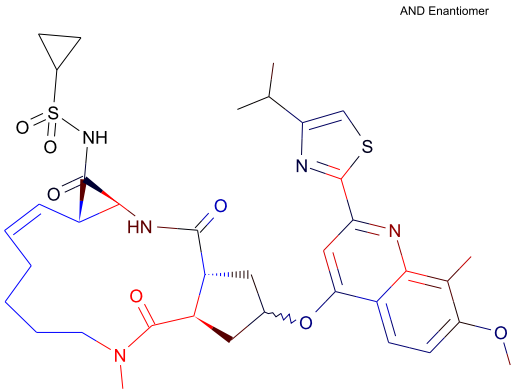
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1043250487	 [*]N[C@@H](C)C([*])([*]	1.15

FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137
FCFP_6	-1272798659	 [*]C([*])CC(=[*])[*]	0.11

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1272709286	 [*][C@H]1[*][*]NC1	-0.526
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422

Simeprevir



$C_{38}H_{47}N_5O_7S_2$
Molecular Weight: 749.93908
ALogP: 4.796
Rotatable Bonds: 8
Acceptors: 9
Donors: 2

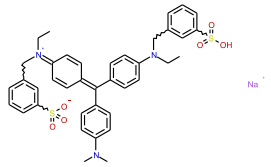
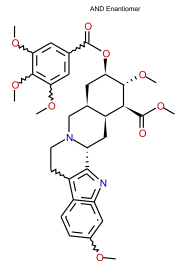
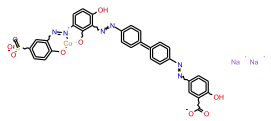
Model Prediction

Prediction: 0.28
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 24.1
Mahalanobis Distance p-value: 1.44e-046

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

TOPKAT_Carcinogenic_Potency_TD50_Rat

Structural Similar Compounds

Name	411	223	188
Structure			
Actual Endpoint (-log C)	3.06566	6.29867	5.5378
Predicted Endpoint (-log C)	4.8672	7.5657	5.71925
Distance	0.846	0.861	0.891
Reference	CPDB	CPDB	CPDB

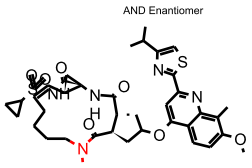
Model Applicability

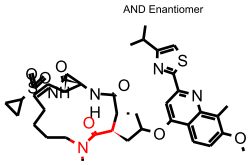
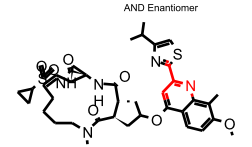
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC27 out of range. Value: -3.3094. Training min, max, SD, explained variance: -2.8642, 4.2058, 1.002, 0.0096.

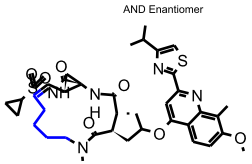
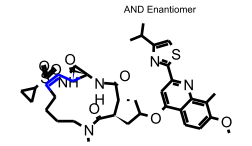
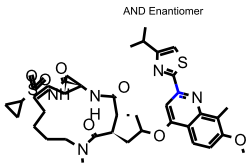
Feature Contribution

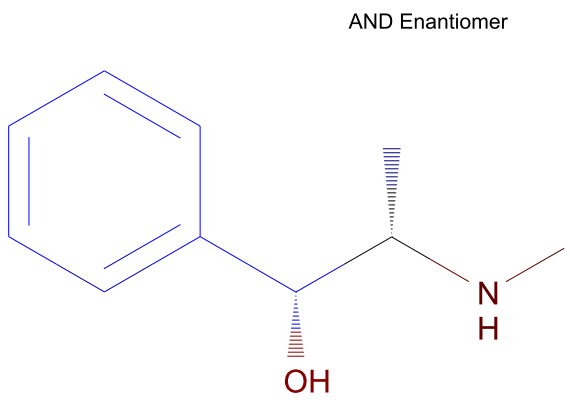
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	565998553	 [*]OC(=O)C=[*]	0.357
FCFP_6	690511177	 [*]:[cH]:[c](:n:[*])[c](:[*]):[*]	0.293

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1175638033	 [*]CCCCC=[*]	-0.512
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	16	 [*][c](:[*]):[*]	-0.354



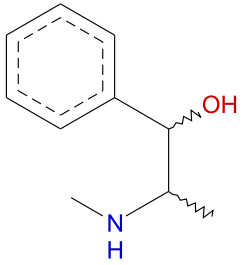
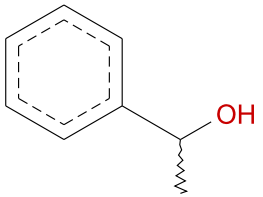
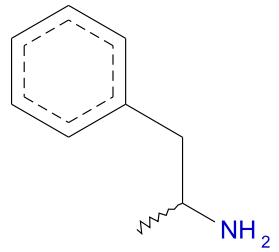
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 0.109
Unit: g/kg_body_weight
Mahalanobis Distance: 15
Mahalanobis Distance p-value: 0.161

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE SULPHATE	ALPHA-METHYLBENZYL ALCOHOL	DL-AMPHETAMINE SULFATE
Structure			
Actual Endpoint (-log C)	4.42222	2.65882	4.43204
Predicted Endpoint (-log C)	3.18027	2.85	3.16317
Distance	0.000	0.456	0.496
Reference	NTP 307 41	NTP REPORT # 369	NTP 387 58

Model Applicability

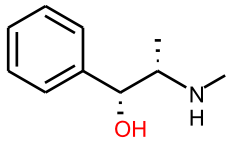
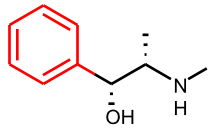
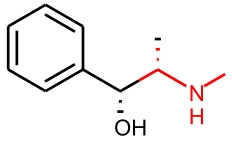
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 2023785560: [*]C([*])O
3. Unknown ECFP_6 feature: -222940837: [*]C([*])[C@H](O)[c](:[*]):[*]
4. Unknown ECFP_6 feature: -871548073: [*]N[C@@H](C)C([*])[*]
5. Unknown ECFP_6 feature: 1336634178: [*]C([*])NC
6. Unknown ECFP_6 feature: -176846085: [*]C([*])[c](:[cH]:[*]):[cH]:[*]
7. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
8. Unknown ECFP_6 feature: 864287155: [*]NC

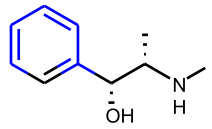
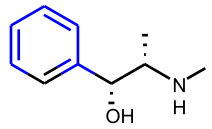
Feature Contribution

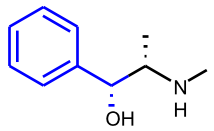
Top features for positive contribution

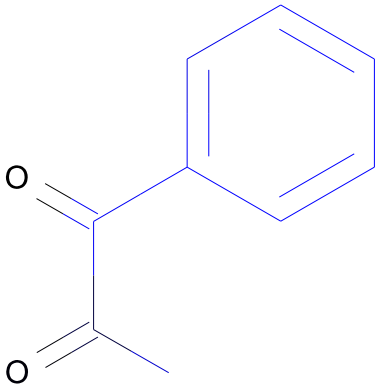
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	3	AND Enantiomer  [*]O	0.0924
FCFP_6	-2093839777	AND Enantiomer  [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.078
FCFP_6	-885550502	AND Enantiomer  [*]C([*])NC	0.0684

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer  [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	AND Enantiomer  [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133

FCFP_6	-1698724694	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0944



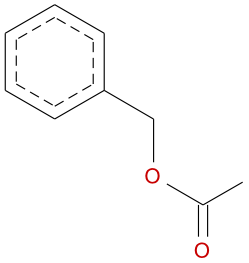
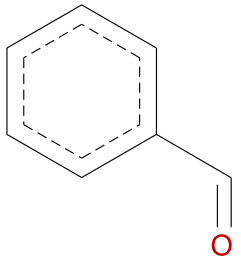
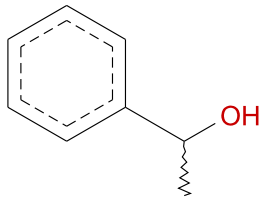
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: 0.587
Unit: g/kg_body_weight
Mahalanobis Distance: 14.6
Mahalanobis Distance p-value: 0.257

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	BENZYL ACETATE	BENZALDEHYDE	ALPHA-METHYLBENZYL ALCOHOL
Structure			
Actual Endpoint (-log C)	2.62394	2.87048	2.65882
Predicted Endpoint (-log C)	2.60518	2.6326	2.85
Distance	0.397	0.441	0.485
Reference	NTP REPORT # 250	NTP REPORT # 378	NTP REPORT # 369

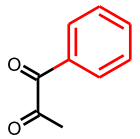
Model Applicability

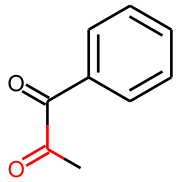
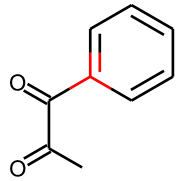
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: 1432101658: [*]C(=[*])C(=O)[c](:[*]):[*]
5. Unknown ECFP_6 feature: -472613004: [*]C(=[*])C(=O)C

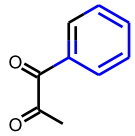
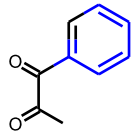
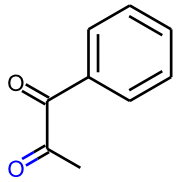
Feature Contribution

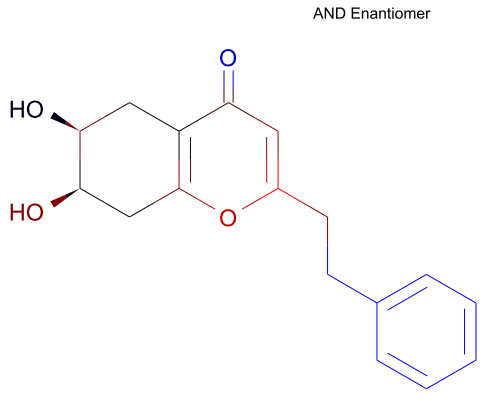
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-2093839777	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.078

ECFP_6	2099970318	 [*]C(=O)[*]	0.0766
ECFP_6	642810091	 [*][c](:[*]):[*]	0.0424

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134
ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133
FCFP_6	1	 [*]=O	-0.102



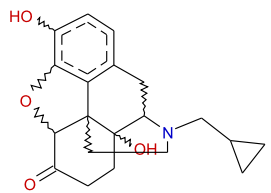
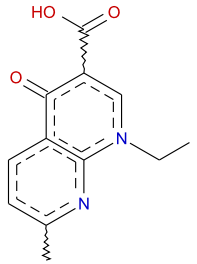
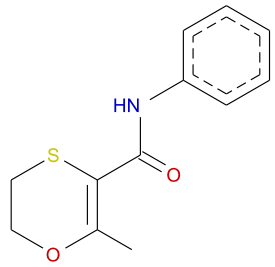
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: 0.0229
Unit: g/kg_body_weight
Mahalanobis Distance: 25.8
Mahalanobis Distance p-value: 6.13e-017

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	NALTREXONE.HCL	NALIDIXIC ACID	CARBOXIN
Structure			
Actual Endpoint (-log C)	4.05615	3.36594	3.8945
Predicted Endpoint (-log C)	4.77886	4.24773	3.7395
Distance	0.524	0.595	0.602
Reference	NDA-18932	NTP REPORT # 368	EPA COVER SHEET 0022;890701;(1)

Model Applicability

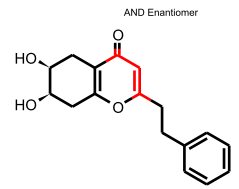
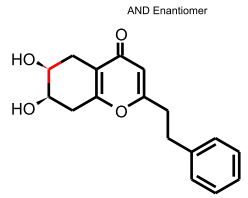
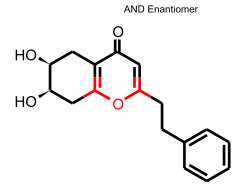
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 1875238785: [*]C\C(=C\[*])\O[*]
3. Unknown ECFP_6 feature: 464808839: [*]C(=CC(=[*])[*])[*]
4. Unknown ECFP_6 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
5. Unknown ECFP_6 feature: 1794461805: [*]CC(=C([*])[*])C(=[*])[*]
6. Unknown ECFP_6 feature: 53207596: [*]C([*])CC(=[*])[*]
7. Unknown ECFP_6 feature: -329826665: [*]C[C@H](O)C([*])[*]
8. Unknown ECFP_6 feature: 1650944136: [*]CC(=C([*])[*])O[*]
9. Unknown ECFP_6 feature: -560785749: [*]C(=[*])OC(=[*])[*]
10. Unknown ECFP_6 feature: -1795525632: [*]CCC(=[*])[*]
11. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
12. Unknown ECFP_6 feature: 2024749573: [*]C([*])O

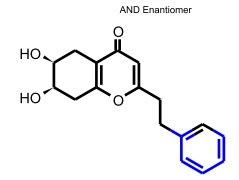
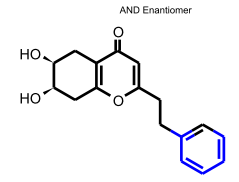
Feature Contribution

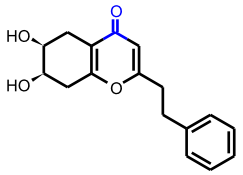
Top features for positive contribution

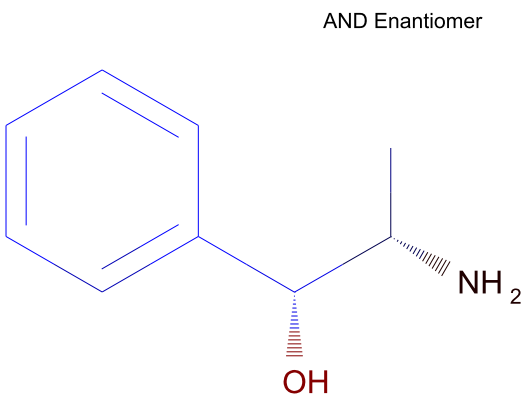
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	451847724		0.16
ECFP_6	-167460056		0.136
FCFP_6	-1143715940		0.13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244		-0.134
ECFP_6	1564392544		-0.133

ECFP_6	2106656448	AND Enantiomer  [*]C(=O)[*]	-0.11



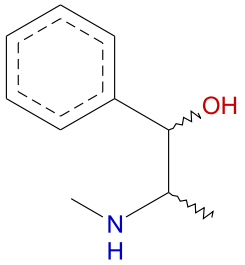
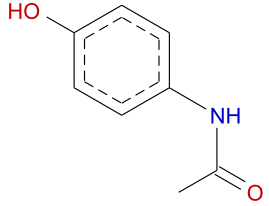
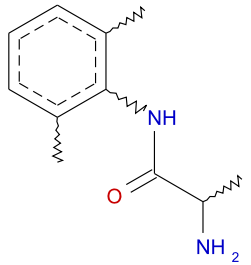
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 0.153
Unit: g/kg_body_weight
Mahalanobis Distance: 17.5
Mahalanobis Distance p-value: 0.000811

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE SULPHATE	ACETAMINOPHEN	TOCAINIDE.HCL
Structure			
Actual Endpoint (-log C)	4.42222	2.74809	3.28389
Predicted Endpoint (-log C)	3.18027	2.92206	2.93728
Distance	0.278	0.431	0.445
Reference	NTP 307 41	NTP REPORT # 394	NDA-18257

Model Applicability

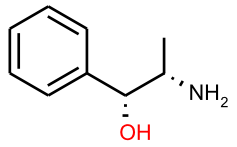
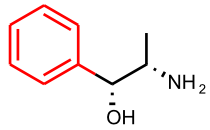
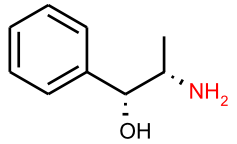
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 2023785560: [*]C([*])O
3. Unknown ECFP_6 feature: -222940837: [*]C([*])[C@H](O)[c](:[*]):[*]
4. Unknown ECFP_6 feature: -261056708: [*]C([*])[C@H](C)N
5. Unknown ECFP_6 feature: -933808133: [*]C([*])N
6. Unknown ECFP_6 feature: -176846085: [*]C([*])[c](:[cH]:[*]):[cH]:[*]
7. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

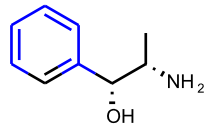
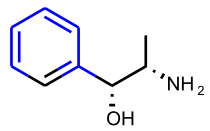
Feature Contribution

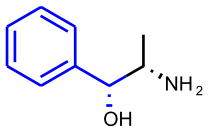
Top features for positive contribution

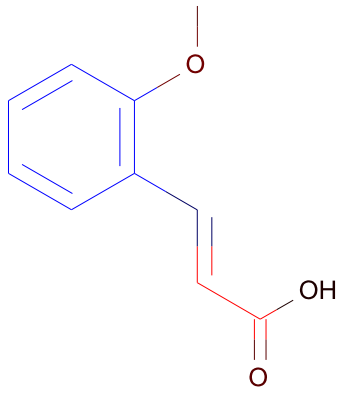
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	3	AND Enantiomer  [*]O	0.0924
FCFP_6	-2093839777	AND Enantiomer  [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.078
ECFP_6	1572579716	AND Enantiomer  [*]N	0.0576

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer  [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	AND Enantiomer  [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133

FCFP_6	-1698724694	AND Enantiomer  [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.0944



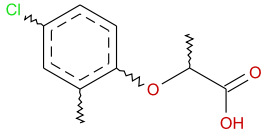
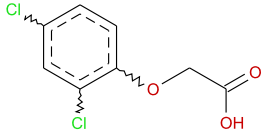
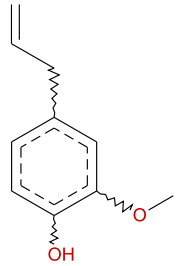
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.069
Unit: g/kg_body_weight
Mahalanobis Distance: 14.9
Mahalanobis Distance p-value: 0.177

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	11	2; 4-DICHLOROPHENOXY ACETIC ACID	EUGENOL
Structure			
Actual Endpoint (-log C)	3.91675	4.64549	2.73826
Predicted Endpoint (-log C)	3.52639	3.88474	3.36098
Distance	0.410	0.421	0.428
Reference	EPA COVER SHEET 0067;890101;(1)	HEED ECAO CIN G035;8803;(1)	NTP REPORT # 223

Model Applicability

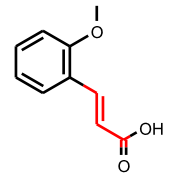
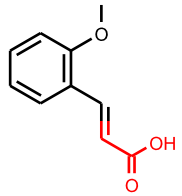
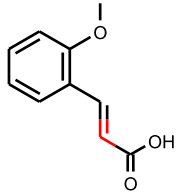
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -1905093414: [*]=CC(=O)O
3. Unknown ECFP_6 feature: -470416293: [*]\C=C\C(=[*])[*]
4. Unknown ECFP_6 feature: -1831055759: [*]\C=C\[c](:[*]):[*]
5. Unknown ECFP_6 feature: 1335702447: [*][c](:[*]):[c](C=[*]):[cH]:[*]
6. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
7. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

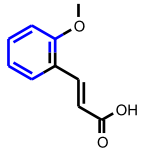
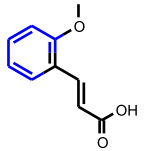
Feature Contribution

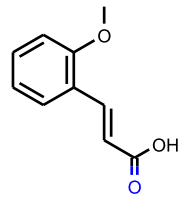
Top features for positive contribution

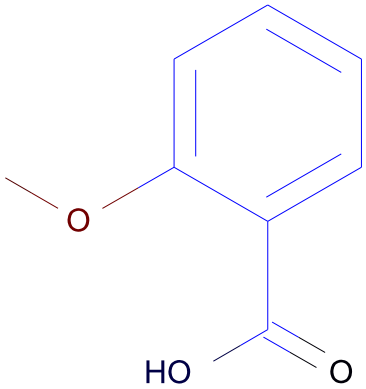
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.16
FCFP_6	-1176841573	 [*]=CC(=O)O	0.0963
ECFP_6	-1925046727	 [*]C=[*]	0.0915

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133

FCFP_6	1	 [*]=O	-0.102



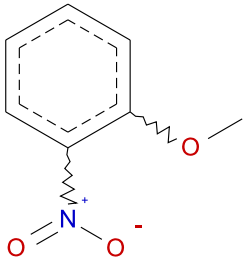
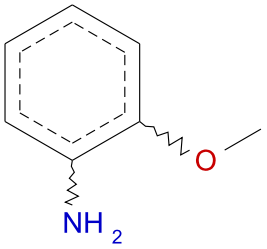
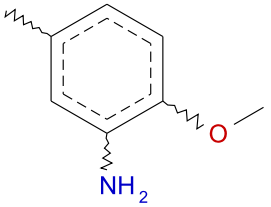
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.269
Unit: g/kg_body_weight
Mahalanobis Distance: 16.4
Mahalanobis Distance p-value: 0.0121

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	O-NITROANISOLE	O-ANISIDINE .HCL	P-CRESIDINE
Structure			
Actual Endpoint (-log C)	3.74733	2.69251	3.26224
Predicted Endpoint (-log C)	3.48518	3.50794	3.63547
Distance	0.401	0.415	0.434
Reference	NTP REPORT # 416	NTP REPORT # 89	NTP REPORT # 142

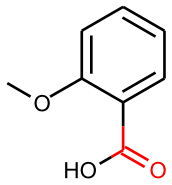
Model Applicability

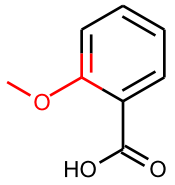
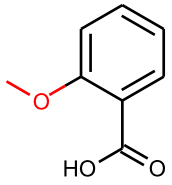
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
3. Unknown ECFP_6 feature: 1429461619: [*]:[c](:[*])C(=O)O
4. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

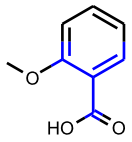
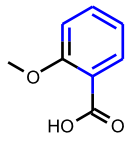
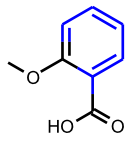
Feature Contribution

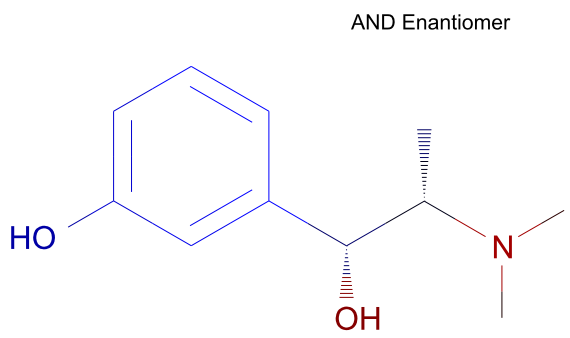
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2099970318	 [*]C(=O)[*]	0.0766

FCFP_6	1036089772	 [*]:[c](:[*])OC	0.073
FCFP_6	136627117	 [*]OC	0.0538

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1337040050	 [*]C(=[*])[c](:[cH]:[*]):[c]([*]):[*]	-0.158
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133



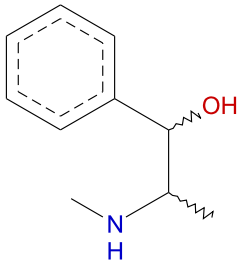
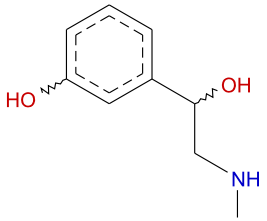
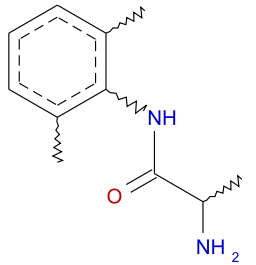
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 0.0937
Unit: g/kg_body_weight
Mahalanobis Distance: 22.4
Mahalanobis Distance p-value: 5.03e-011

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE SULPHATE	PHENYLEPHRINE .HCL	TOCAINIDE.HCL
Structure			
Actual Endpoint (-log C)	4.42222	3.7319	3.28389
Predicted Endpoint (-log C)	3.18027	3.37392	2.93728
Distance	0.423	0.423	0.503
Reference	NTP 307 41	NTP REPORT # 322	NDA-18257

Model Applicability

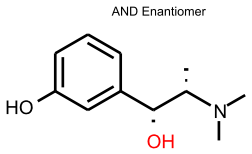
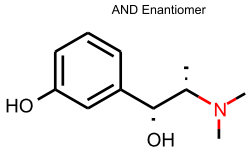
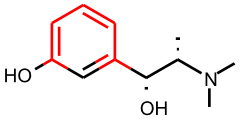
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC11 out of range. Value: 6.9229. Training min, max, SD, explained variance: -6.5761, 6.1703, 2.258, 0.0223.
2. Unknown ECFP_6 feature: 2023785560: [*]C([*])O
3. Unknown ECFP_6 feature: -222940837: [*]C([*])[C@H](O)[c](:[*]):[*]
4. Unknown ECFP_6 feature: -1280034494: [*]C([*])[C@H](C)N([*])[*]
5. Unknown ECFP_6 feature: 1064495017: [*]C([*])N(C)C
6. Unknown ECFP_6 feature: -176846085: [*]C([*])[c](:[cH]:[*]):[cH]:[*]
7. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
8. Unknown ECFP_6 feature: -177786161: [*]:[cH]:[c](O):[cH]:[*]
9. Unknown ECFP_6 feature: 2019062761: [*]:[c](:[*])O

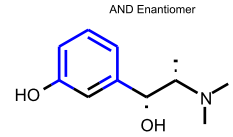
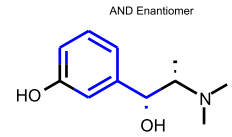
Feature Contribution

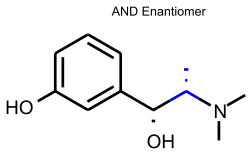
Top features for positive contribution

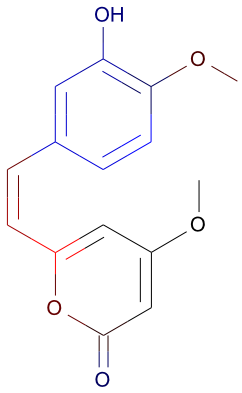
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	3	 [*]O	0.0924
FCFP_6	9	 [*]N([*])[*]	0.0797
ECFP_6	2007300961	 [*][c]1:[*]:[c]([*]):[cH]:[cH]:[cH]:1	0.0564

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134
FCFP_6	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.0906

FCFP_6	136597326	 [*]C([*])C	-0.0815



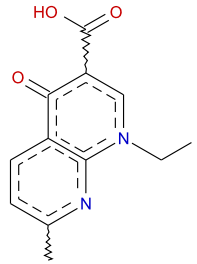
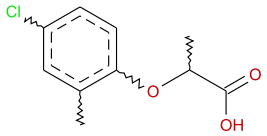
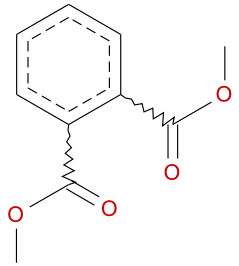
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: 0.0261
Unit: g/kg_body_weight
Mahalanobis Distance: 26.7
Mahalanobis Distance p-value: 1.24e-018

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	NALIDIXIC ACID	11	DIMETHYL PHTHALATE
Structure			
Actual Endpoint (-log C)	3.36594	3.91675	1.68616
Predicted Endpoint (-log C)	4.24773	3.52639	2.46505
Distance	0.547	0.585	0.587
Reference	NTP REPORT # 368	EPA COVER SHEET 0067;890101;(1)	AWQCD ECAO CIN H066;8710;(5)

Model Applicability

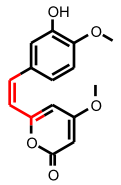
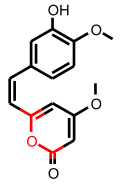
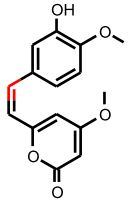
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -176483725: [*]=C[c](:[cH]:[*]):[cH]:[*]
3. Unknown ECFP_6 feature: 2019062761: [*]:[c](:[*])O
4. Unknown ECFP_6 feature: -1831055759: [*]\C=C\[c](:[*]):[*]
5. Unknown ECFP_6 feature: -1053980253: [*]O\C(=C/[*])\C=[*]
6. Unknown ECFP_6 feature: -560785749: [*]C(=[*])OC(=[*])[*]
7. Unknown ECFP_6 feature: -1885846789: [*]OC(=O)C=[*]
8. Unknown ECFP_6 feature: 464808839: [*]C(=CC(=[*])[*])[*]
9. Unknown ECFP_6 feature: -444332269: [*]O\C(=C\[*])\C=[*]
10. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

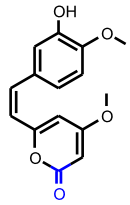
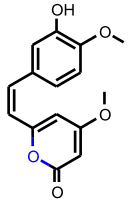
Feature Contribution

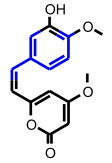
Top features for positive contribution

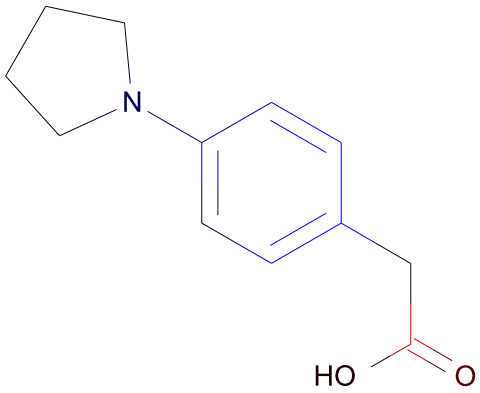
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.16
FCFP_6	-1143715940	 [*]C(=[*])OC(=[*])[*]	0.13
ECFP_6	-1925046727	 [*]C=[*]	0.0915

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11
FCFP_6	1	 [*]=O	-0.102

FCFP_6	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.0906



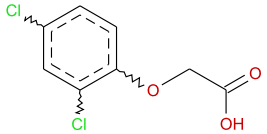
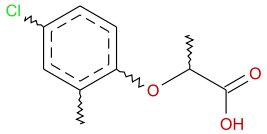
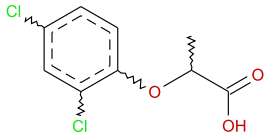
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.0467
Unit: g/kg_body_weight
Mahalanobis Distance: 25.2
Mahalanobis Distance p-value: 5.48e-016

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2; 4-DICHLOROPHENOXY ACETIC ACID	11	DICHLOROPROP
Structure			
Actual Endpoint (-log C)	4.64549	3.91675	3.97324
Predicted Endpoint (-log C)	3.88474	3.52639	3.57411
Distance	0.456	0.488	0.510
Reference	HEED ECAO CIN G035;8803;(1)	EPA COVER SHEET 0067;890101;(1)	HEEP ECAO CIN P079;8408;(2)

Model Applicability

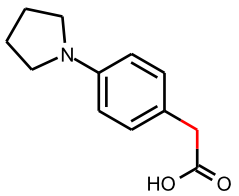
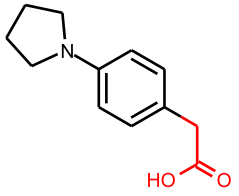
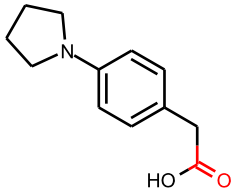
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC13 out of range. Value: 6.8267. Training min, max, SD, explained variance: -5.6594, 6.6668, 2.119, 0.0196.
2. Unknown ECFP_6 feature: 1731135544: [*]CC(=O)O
3. Unknown ECFP_6 feature: 771857573: [*]C(=[*])C[c](:[*]):[*]
4. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 1951894094: [*]:[c](:[*])N1C[*][*]C1
6. Unknown ECFP_6 feature: -757679000: [*]N1[*][*]CC1
7. Unknown ECFP_6 feature: -1332781180: [*]1[*]CCC1

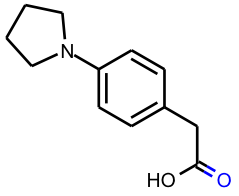
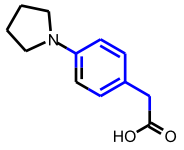
Feature Contribution

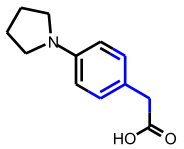
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

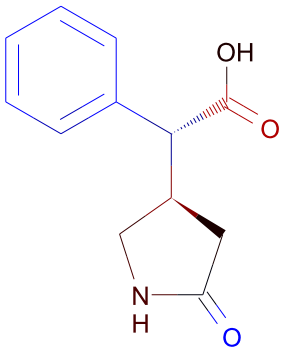
ECFP_6	1559650422	 [*]C[*]	0.129
FCFP_6	-1176841573	 [*]=CC(=O)O	0.0963
ECFP_6	2099970318	 [*]C(=O)[*]	0.0766

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	-0.102
FCFP_6	-453677277	 [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.0906

FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0713

AND Enantiomer



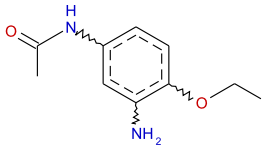
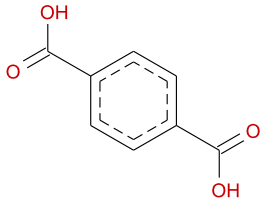
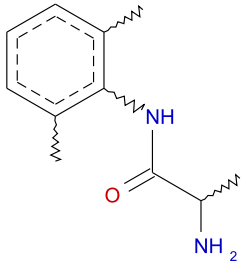
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 0.0959
Unit: g/kg_body_weight
Mahalanobis Distance: 26.8
Mahalanobis Distance p-value: 7.21e-019

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3-AMINO-4-ETHOXYACETANILIDE	P-PHTHALIC ACID	TOCAINIDE.HCL
Structure			
Actual Endpoint (-log C)	2.9873	2.52148	3.28389
Predicted Endpoint (-log C)	3.01152	2.63704	2.93728
Distance	0.461	0.519	0.527
Reference	NTP 112 C-11	HEEP ECAO CIN P190;8608;(2)	NDA-18257

Model Applicability

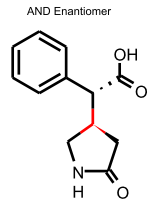
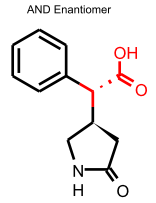
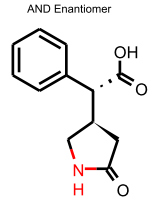
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -154530762: [*]N[*]
3. Unknown ECFP_6 feature: -1694930393: [*]=C1[*][*]CN1
4. Unknown ECFP_6 feature: -742538367: O=C1C[*][*]N1
5. Unknown ECFP_6 feature: 53207596: [*]C([*])CC(=[*])[*]
6. Unknown ECFP_6 feature: -858846751: [*]C([*])C1C[*][*]C1
7. Unknown ECFP_6 feature: -1457159889: [*][C@H]1[*][*]NC1
8. Unknown ECFP_6 feature: -1905455774: [*]C([*])C(=O)O
9. Unknown ECFP_6 feature: 1603312431: [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]
10. Unknown ECFP_6 feature: -176846085: [*]C([*])[c](:[cH]:[*]):[cH]:[*]
11. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

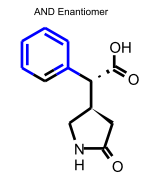
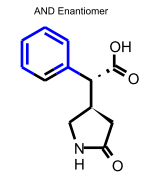
Feature Contribution

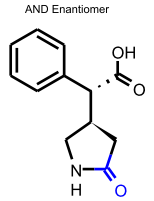
Top features for positive contribution

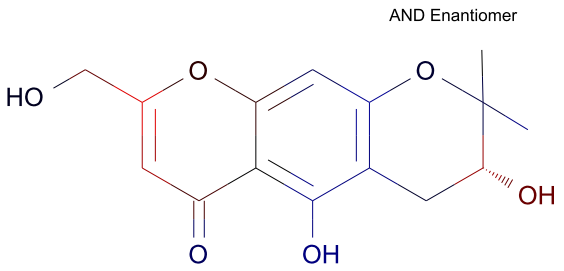
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	-167460056	 [*]C([*])[*]	0.136
FCFP_6	-1176841573	 [*]=CC(=O)O	0.0963
FCFP_6	3	 [*]O	0.0924

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133

ECFP_6	2106656448	AND Enantiomer  [*]C(=O)[*]	-0.11



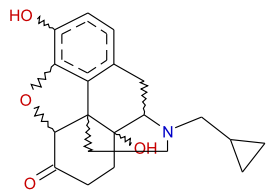
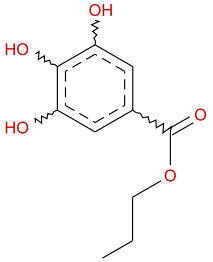
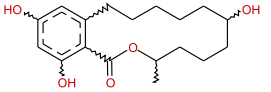
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: 0.0104
Unit: g/kg_body_weight
Mahalanobis Distance: 28.5
Mahalanobis Distance p-value: 5.37e-022

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	NALTREXONE.HCL	PROPYL GALLATE	ZERANOL
Structure			
Actual Endpoint (-log C)	4.05615	2.84963	4.20737
Predicted Endpoint (-log C)	4.77886	2.91297	3.4948
Distance	0.612	0.619	0.646
Reference	NDA-18932	NTP REPORT # 240	REGULAT. TOXICOL. PHARMACOL. 1983; 3: 9-

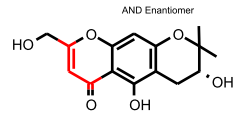
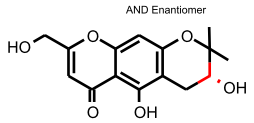
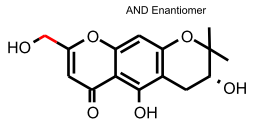
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

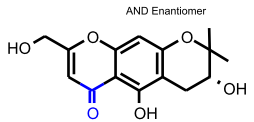
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -1660340418: [*]C[c](:[c]([*]):[*]):[c]([*]):[*]
3. Unknown ECFP_6 feature: -570915357: [*]O[c](:[cH]:[*]):[c]([*]):[*]
4. Unknown ECFP_6 feature: -813997308: [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]
5. Unknown ECFP_6 feature: -1660913849: [*][c](:[*]):[c](O):[c]([*]):[*]
6. Unknown ECFP_6 feature: -560785749: [*]C(=[*])OC(=[*])[*]
7. Unknown ECFP_6 feature: 1875238785: [*]C\C(=C\[*])\O[*]
8. Unknown ECFP_6 feature: 464808839: [*]C(=CC(=[*])[*])[*]
9. Unknown ECFP_6 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
10. Unknown ECFP_6 feature: 2019062761: [*]:[c](:[*])O
11. Unknown ECFP_6 feature: 53207596: [*]C([*])CC(=[*])[*]
12. Unknown ECFP_6 feature: -1051556861: [*]C[C@@H](O)C([*])([*])[*]
13. Unknown ECFP_6 feature: 1778376725: [*]OC(C)(C)C([*])[*]
14. Unknown ECFP_6 feature: -200406221: [*]C([*])([*])O[c](:[*]):[*]
15. Unknown ECFP_6 feature: 2024749573: [*]C([*])O
16. Unknown ECFP_6 feature: 859433814: [*]C([*])([*])C
17. Unknown ECFP_6 feature: 769217534: [*]C(=[*])CO
18. Unknown ECFP_6 feature: 2022454958: [*]CO

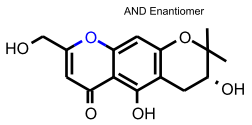
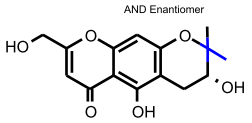
Feature Contribution

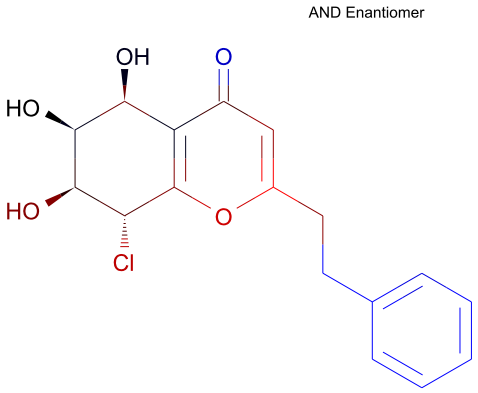
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.16
ECFP_6	-167460056	 [*]C([*])[*]	0.136
ECFP_6	1559650422	 [*]C[*]	0.129

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11

FCFP_6	1	 [*]=O	-0.102
FCFP_6	136597326	 [*]C([*])C	-0.0815



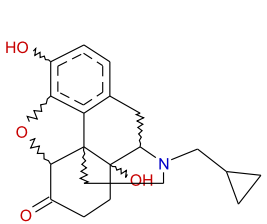
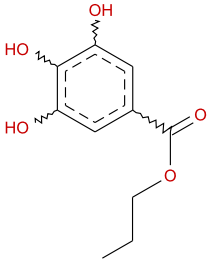
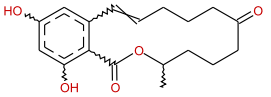
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: 0.0201
Unit: g/kg_body_weight
Mahalanobis Distance: 26.8
Mahalanobis Distance p-value: 8.04e-019

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	NALTREXONE.HCL	PROPYL GALLATE	ZEARALENONE
Structure			
Actual Endpoint (-log C)	4.05615	2.84963	5.40602
Predicted Endpoint (-log C)	4.77886	2.91297	3.57081
Distance	0.585	0.627	0.653
Reference	NDA-18932	NTP REPORT # 240	NTP REPORT # 235

Model Applicability

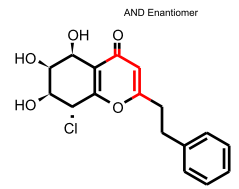
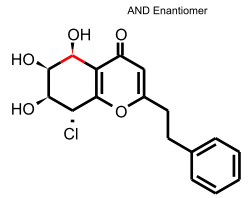
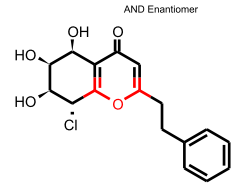
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 1875238785: [*]C\C(=C\[*])\O[*]
3. Unknown ECFP_6 feature: 464808839: [*]C(=CC(=[*])[*])[*]
4. Unknown ECFP_6 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
5. Unknown ECFP_6 feature: 1795792463: [*]C([*])C(=C([*])[*])C(=[*])[*]
6. Unknown ECFP_6 feature: -1567907747: [*]C([*])[C@@H](O)C(=[*])[*]
7. Unknown ECFP_6 feature: 305695353: [*]C([*])C(O)C([*])[*]
8. Unknown ECFP_6 feature: -7106223: [*]C([*])[C@@H](Cl)C(=[*])[*]
9. Unknown ECFP_6 feature: 1652274794: [*]OC(=C([*])[*])C([*])[*]
10. Unknown ECFP_6 feature: -560785749: [*]C(=[*])OC(=[*])[*]
11. Unknown ECFP_6 feature: 2024749573: [*]C([*])O
12. Unknown ECFP_6 feature: -1795525632: [*]CCC(=[*])[*]
13. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
14. Unknown ECFP_6 feature: 105634199: [*]C([*])Cl

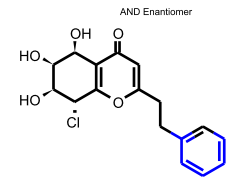
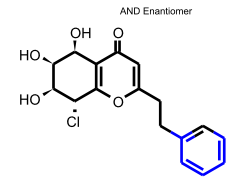
Feature Contribution

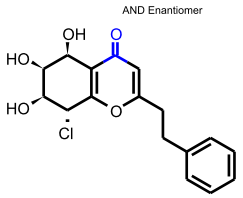
Top features for positive contribution

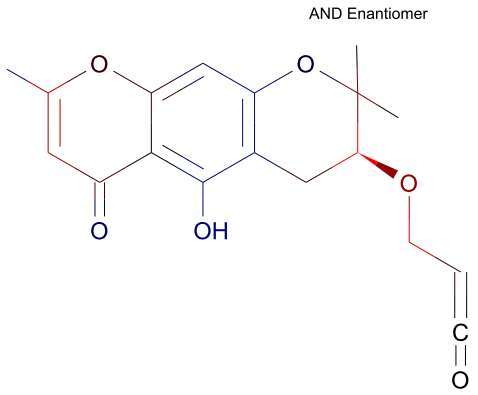
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.16
ECFP_6	-167460056	 [*]C([*])[*]	0.136
FCFP_6	-1143715940	 [*]C(=[*])OC(=[*])[*]	0.13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134
ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133

ECFP_6	2106656448	AND Enantiomer  [*]C(=O)[*]	-0.11



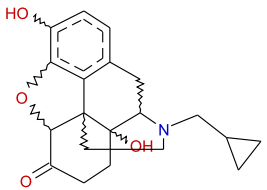
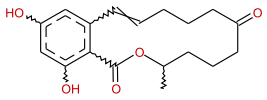
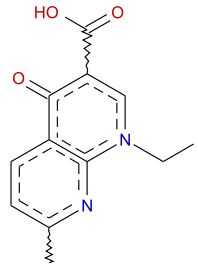
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.0112
Unit: g/kg_body_weight
Mahalanobis Distance: 31
Mahalanobis Distance p-value: 1.57e-026

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	NALTREXONE.HCL	ZEARALENONE	NALIDIXIC ACID
Structure			
Actual Endpoint (-log C)	4.05615	5.40602	3.36594
Predicted Endpoint (-log C)	4.77886	3.57081	4.24773
Distance	0.462	0.571	0.644
Reference	NDA-18932	NTP REPORT # 235	NTP REPORT # 368

Model Applicability

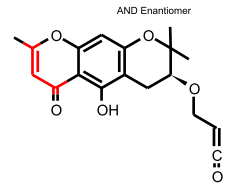
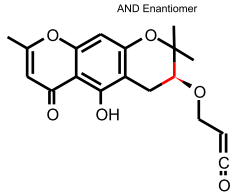
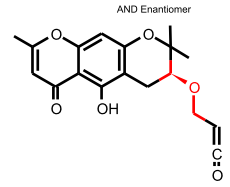
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 444624378: [*]C=C=O
3. Unknown ECFP_6 feature: -1114776580: [*]=C=[*]
4. Unknown ECFP_6 feature: -1660340418: [*]C[c](:[c]([*]):[*]):[c]([*]):[*]
5. Unknown ECFP_6 feature: -570915357: [*]O[c](:[cH]:[*]):[c]([*]):[*]
6. Unknown ECFP_6 feature: -813997308: [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]
7. Unknown ECFP_6 feature: -1660913849: [*][c](:[*]):[c](O):[c]([*]):[*]
8. Unknown ECFP_6 feature: -560785749: [*]C(=[*])OC(=[*])[*]
9. Unknown ECFP_6 feature: -331149802: [*]O\C(=C/[*])\C
10. Unknown ECFP_6 feature: 464808839: [*]C(=CC(=[*])[*])[*]
11. Unknown ECFP_6 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
12. Unknown ECFP_6 feature: 53207596: [*]C([*])CC(=[*])[*]
13. Unknown ECFP_6 feature: -2124995946: [*]C[C@H](O[*])C([*])([*])[*]
14. Unknown ECFP_6 feature: 1778376725: [*]OC(C)(C)C([*])[*]
15. Unknown ECFP_6 feature: -200406221: [*]C([*])([*])O[c](:[*]):[*]
16. Unknown ECFP_6 feature: 859433814: [*]C([*])([*])C
17. Unknown ECFP_6 feature: 2019062761: [*]:[c](:[*])O
18. Unknown ECFP_6 feature: -1591590376: [*]C=C=O
19. Unknown ECFP_6 feature: 2106995136: [*]=C=O
20. Unknown ECFP_6 feature: -91536905: [*]CC=C=[*]

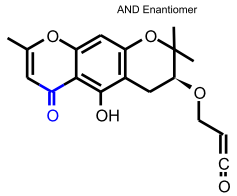
21. Unknown ECFP_6 feature: -1688150664: [*]OCC=[*]
22. Unknown ECFP_6 feature: -1250019913: [*]COC([*])[*]

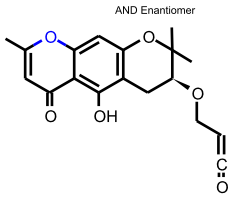
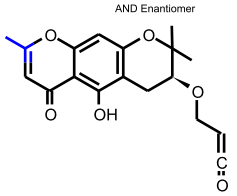
Feature Contribution

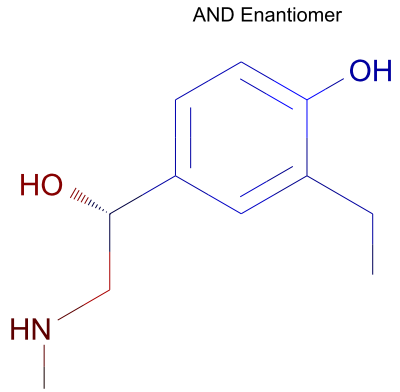
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.16
ECFP_6	-167460056	 [*]C([*])[*]	0.136
FCFP_6	-1143715940	 [*]C(=[*])OC(=[*])[*]	0.13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11

FCFP_6	1	 [*]=O	-0.102
FCFP_6	136597326	 [*]C([*])C	-0.0815



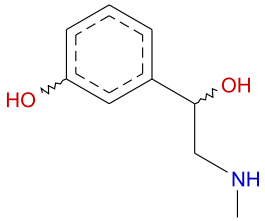
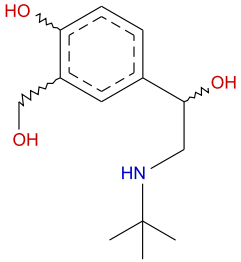
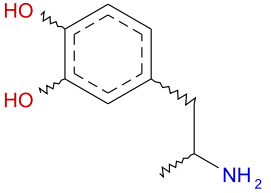
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 0.0809
Unit: g/kg_body_weight
Mahalanobis Distance: 19
Mahalanobis Distance p-value: 1.11e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	PHENYLEPHRINE .HCL	ALBUTEROL.SULFATE	ALPHA-METHYLDOPASESQUITTYDRATE
Structure			
Actual Endpoint (-log C)	3.7319	4.37897	3.03293
Predicted Endpoint (-log C)	3.37392	3.92079	2.95075
Distance	0.307	0.427	0.475
Reference	NTP REPORT # 322	NDA-19269	NTP REPORT # 348

Model Applicability

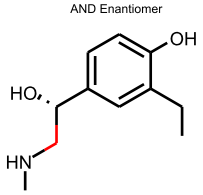
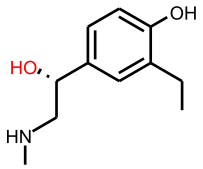
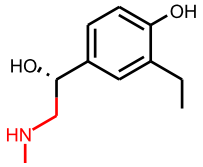
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -2024509555: [*]C[c](:[cH]:[*]):[c]([*]):[*]
3. Unknown ECFP_6 feature: -176846085: [*]C([*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: 2019062761: [*]:[c](:[*])O
5. Unknown ECFP_6 feature: -1537923023: [*]C[C@H](O)[c](:[*]):[*]
6. Unknown ECFP_6 feature: -652986225: [*]NCC([*])[*]
7. Unknown ECFP_6 feature: 2023785560: [*]C([*])O
8. Unknown ECFP_6 feature: 493154328: [*]CNC
9. Unknown ECFP_6 feature: 864287155: [*]NC
10. Unknown ECFP_6 feature: 767488533: [*]:[c](:[*])CC

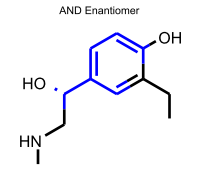
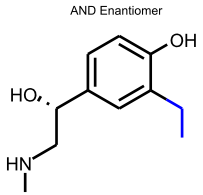
Feature Contribution

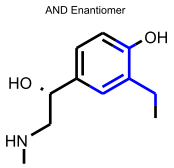
Top features for positive contribution

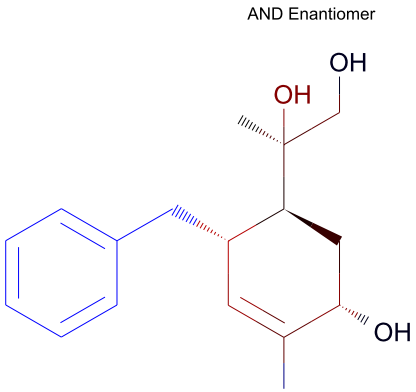
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	1559650422	AND Enantiomer  [*]C[*]	0.129
FCFP_6	3	AND Enantiomer  [*]O	0.0924
FCFP_6	-885550502	AND Enantiomer  [*]C([*])NC	0.0684

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-453677277	AND Enantiomer  [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.0906
FCFP_6	136597326	AND Enantiomer  [*]C([*])C	-0.0815

FCFP_6	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0713



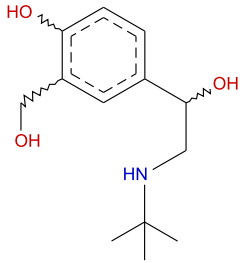
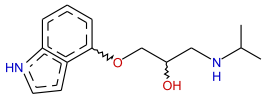
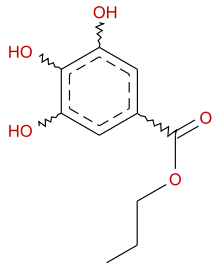
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 0.0495
Unit: g/kg_body_weight
Mahalanobis Distance: 30.6
Mahalanobis Distance p-value: 1.04e-025

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ALBUTEROL.SULFATE	PINDOLOL	PROPYL GALLATE
Structure			
Actual Endpoint (-log C)	4.37897	3.4038	2.84963
Predicted Endpoint (-log C)	3.92079	4.03315	2.91297
Distance	0.575	0.617	0.620
Reference	NDA-19269	NDA-18285	NTP REPORT # 240

Model Applicability

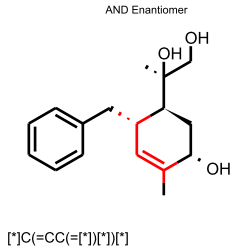
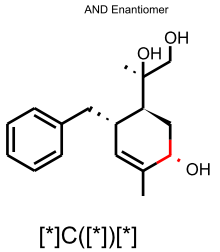
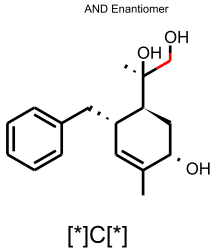
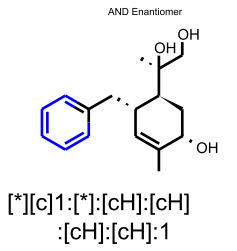
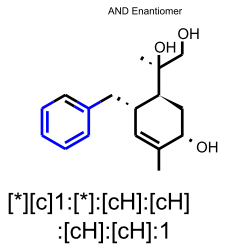
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

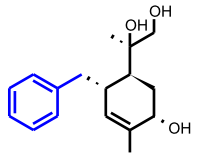
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -2097867909: [*]C[C@H](O)C(=[*])[*]
3. Unknown ECFP_6 feature: -801490360: [*]C([*])CC([*])[*]
4. Unknown ECFP_6 feature: -327548242: [*]C[C@@H](C([*])[*])C([*])([*])[*]
5. Unknown ECFP_6 feature: -1263967621: [*]C[C@@H](C=[*])C([*])[*]
6. Unknown ECFP_6 feature: 470495651: [*]C([*])C=C([*])[*]
7. Unknown ECFP_6 feature: -1042330089: [*]\C=C(\C)/C([*])[*]
8. Unknown ECFP_6 feature: 1280892564: [*]C[C@](C)(O)C([*])[*]
9. Unknown ECFP_6 feature: -1907755304: [*]C([*])([*])CO
10. Unknown ECFP_6 feature: 865857320: [*]C([*])([*])C
11. Unknown ECFP_6 feature: 2025123907: [*]C([*])([*])O
12. Unknown ECFP_6 feature: 2024749573: [*]C([*])O
13. Unknown ECFP_6 feature: 771121623: [*]C([*])C[c](:[*]):[*]
14. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
15. Unknown ECFP_6 feature: 2022454958: [*]CO

Feature Contribution

Top features for positive contribution

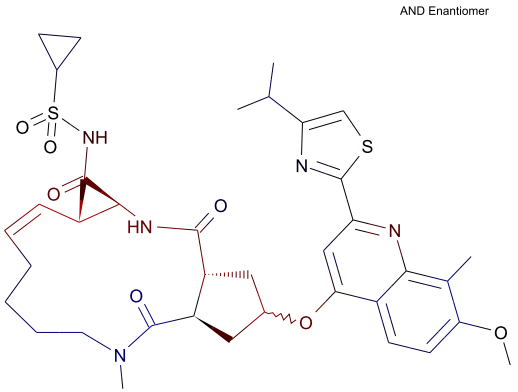
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	451847724		0.16
ECFP_6	-167460056		0.136
ECFP_6	1559650422		0.129
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	991735244		-0.134
ECFP_6	1564392544		-0.133

FCFP_6	-1698724694	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0944

Simeprevir

TOPKAT_Chronic_LOAEL



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: 0.00211

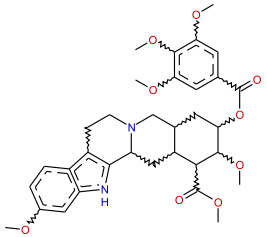
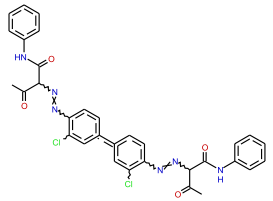
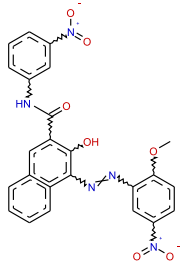
Unit: g/kg_body_weight

Mahalanobis Distance: 55.7

Mahalanobis Distance p-value: 1.58e-065

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	RESERPINE	DIARYLANILIDE YELLOW	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	6.38645	2.70208	2.28997
Predicted Endpoint (-log C)	5.548	3.76154	3.52921
Distance	0.885	0.887	0.922
Reference	NTP 193 22	NTP 30 C-4	NTP 411 146

Model Applicability

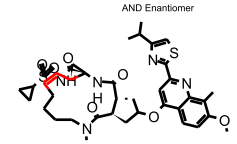
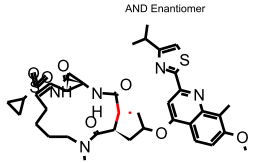
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

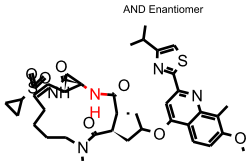
1. OPS PC1 out of range. Value: 18.64. Training min, max, SD, explained variance: -9.2986, 15.594, 5, 0.1094.
2. Unknown FCFP_2 feature: 1018942292: [*]NS(=O)(=O)C1[*][*]1
3. Unknown ECFP_6 feature: -154530762: [*]N[*]
4. Unknown ECFP_6 feature: 914325265: [*]:s:[*]
5. Unknown ECFP_6 feature: -797085356: [*]S(=[*])(=[*])[*]
6. Unknown ECFP_6 feature: -2095963820: [*][C@@H]1[*][*]C[C@H]1C(=[*])[*]
7. Unknown ECFP_6 feature: -867777309: [*]NC(=O)C([*])[*]
8. Unknown ECFP_6 feature: -1338907019: [*]C(=[*])NC1([*])[*][*]1
9. Unknown ECFP_6 feature: -309676563: [*]N[C@]1(C[C@H]1[*])C(=[*])[*]
10. Unknown ECFP_6 feature: 413587124: [*][C@@H]1CC1([*])[*]
11. Unknown ECFP_6 feature: -1049290660: [*]C1([*])C[C@H]1C=[*]
12. Unknown ECFP_6 feature: 890368401: [*]\C=C/C1[*][*]1
13. Unknown ECFP_6 feature: 360408239: [*]C\C=C/[*]
14. Unknown ECFP_6 feature: -1331088410: [*]CCC=[*]
15. Unknown ECFP_6 feature: -1332781180: [*]1[*]CCC1
16. Unknown ECFP_6 feature: -757679000: [*]N1[*][*]CC1
17. Unknown ECFP_6 feature: 1616402542: [*]CN(C)C(=[*])[*]
18. Unknown ECFP_6 feature: 1526862590: [*]C([*])C(=O)N([*])[*]
19. Unknown ECFP_6 feature: -801490360: [*]C([*])CC([*])[*]

20. Unknown ECFP_6 feature: 1480368712: [*]OC1C[*][*]C1
21. Unknown ECFP_6 feature: -1818486371: [*]NC(=O)C1([*])[*][*]1
22. Unknown ECFP_6 feature: 946167604: [*]C(=[*])NS(=[*])(=[*])[*]
23. Unknown ECFP_6 feature: 866343404: [*]N([*])C
24. Unknown ECFP_6 feature: 1310940530: [*]C([*])O[c](:[*]):[*]
25. Unknown ECFP_6 feature: -428002189: [*]:[cH]:[c](:n:[*])[c](:[*]):[*]
26. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
27. Unknown ECFP_6 feature: 1576255326: [*][c](:[*]):[c](C):[c](:[*]):[*]
28. Unknown ECFP_6 feature: 1048320787: [*][c](:[*]):[c](:n:[*]):[c](:[*]):[*]
29. Unknown ECFP_6 feature: -1426923364: [*][c]1:[*]:[*]:[cH]:s:1
30. Unknown ECFP_6 feature: -253227249: [*]:[c](:[*])[c]1:n:[*]:[*]:s:1
31. Unknown ECFP_6 feature: 1411720546: [*]C([*])[c]1:[cH]:[*]:[*]:n:1
32. Unknown ECFP_6 feature: -224638920: [*][c]1:[*]:[*]:s:[cH]:1
33. Unknown ECFP_6 feature: 733491677: [*]:[c](:[*])C(C)C
34. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC
35. Unknown ECFP_6 feature: 2121941848: [*]NS(=O)(=O)C1[*][*]1
36. Unknown ECFP_6 feature: 2102150379: [*]S(=[*])(=O)[*]
37. Unknown ECFP_6 feature: -622223421: [*]S(=[*])(=[*])C1CC1

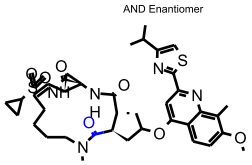
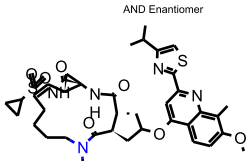
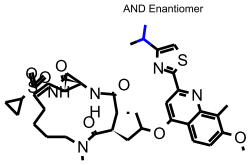
Feature Contribution

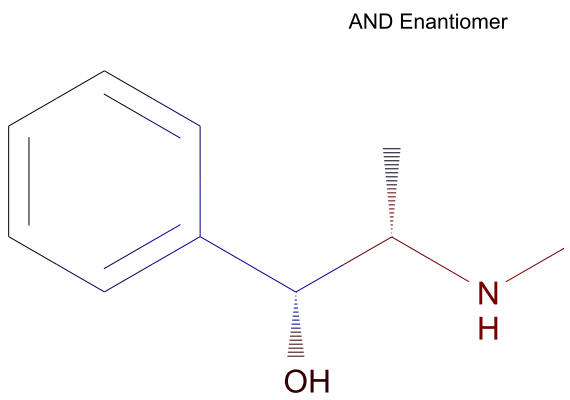
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.16
ECFP_6	-167460056	 [*]C([*])[*]	0.136

FCFP_6	3	 [*]O	0.0924

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11
FCFP_6	1	 [*]=O	-0.102
FCFP_6	136597326	 [*]C([*])C	-0.0815



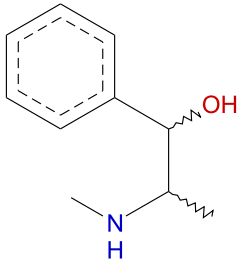
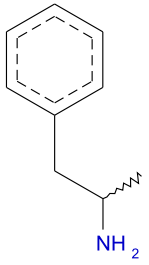
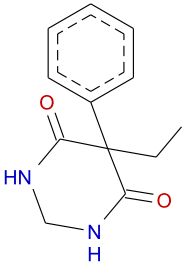
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 0.127
Unit: g/kg_body_weight
Mahalanobis Distance: 6.7
Mahalanobis Distance p-value: 0.308

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE SULFATE	DL-AMPHETAMINE SULFATE	PRIMIDONE (PRIMACLONE)
Structure			
Actual Endpoint (-log C)	4.16695	4.43204	3.94103
Predicted Endpoint (-log C)	3.11455	3.18009	3.24349
Distance	0.000	0.355	0.474
Reference	NCI/NTP TR-307	NCI/NTP TR-387	NCI/NTP TR-476

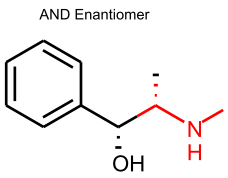
Model Applicability

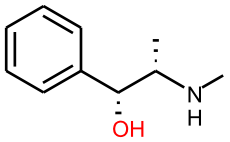
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

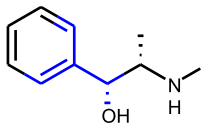
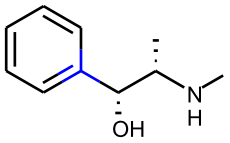
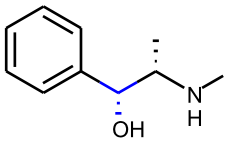
Feature Contribution

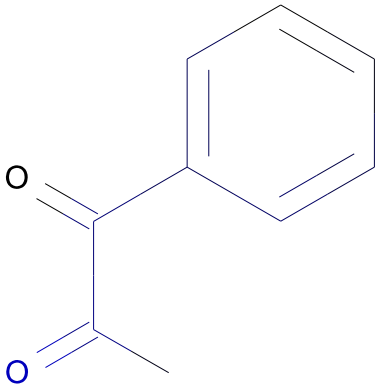
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-885550502	 [*]C([*])NC	0.115

FCFP_2	3	AND Enantiomer  [*]O	0.0737

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829
FCFP_2	16	AND Enantiomer  [*][c](:[*]):[*]	-0.0512
FCFP_2	0	AND Enantiomer  [*]C([*])[*]	-0.0314



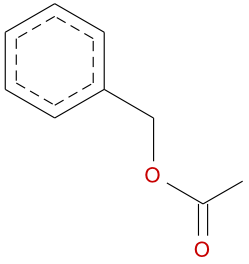
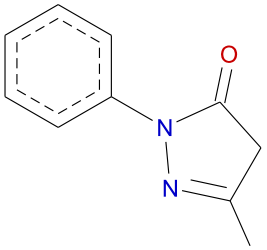
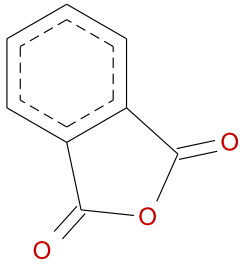
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: 0.092
Unit: g/kg_body_weight
Mahalanobis Distance: 4.63
Mahalanobis Distance p-value: 0.988

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	BENZYL ACETATE	1-PHENYL-3-METHYL-5-PYRAZOLONE	PHTHALIC ANHYDRIDE
Structure			
Actual Endpoint (-log C)	2.46903	2.88887	2.3413
Predicted Endpoint (-log C)	3.22857	3.45184	3.37248
Distance	0.296	0.348	0.371
Reference	NCI/NTP TR-431	NCI/NTP TR-141	NCI/NTP TR-159

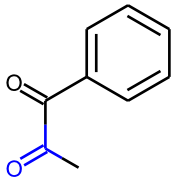
Model Applicability

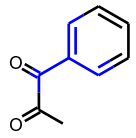
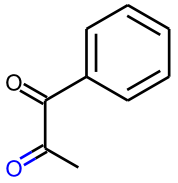
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

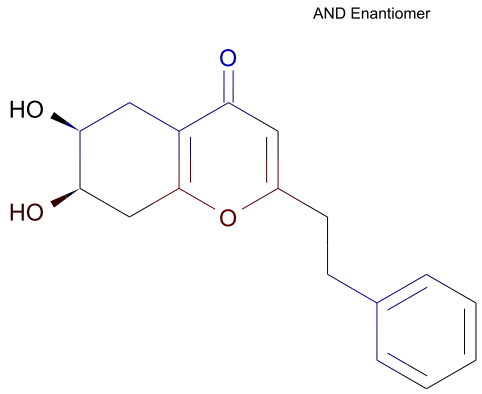
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105

FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829
FCFP_2	1	 [*]=O	-0.0796



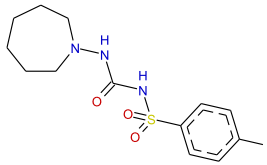
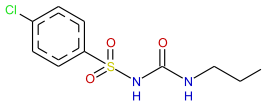
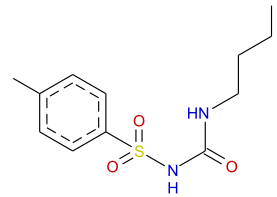
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: 0.124
Unit: g/kg_body_weight
Mahalanobis Distance: 6.78
Mahalanobis Distance p-value: 0.275

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	TOLAZAMIDE	CHLORPROPAMIDE	TOLBUTAMIDE
Structure			
Actual Endpoint (-log C)	2.84011	3.0107	2.3985
Predicted Endpoint (-log C)	3.59315	3.18321	3.32272
Distance	0.441	0.474	0.498
Reference	NCI/NTP TR-051	NCI/NTP TR-045	NCI/NTP TR-031

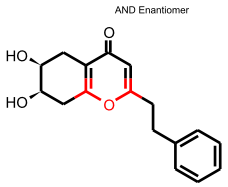
Model Applicability

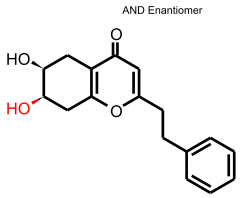
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

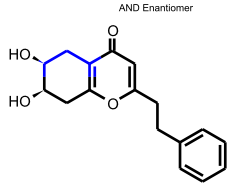
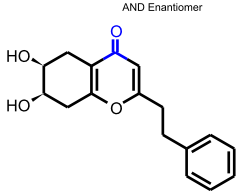
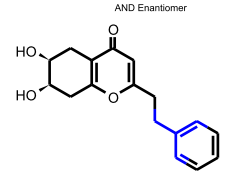
Feature Contribution

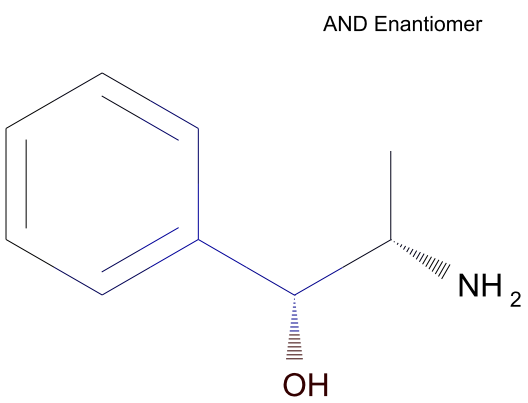
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C(=[*])OC(=[*])[*]	0.095

FCFP_2	3	 [*]O	0.0737

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	 [*]C([*])CC(=[*])[*]	-0.111
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



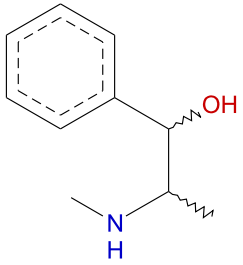
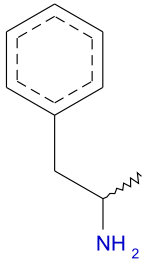
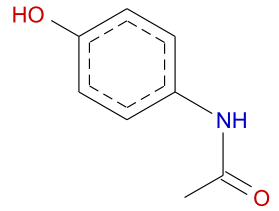
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 0.131
Unit: g/kg_body_weight
Mahalanobis Distance: 4.61
Mahalanobis Distance p-value: 0.989

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE SULFATE	DL-AMPHETAMINE SULFATE	ACETAMINOPHEN (4-HYDROXYACETANILIDE)
Structure			
Actual Endpoint (-log C)	4.16695	4.43204	2.74809
Predicted Endpoint (-log C)	3.11455	3.18009	2.67552
Distance	0.229	0.326	0.417
Reference	NCI/NTP TR-307	NCI/NTP TR-387	NCI/NTP TR-394

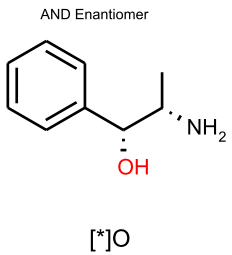
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

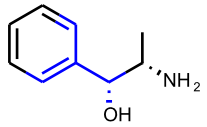
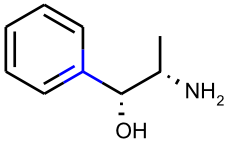
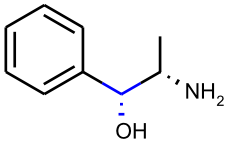
Feature Contribution

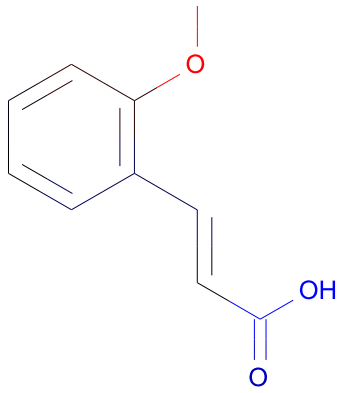
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3	 [*]O	0.0737

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829
FCFP_2	16	AND Enantiomer  [*][c](:[*]):[*]	-0.0512
FCFP_2	0	AND Enantiomer  [*]C([*])[*]	-0.0314



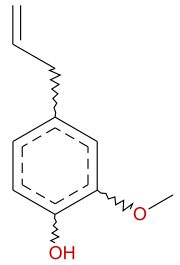
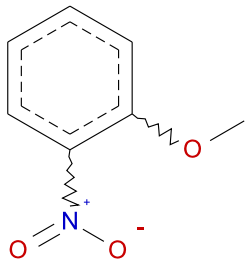
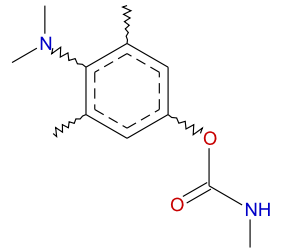
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.186
Unit: g/kg_body_weight
Mahalanobis Distance: 6.15
Mahalanobis Distance p-value: 0.566

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EUGENOL	O-NITROANISOLE	MEXACARBATE
Structure			
Actual Endpoint (-log C)	2.78402	3.18813	4.07253
Predicted Endpoint (-log C)	2.98617	3.55234	3.7152
Distance	0.369	0.369	0.401
Reference	NCI/NTP TR-223	NCI/NTP TR-416	NCI/NTP TR-147

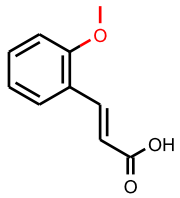
Model Applicability

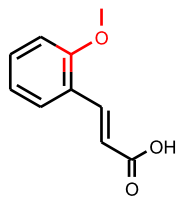
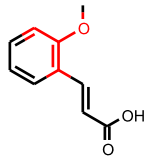
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

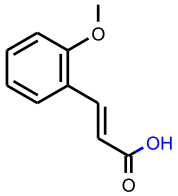
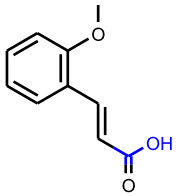
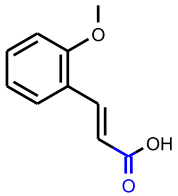
Feature Contribution

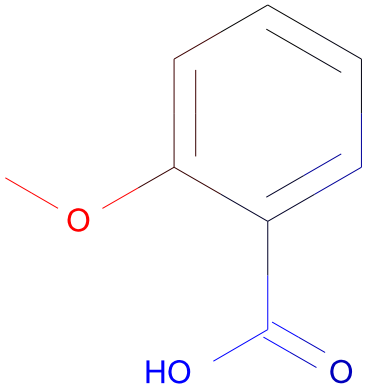
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.0611

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	 [*]O	-0.214
FCFP_2	-548632217	 [*]C(=[*])O	-0.119
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105



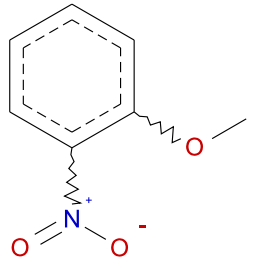
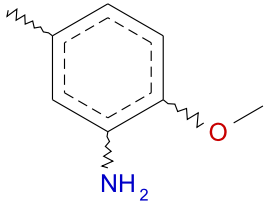
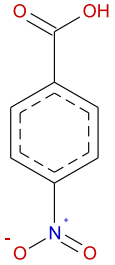
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.178
Unit: g/kg_body_weight
Mahalanobis Distance: 6.16
Mahalanobis Distance p-value: 0.565

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	O-NITROANISOLE	P-CRESIDINE	P-NITROBENZOIC ACID
Structure			
Actual Endpoint (-log C)	3.18813	2.48408	2.90081
Predicted Endpoint (-log C)	3.55234	3.53389	2.77599
Distance	0.309	0.366	0.410
Reference	NCI/NTP TR-416	NCI/NTP TR-142	NCI/NTP TR-442

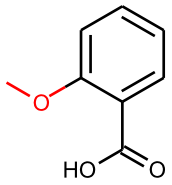
Model Applicability

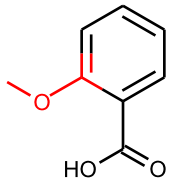
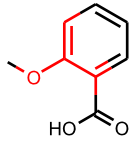
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

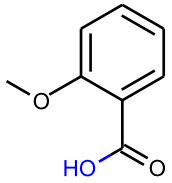
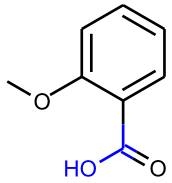
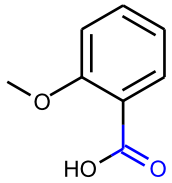
Feature Contribution

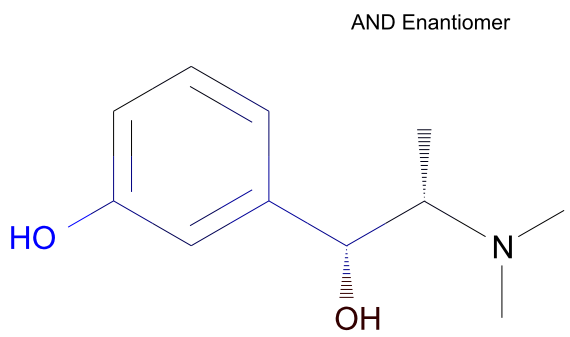
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.0611

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	 [*]O	-0.214
FCFP_2	-548632217	 [*]C(=[*])O	-0.119
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105



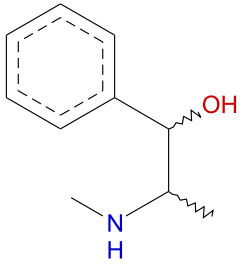
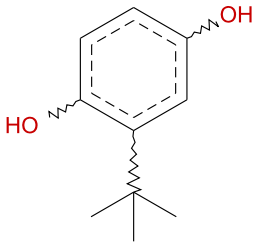
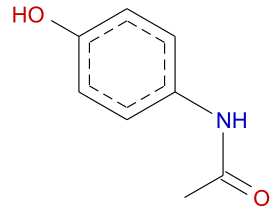
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 0.356
Unit: g/kg_body_weight
Mahalanobis Distance: 6.25
Mahalanobis Distance p-value: 0.518

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE SULFATE	t-BUTYLHYDROQUINONE	ACETAMINOPHEN (4-HYDROXYACETANILIDE)
Structure			
Actual Endpoint (-log C)	4.16695	2.44253	2.74809
Predicted Endpoint (-log C)	3.11455	2.62758	2.67552
Distance	0.396	0.433	0.476
Reference	NCI/NTP TR-307	NCI/NTP TR-459	NCI/NTP TR-394

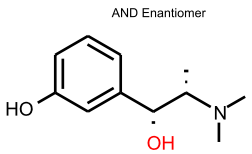
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

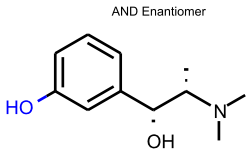
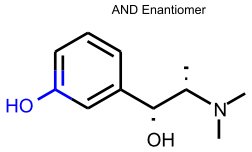
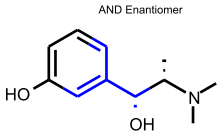
Feature Contribution

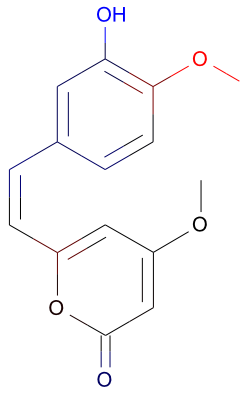
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3	 [*]O	0.0737

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	7	 [*]O	-0.214
FCFP_2	-549108873	 [*]:[c](:[*])O	-0.127
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



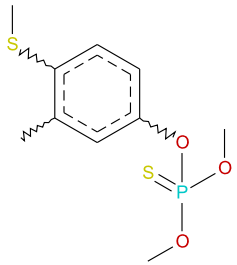
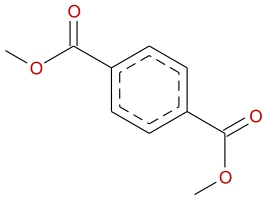
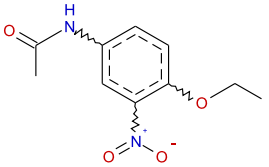
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: 0.109
Unit: g/kg_body_weight
Mahalanobis Distance: 8.17
Mahalanobis Distance p-value: 0.0158

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	FENTHION	DIMETHYL TEREPHTHALATE	3-NITRO-P-ACETOPHENETIDE
Structure			
Actual Endpoint (-log C)	5.4903	2.93604	3.11767
Predicted Endpoint (-log C)	4.18234	3.50861	3.42808
Distance	0.486	0.505	0.527
Reference	NCI/NTP TR-103	NCI/NTP TR-121	NCI/NTP TR-133

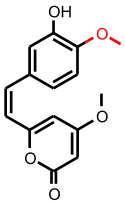
Model Applicability

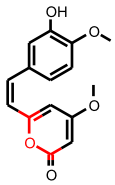
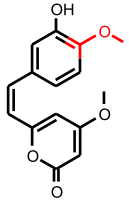
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

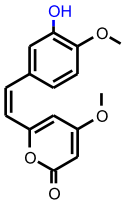
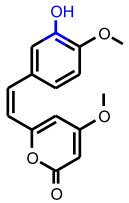
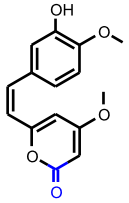
Feature Contribution

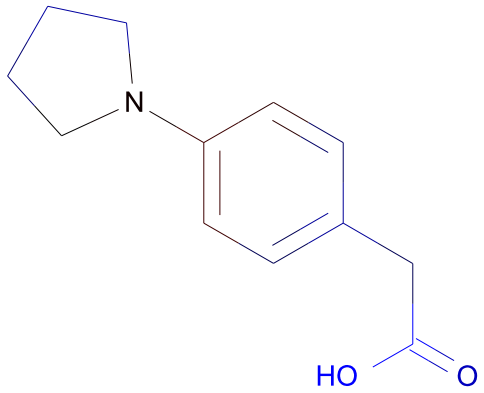
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

FCFP_2	-1143715940	 [*]C(=[*])OC(=[*])[*]	0.095
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	 [*]O	-0.214
FCFP_2	-549108873	 [*]:[c](:[*])O	-0.127
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105



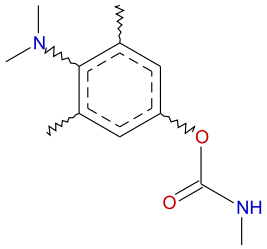
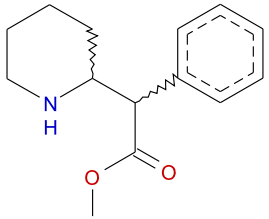
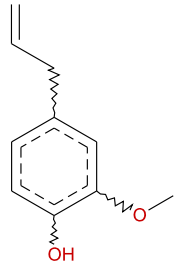
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.432
Unit: g/kg_body_weight
Mahalanobis Distance: 8.71
Mahalanobis Distance p-value: 0.00349

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	MEXACARBATE	METHYLPHENIDATE HYDROCHLORIDE	EUGENOL
Structure			
Actual Endpoint (-log C)	4.07253	3.66897	2.78402
Predicted Endpoint (-log C)	3.7152	3.4956	2.98617
Distance	0.383	0.419	0.431
Reference	NCI/NTP TR-147	NCI/NTP TR-439	NCI/NTP TR-223

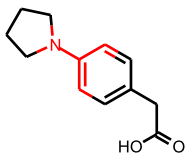
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

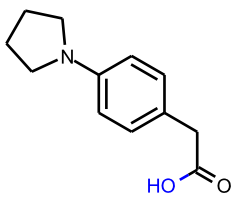
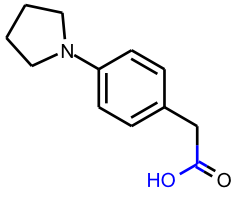
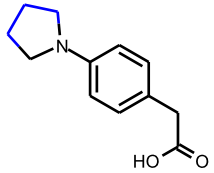
Feature Contribution

Top features for positive contribution

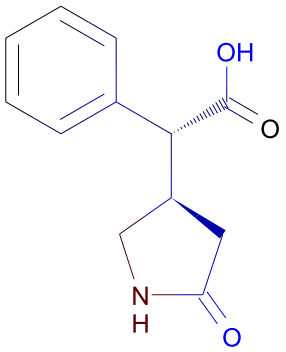
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.0611

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	7	 [*]O	-0.214
FCFP_2	-548632217	 [*]C(=[*])O	-0.119
FCFP_2	-1272798659	 [*]C([*])CC(=[*])[*]	-0.111

AND Enantiomer



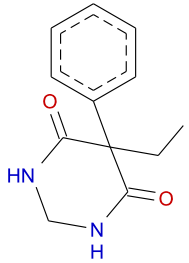
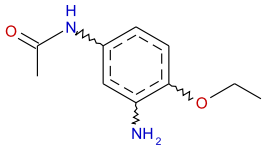
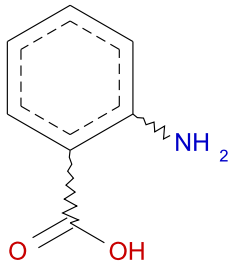
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 0.356
Unit: g/kg_body_weight
Mahalanobis Distance: 9.81
Mahalanobis Distance p-value: 8.82e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	PRIMIDONE (PRIMACLONE)	3-AMINO-4-ETHOXYACETANILIDE	O-ANTHRANILIC ACID
Structure			
Actual Endpoint (-log C)	3.94103	2.45902	2.00682
Predicted Endpoint (-log C)	3.24349	3.20661	2.55301
Distance	0.358	0.392	0.442
Reference	NCI/NTP TR-476	NCI/NTP TR-112	NCI/NTP TR-036

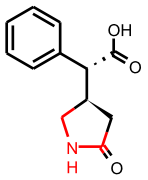
Model Applicability

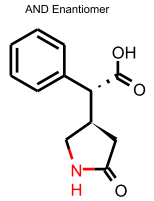
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

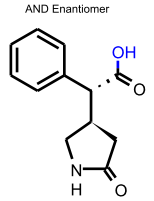
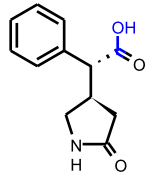
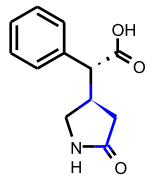
Feature Contribution

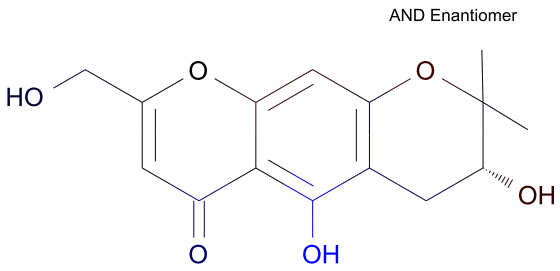
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-885550502	AND Enantiomer  [*]C([*])NC	0.115

FCFP_2	3	AND Enantiomer  [*]O	0.0737

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	AND Enantiomer  [*]O	-0.214
FCFP_2	-548632217	AND Enantiomer  [*]C(=[*])O	-0.119
FCFP_2	-1272798659	AND Enantiomer  [*]C([*])CC(=[*])[*]	-0.111



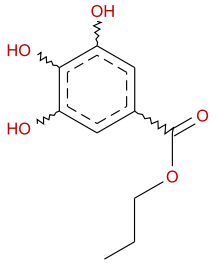
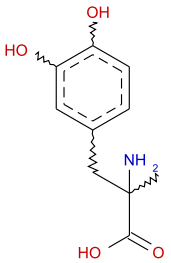
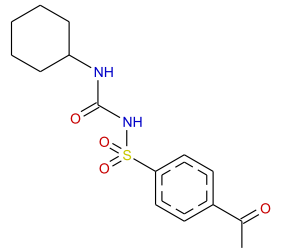
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: 0.321
Unit: g/kg_body_weight
Mahalanobis Distance: 9.19
Mahalanobis Distance p-value: 0.000755

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	PROPYL GALLATE	METHYLDOPA SESQUIHYDRATE	ACETOHEXAMIDE
Structure			
Actual Endpoint (-log C)	2.59435	2.94452	2.55683
Predicted Endpoint (-log C)	2.18569	2.32114	3.62413
Distance	0.546	0.588	0.600
Reference	NCI/NTP TR-240	NCI/NTP TR-348	NCI/NTP TR-050

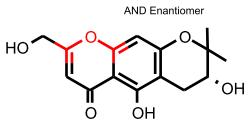
Model Applicability

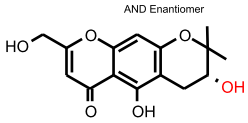
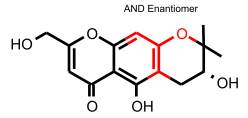
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

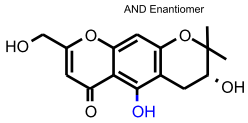
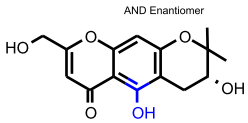
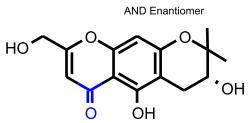
Feature Contribution

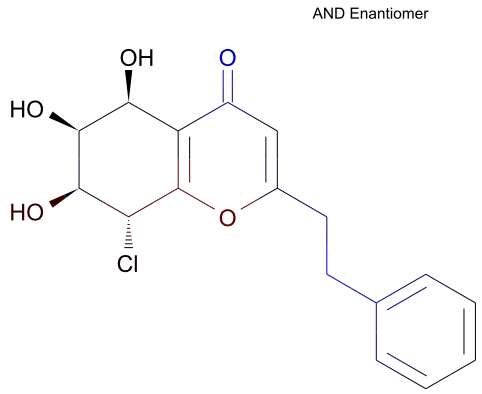
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

FCFP_2	3	 [*]O	0.0737
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c](:[*]):[*]	0.0611

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	 [*]O	-0.214
FCFP_2	-549108873	 [*]:[c](:[*])O	-0.127
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105



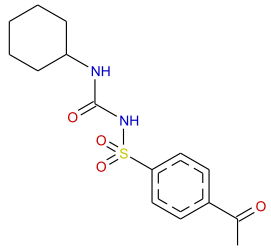
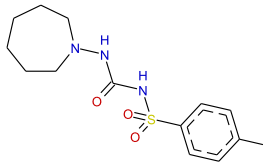
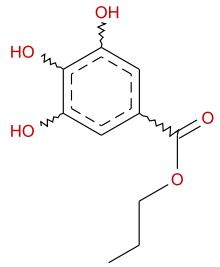
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: 0.137
Unit: g/kg_body_weight
Mahalanobis Distance: 8.4
Mahalanobis Distance p-value: 0.00859

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACETOHEXAMIDE	TOLAZAMIDE	PROPYL GALLATE
Structure			
Actual Endpoint (-log C)	2.55683	2.84011	2.59435
Predicted Endpoint (-log C)	3.62413	3.59315	2.18569
Distance	0.513	0.538	0.559
Reference	NCI/NTP TR-050	NCI/NTP TR-051	NCI/NTP TR-240

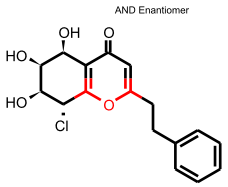
Model Applicability

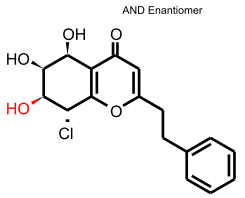
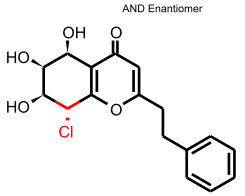
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

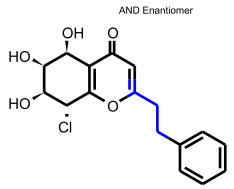
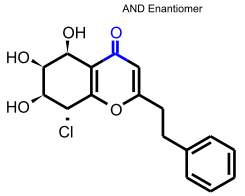
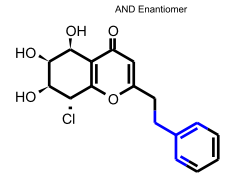
Feature Contribution

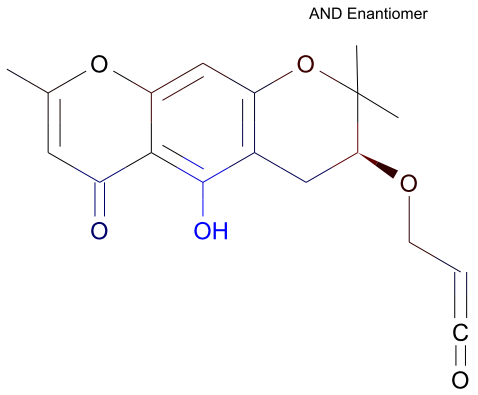
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C(=[*])OC(=[*])[*]	0.095

FCFP_2	3	 [*]O	0.0737
FCFP_2	71953198	 [*]C([*])Cl	0.058

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	 [*]C([*])CC(=[*])[*]	-0.111
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



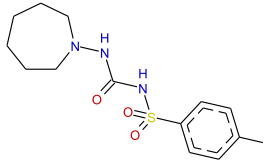
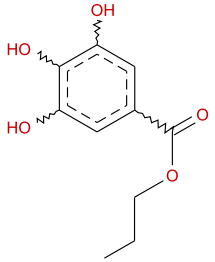
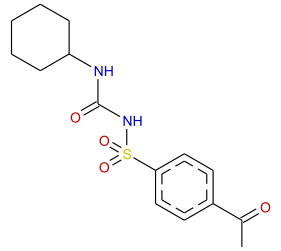
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.158
Unit: g/kg_body_weight
Mahalanobis Distance: 7.93
Mahalanobis Distance p-value: 0.0293

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	TOLAZAMIDE	PROPYL GALLATE	ACETOHEXAMIDE
Structure			
Actual Endpoint (-log C)	2.84011	2.59435	2.55683
Predicted Endpoint (-log C)	3.59315	2.18569	3.62413
Distance	0.476	0.510	0.525
Reference	NCI/NTP TR-051	NCI/NTP TR-240	NCI/NTP TR-050

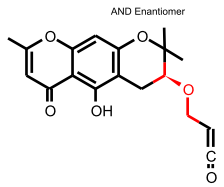
Model Applicability

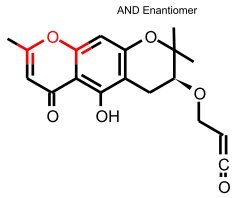
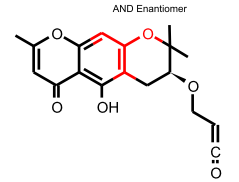
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 444624378: [*]C=C=O

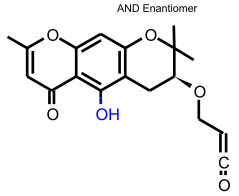
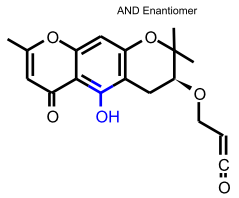
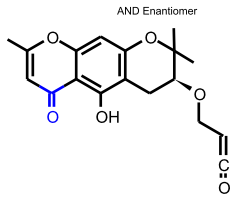
Feature Contribution

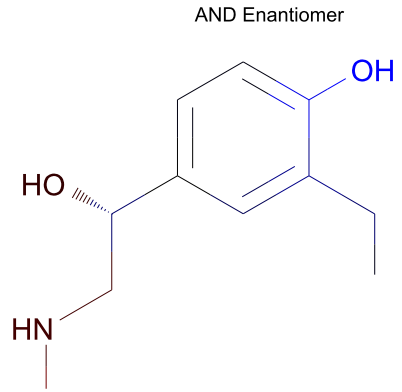
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C(=[*])OC(=[*])[*]	0.095

FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.0611

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	 [*]O	-0.214
FCFP_2	-549108873	 [*]:[c](:[*])O	-0.127
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105



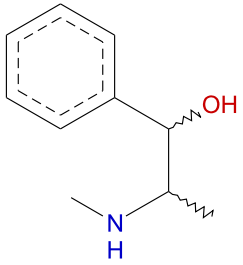
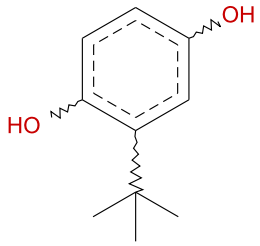
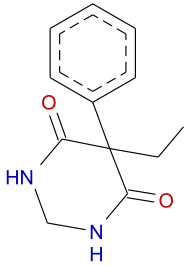
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 0.56
Unit: g/kg_body_weight
Mahalanobis Distance: 8.32
Mahalanobis Distance p-value: 0.0106

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE SULFATE	t-BUTYLHYDROQUINONE	PRIMIDONE (PRIMACLONE)
Structure			
Actual Endpoint (-log C)	4.16695	2.44253	3.94103
Predicted Endpoint (-log C)	3.11455	2.62758	3.24349
Distance	0.414	0.521	0.546
Reference	NCI/NTP TR-307	NCI/NTP TR-459	NCI/NTP TR-476

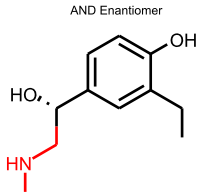
Model Applicability

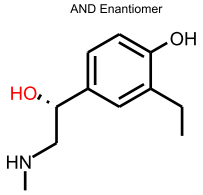
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

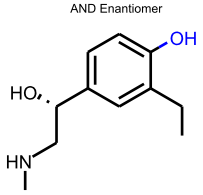
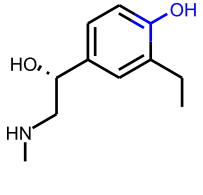
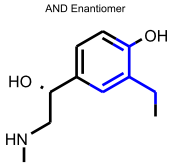
Feature Contribution

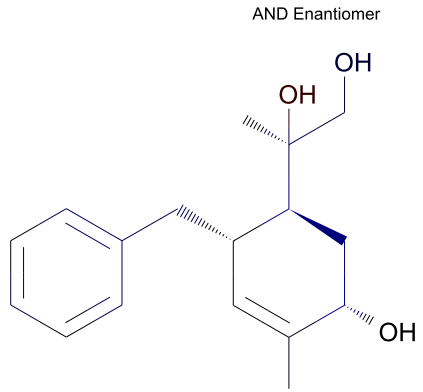
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-885550502	 [*]C([*])NC	0.115

FCFP_2	3	AND Enantiomer  [*]O	0.0737

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	AND Enantiomer  [*]O	-0.214
FCFP_2	-549108873	AND Enantiomer  [*]:[c](:[*])O	-0.127
FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



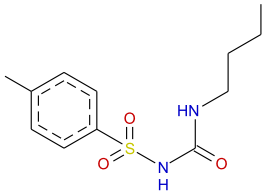
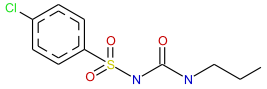
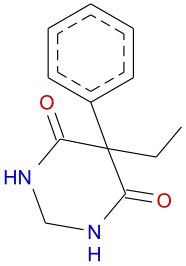
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 0.243
Unit: g/kg_body_weight
Mahalanobis Distance: 8.19
Mahalanobis Distance p-value: 0.0149

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	TOLBUTAMIDE	CHLORPROPAMIDE	PRIMIDONE (PRIMACLONE)
Structure			
Actual Endpoint (-log C)	2.3985	3.0107	3.94103
Predicted Endpoint (-log C)	3.32272	3.18321	3.24349
Distance	0.501	0.509	0.589
Reference	NCI/NTP TR-031	NCI/NTP TR-045	NCI/NTP TR-476

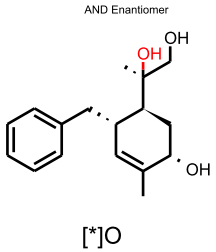
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

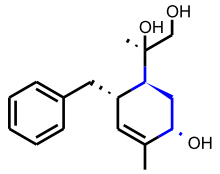
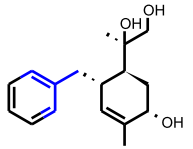
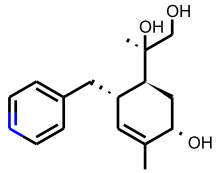
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3	 [*]O	0.0737

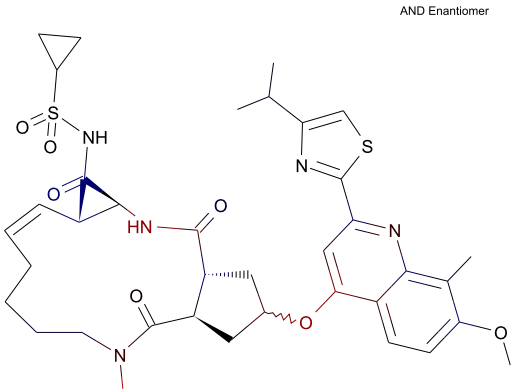
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	-1272798659	AND Enantiomer  [*]C([*])CC(=[*])[*]	-0.111
FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829
FCFP_2	16	AND Enantiomer  [*][c](:[*]):[*]	-0.0512

Simeprevir

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: 0.00297

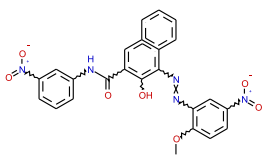
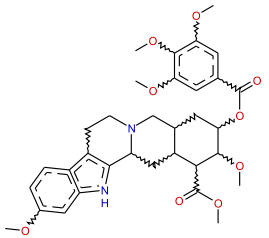
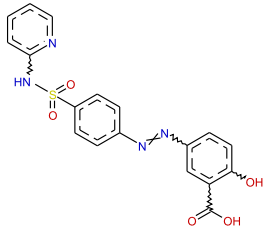
Unit: g/kg_body_weight

Mahalanobis Distance: 14.2

Mahalanobis Distance p-value: 5.46e-013

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	C.I.PIGMENT RED 23	RESERPINE	SALICYLAZOSULFAPYRIDINE
Structure			
Actual Endpoint (-log C)	2.30052	6.13118	3.375
Predicted Endpoint (-log C)	3.55333	4.38304	2.80292
Distance	0.861	0.934	1.035
Reference	NCI/NTP TR-411	NCI/NTP TR-193	NCI/NTP TR-457

Model Applicability

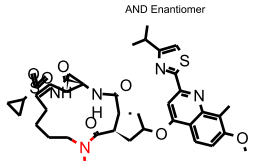
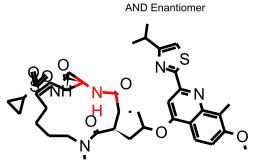
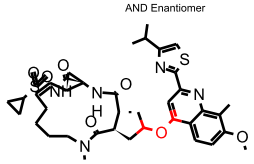
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 749.94. Training min, max, mean, SD: 74.122, 731.95, 245.25, 106.4.
2. OPS PC9 out of range. Value: 5.11. Training min, max, SD, explained variance: -2.8548, 3.3954, 1.263, 0.0360.
3. OPS PC12 out of range. Value: -3.0299. Training min, max, SD, explained variance: -2.364, 2.9228, 1.079, 0.0263.
4. OPS PC14 out of range. Value: 4.236. Training min, max, SD, explained variance: -2.0656, 3.3808, 1.011, 0.0231.
5. Unknown FCFP_2 feature: 690481386: [*]:[c](:[*])[c]1:n:[*]:[*]:s:1
6. Unknown FCFP_2 feature: 1018942292: [*]NS(=O)(=O)C1[*][*]1

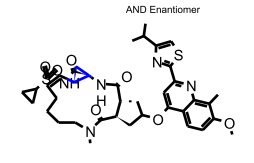
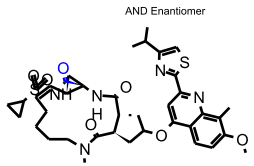
Feature Contribution

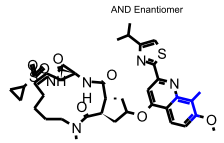
Top features for positive contribution

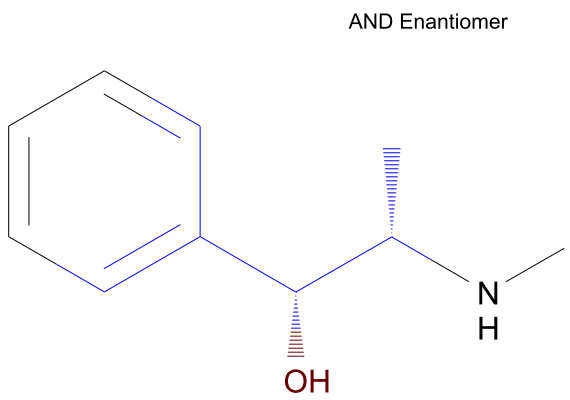
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	136627117	 [*]OC	0.173
FCFP_2	-885550502	 [*]C([*])NC	0.115
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	 [*]C([*])CC(=[*])[*]	-0.111
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105

FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



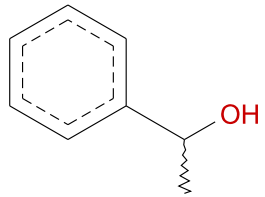
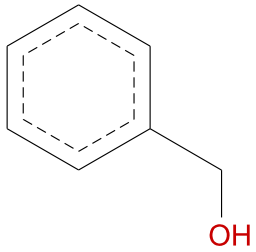
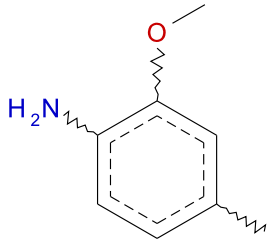
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 1.98
Unit: g/kg_body_weight
Mahalanobis Distance: 5.65
Mahalanobis Distance p-value: 0.281

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ALPHA-METHYLBENZYL ALCOHOL	BENZYL ALCOHOL	M-CRESIDINE
Structure			
Actual Endpoint (-log C)	2.21189	2.43193	5.93318
Predicted Endpoint (-log C)	1.71063	2.09547	3.87056
Distance	0.437	0.558	0.621
Reference	NCI/NTP TR-369	NCI/NTP TR-343	NCI/NTP TR-105

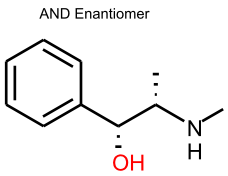
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 136686699: [*]NC

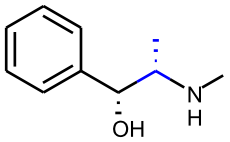
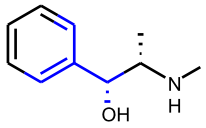
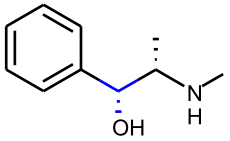
Feature Contribution

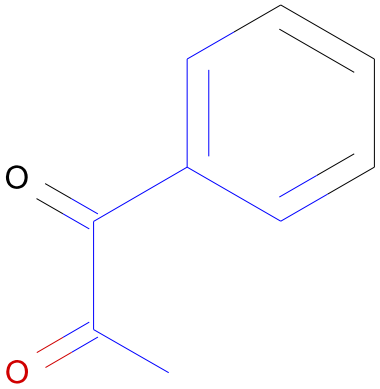
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3	 [*]O	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	136597326	AND Enantiomer  [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	0	AND Enantiomer  [*]C([*])[*]	-0.29



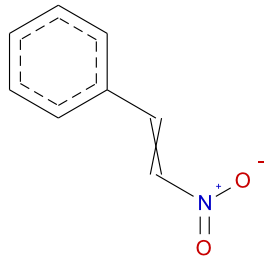
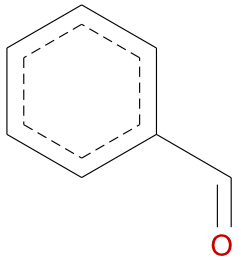
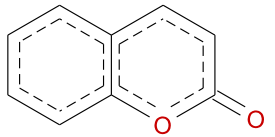
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: 1.39
Unit: g/kg_body_weight
Mahalanobis Distance: 5.05
Mahalanobis Distance p-value: 0.552

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	BETA-NITROSTYRENE	BENZALDEHYDE	COUMARIN
Structure			
Actual Endpoint (-log C)	2.6965	2.38236	3.76684
Predicted Endpoint (-log C)	2.69255	2.25838	3.6624
Distance	0.348	0.400	0.444
Reference	NCI/NTP TR-170	NCI/NTP TR-378	NCI/NTP TR-422

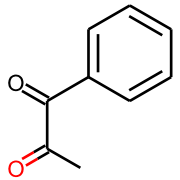
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1549192822: [*]C(=[*])C(=O)[c](:[*]):[*]

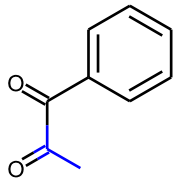
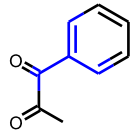
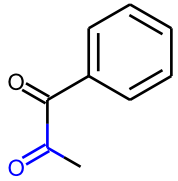
Feature Contribution

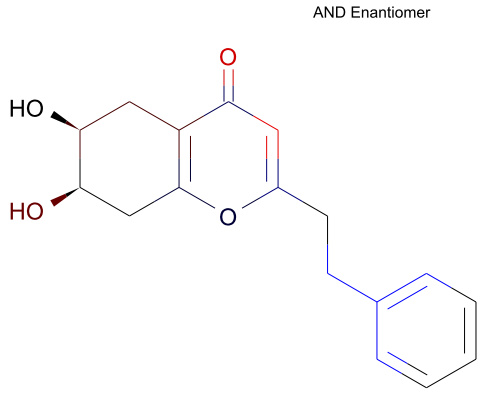
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1	 [*]=O	0.511

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	136597326	 [*]C([*])C	-0.489
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307



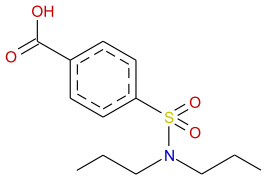
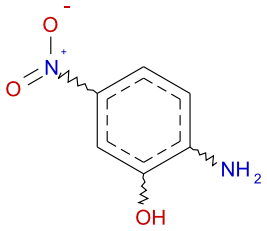
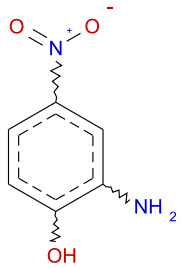
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: 0.0368
Unit: g/kg_body_weight
Mahalanobis Distance: 9.89
Mahalanobis Distance p-value: 8.04e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	PROBENECID	2-AMINO-5-NITROPHENOL	2-AMINO-4-NITROPHENOL
Structure			
Actual Endpoint (-log C)	2.85333	2.88684	2.78993
Predicted Endpoint (-log C)	2.4258	3.28491	3.28491
Distance	0.714	0.734	0.734
Reference	NCI/NTP TR-395	NCI/NTP TR-334	NCI/NTP TR-339

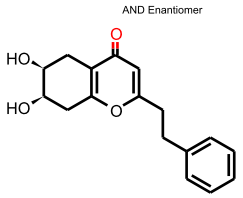
Model Applicability

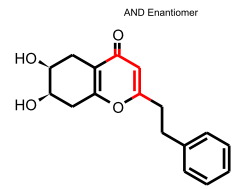
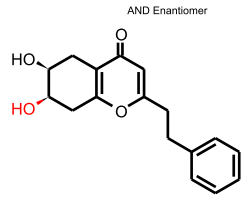
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 436915834: [*]C\C(=C\[*])\O[*]

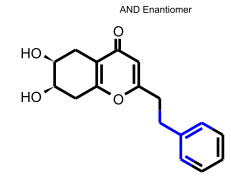
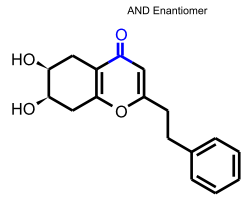
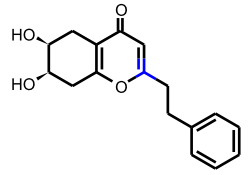
Feature Contribution

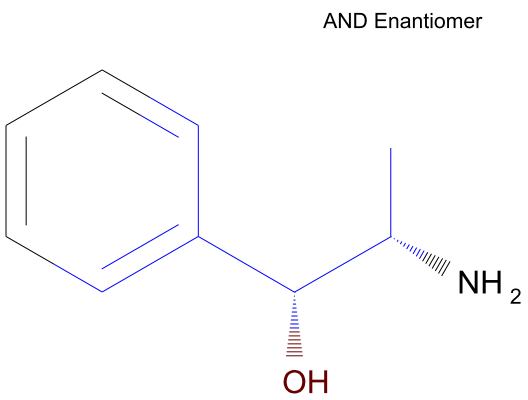
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1	 [*]=O	0.511

FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225
FCFP_2	3	 [*]O	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307
FCFP_2	0	 [*]C([*])[*]	-0.29



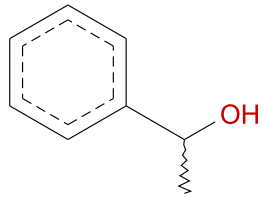
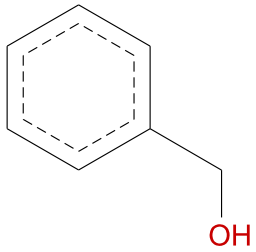
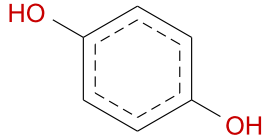
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 3.02
Unit: g/kg_body_weight
Mahalanobis Distance: 5.41
Mahalanobis Distance p-value: 0.384

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ALPHA-METHYLBENZYL ALCOHOL	BENZYL ALCOHOL	HYDROQUINONE
Structure			
Actual Endpoint (-log C)	2.21189	2.43193	3.34286
Predicted Endpoint (-log C)	1.71063	2.09547	3.36172
Distance	0.405	0.548	0.593
Reference	NCI/NTP TR-369	NCI/NTP TR-343	NCI/NTP TR-366

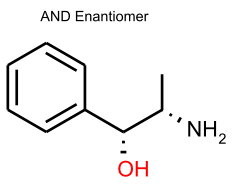
Model Applicability

Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

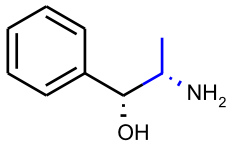
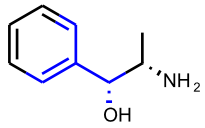
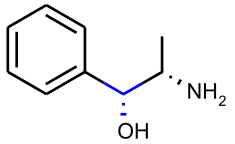
Feature Contribution

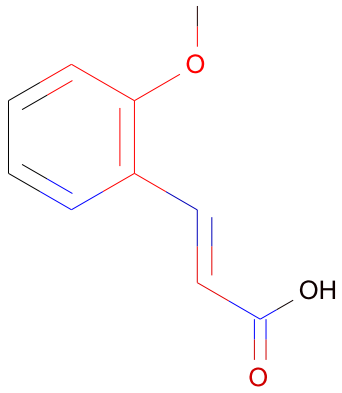
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3	 [*]O	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	136597326	AND Enantiomer  [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	0	AND Enantiomer  [*]C([*])[*]	-0.29



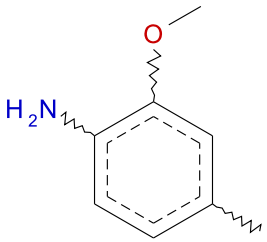
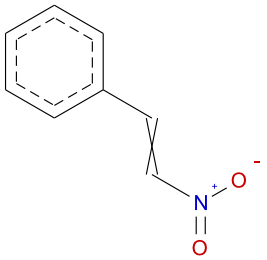
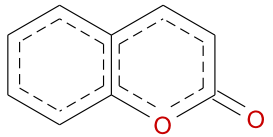
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.0163
Unit: g/kg_body_weight
Mahalanobis Distance: 8.89
Mahalanobis Distance p-value: 0.00017

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	M-CRESIDINE	BETA-NITROSTYRENE	COUMARIN
Structure			
Actual Endpoint (-log C)	5.93318	2.6965	3.76684
Predicted Endpoint (-log C)	3.87056	2.69255	3.6624
Distance	0.482	0.489	0.519
Reference	NCI/NTP TR-105	NCI/NTP TR-170	NCI/NTP TR-422

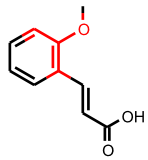
Model Applicability

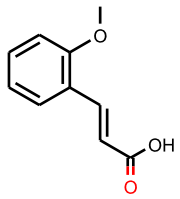
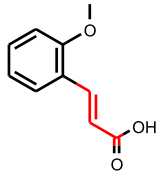
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC9 out of range. Value: 2.9649. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.

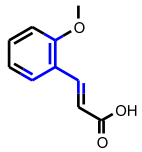
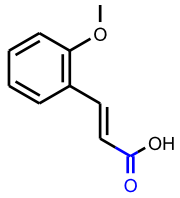
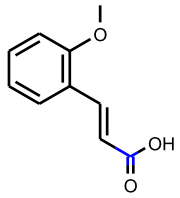
Feature Contribution

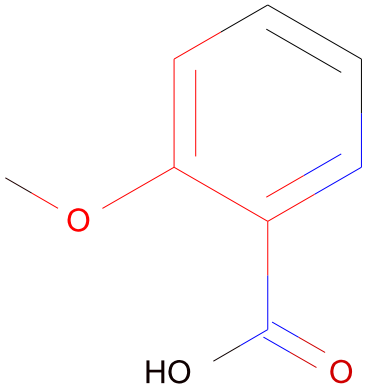
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.672

FCFP_2	1	 [*]=O	0.511
FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307
FCFP_2	0	 [*]C([*])[*]	-0.29



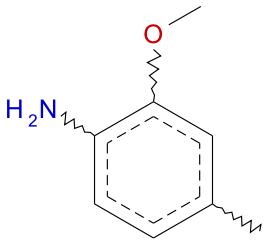
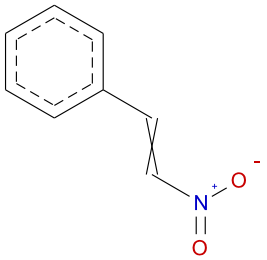
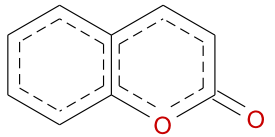
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.0216
Unit: g/kg_body_weight
Mahalanobis Distance: 8.52
Mahalanobis Distance p-value: 0.000497

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	M-CRESIDINE	BETA-NITROSTYRENE	COUMARIN
Structure			
Actual Endpoint (-log C)	5.93318	2.6965	3.76684
Predicted Endpoint (-log C)	3.87056	2.69255	3.6624
Distance	0.397	0.543	0.552
Reference	NCI/NTP TR-105	NCI/NTP TR-170	NCI/NTP TR-422

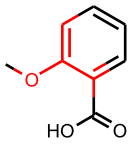
Model Applicability

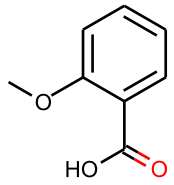
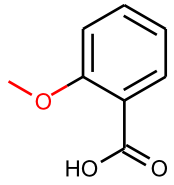
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

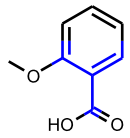
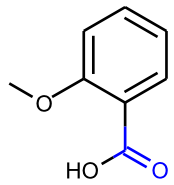
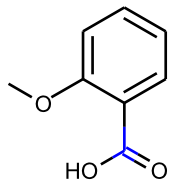
Feature Contribution

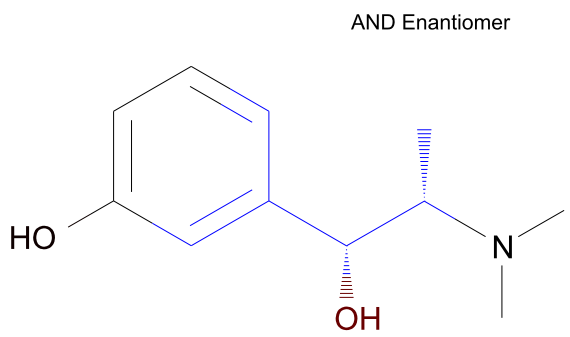
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.672

FCFP_2	1	 [*]=O	0.511
FCFP_2	136627117	 [*]OC	0.0304

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307
FCFP_2	0	 [*]C([*])[*]	-0.29



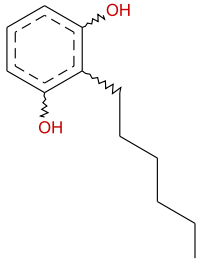
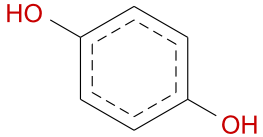
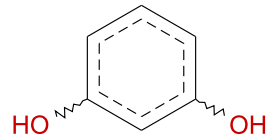
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 0.556
Unit: g/kg_body_weight
Mahalanobis Distance: 7.4
Mahalanobis Distance p-value: 0.00953

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	4-HEXYLRESORCINOL	HYDROQUINONE	RESORCINOL
Structure			
Actual Endpoint (-log C)	3.1915	3.34286	2.99262
Predicted Endpoint (-log C)	2.16134	3.36172	3.36172
Distance	0.592	0.617	0.617
Reference	NCI/NTP TR-330	NCI/NTP TR-366	NCI/NTP TR-403

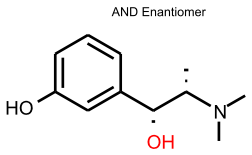
Model Applicability

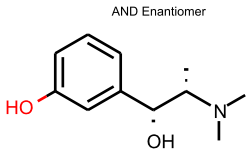
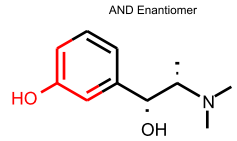
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

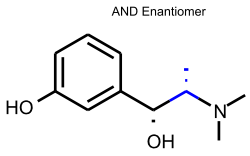
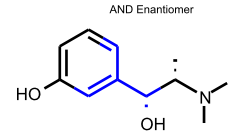
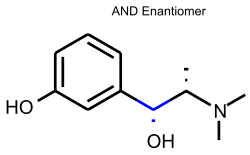
Feature Contribution

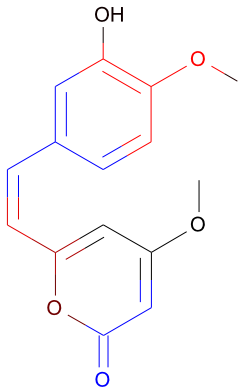
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3	 [*]O	0.104

FCFP_2	7	 [*]O	0.0144
FCFP_2	74595001	 [*]:[cH]:[c](O):[cH]:[*]	0.000246

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C([*])C	-0.489
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	0	 [*]C([*])[*]	-0.29



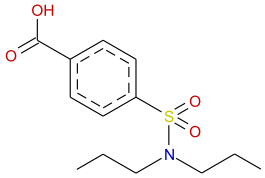
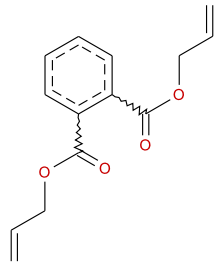
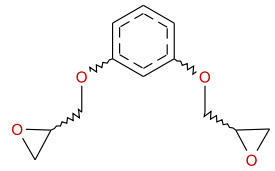
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: 0.00524
Unit: g/kg_body_weight
Mahalanobis Distance: 7.78
Mahalanobis Distance p-value: 0.00366

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	PROBENECID	DIALLYL PHTHALATE	DIGLYCIDYL RESORCINOL ETHER (DGRE)
Structure			
Actual Endpoint (-log C)	2.85333	3.3914	4.99464
Predicted Endpoint (-log C)	2.4258	3.50093	4.37491
Distance	0.642	0.674	0.688
Reference	NCI/NTP TR-395	NCI/NTP TR-284	NCI/NTP TR-257

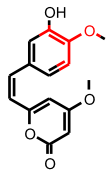
Model Applicability

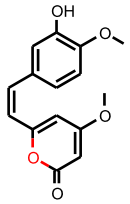
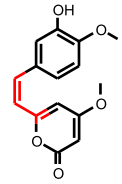
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC9 out of range. Value: 5.1682. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
2. Unknown FCFP_2 feature: 436915834: [*]C\C(=C\[*])\O[*]

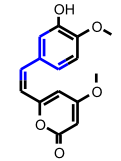
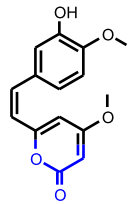
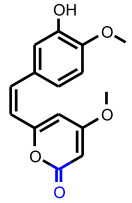
Feature Contribution

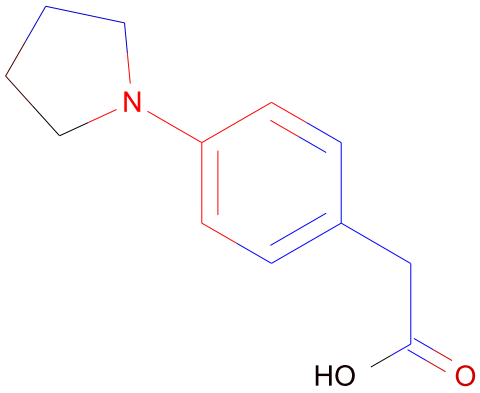
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.672

FCFP_2	1	 [*]=O	0.511
FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	565998553	 [*]OC(=O)C=[*]	-0.348
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307



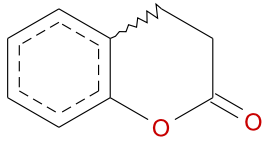
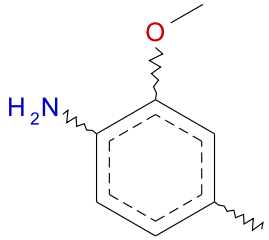
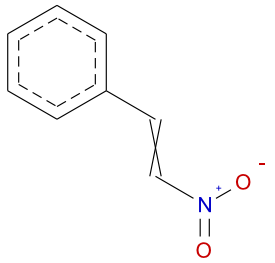
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.038
Unit: g/kg_body_weight
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 4.64e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3,4-DIHYDROCOUMARIN	M-CRESIDINE	BETA-NITROSTYRENE
Structure			
Actual Endpoint (-log C)	2.69361	5.93318	2.6965
Predicted Endpoint (-log C)	3.51534	3.87056	2.69255
Distance	0.567	0.580	0.606
Reference	NCI/NTP TR-423	NCI/NTP TR-105	NCI/NTP TR-170

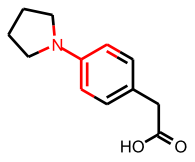
Model Applicability

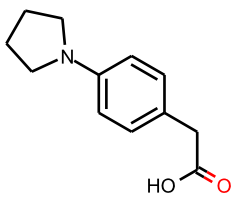
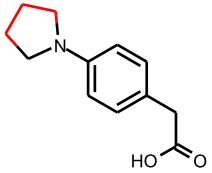
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

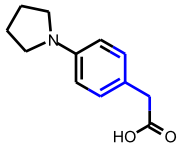
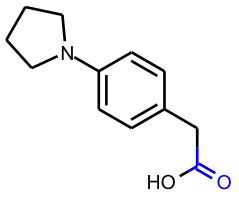
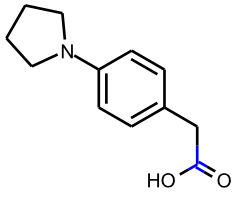
Feature Contribution

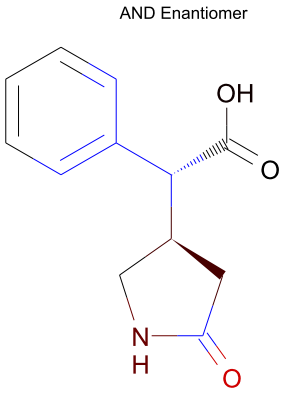
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.672

FCFP_2	1	 [*]=O	0.511
FCFP_2	-1272798659	 [*]C([*])CC(=[*])[*]	0.0703

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307
FCFP_2	0	 [*]C([*])[*]	-0.29



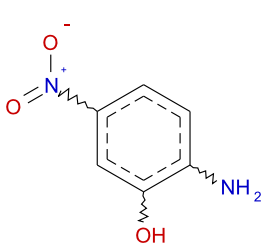
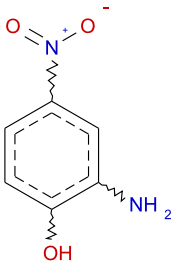
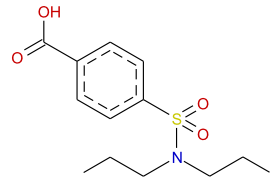
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 0.331
Unit: g/kg_body_weight
Mahalanobis Distance: 7.53
Mahalanobis Distance p-value: 0.00699

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	2-AMINO-5-NITROPHENOL	2-AMINO-4-NITROPHENOL	PROBENECID
Structure			
Actual Endpoint (-log C)	2.88684	2.78993	2.85333
Predicted Endpoint (-log C)	3.28491	3.28491	2.4258
Distance	0.626	0.626	0.729
Reference	NCI/NTP TR-334	NCI/NTP TR-339	NCI/NTP TR-395

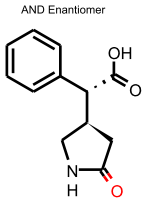
Model Applicability

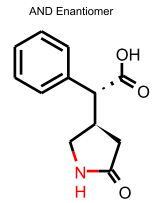
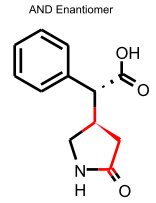
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.

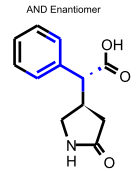
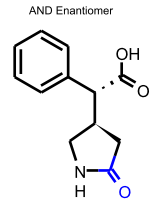
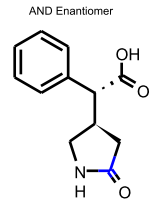
Feature Contribution

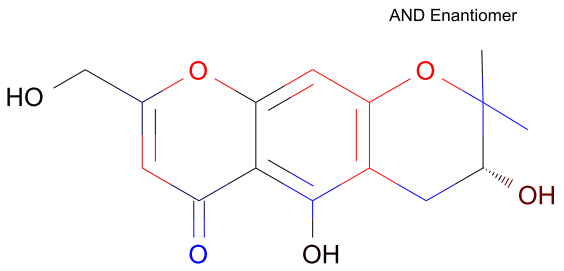
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1	 [*]=O	0.511

FCFP_2	3	 [*]O	0.104
FCFP_2	-1272798659	 [*]C([*])CC(=[*])[*]	0.0703

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307
FCFP_2	0	 [*]C([*])[*]	-0.29



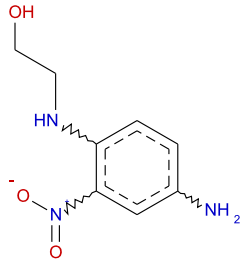
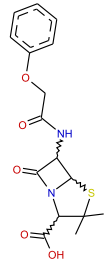
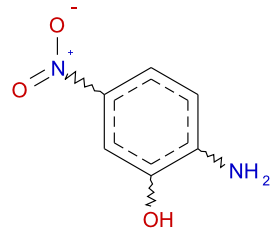
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: 0.00151
Unit: g/kg_body_weight
Mahalanobis Distance: 8.59
Mahalanobis Distance p-value: 0.000403

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	HC RED 3	PENICILLIN VK	2-AMINO-5-NITROPHENOL
Structure			
Actual Endpoint (-log C)	2.59592	2.54455	2.88684
Predicted Endpoint (-log C)	3.285	3.9702	3.28491
Distance	0.643	0.690	0.715
Reference	NCI/NTP TR-281	NCI/NTP TR-336	NCI/NTP TR-334

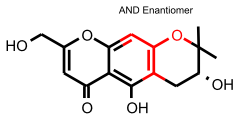
Model Applicability

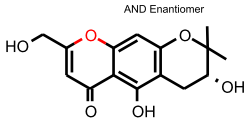
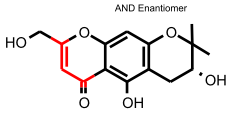
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 436915834: [*]C\C(=C\[*])\O[*]
3. Unknown FCFP_2 feature: -1549192822: [*]C(=[*])C(=O)[c](:[*]):[*]

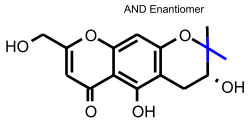
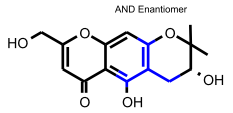
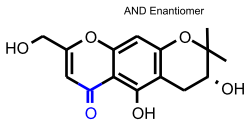
Feature Contribution

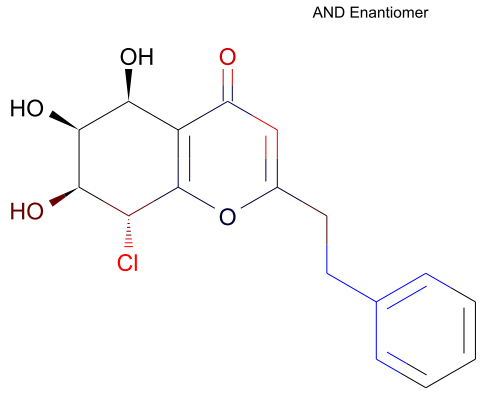
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c](:[*]):[*]	0.672

FCFP_2	1	 [*]=O	0.511
FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C([*])C	-0.489
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307



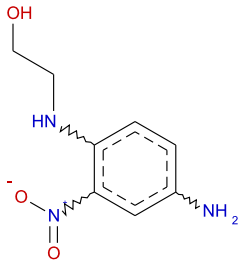
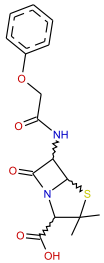
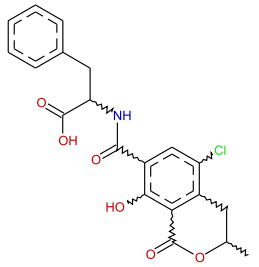
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: 0.00357
Unit: g/kg_body_weight
Mahalanobis Distance: 13.7
Mahalanobis Distance p-value: 3.71e-011

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	HC RED 3	PENICILLIN VK	OCHRATOXIN
Structure			
Actual Endpoint (-log C)	2.59592	2.54455	6.28396
Predicted Endpoint (-log C)	3.285	3.9702	5.12358
Distance	0.690	0.721	0.757
Reference	NCI/NTP TR-281	NCI/NTP TR-336	NCI/NTP TR-358

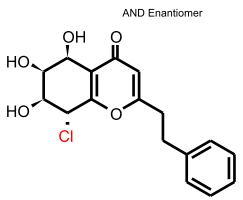
Model Applicability

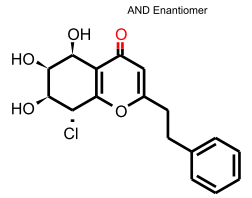
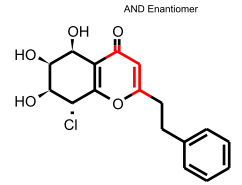
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 436915834: [*]C\C(=C\[*])\O[*]

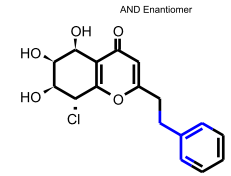
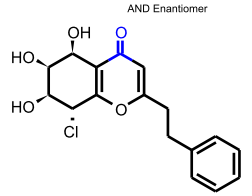
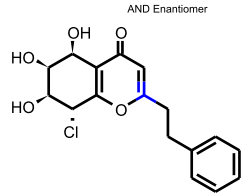
Feature Contribution

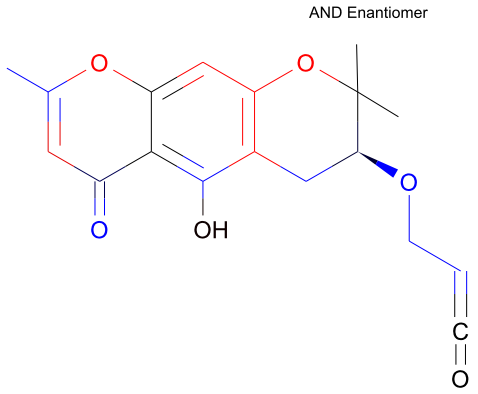
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	32	 [*]Cl	0.526

FCFP_2	1	 [*]=O	0.511
FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*])):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307
FCFP_2	0	 [*]C([*])[*]	-0.29



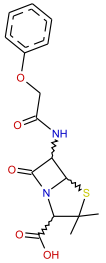
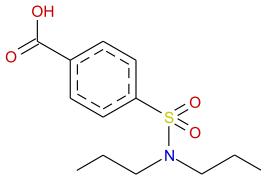
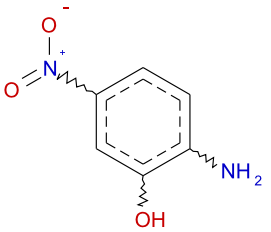
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.0136
Unit: g/kg_body_weight
Mahalanobis Distance: 12.3
Mahalanobis Distance p-value: 3.13e-009

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	PENICILLIN VK	PROBENECID	2-AMINO-5-NITROPHENOL
Structure			
Actual Endpoint (-log C)	2.54455	2.85333	2.88684
Predicted Endpoint (-log C)	3.9702	2.4258	3.28491
Distance	0.629	0.726	0.789
Reference	NCI/NTP TR-336	NCI/NTP TR-395	NCI/NTP TR-334

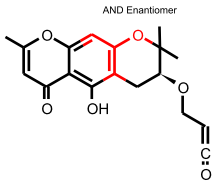
Model Applicability

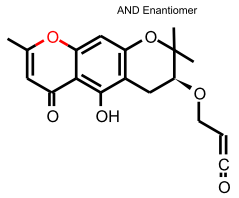
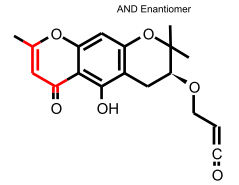
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC9 out of range. Value: 3.0934. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
2. Unknown FCFP_2 feature: 436915834: [*]C\C(=C\[*])\O[*]
3. Unknown FCFP_2 feature: -1549192822: [*]C(=[*])C(=O)[c](:[*]):[*]
4. Unknown FCFP_2 feature: 444624378: [*]C=C=O

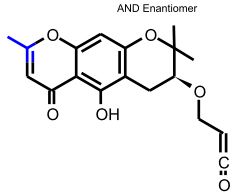
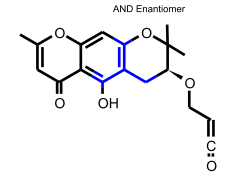
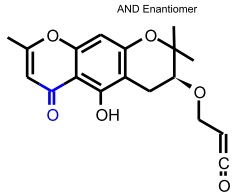
Feature Contribution

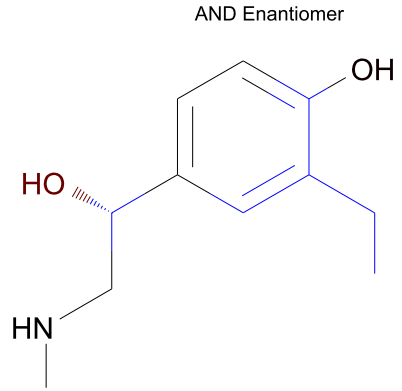
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.672

FCFP_2	1	 [*]=O	0.511
FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C([*])C	-0.489
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307



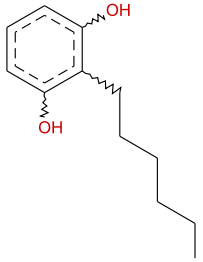
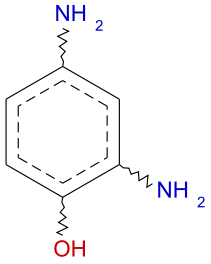
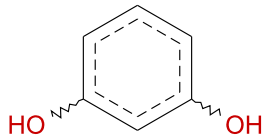
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.557
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 0.91
Unit: g/kg_body_weight
Mahalanobis Distance: 6.08
Mahalanobis Distance p-value: 0.148

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	4-HEXYLRESORCINOL	2,4-DIAMINOPHENOL DIHYDROCHLORIDE	RESORCINOL
Structure			
Actual Endpoint (-log C)	3.1915	3.69598	2.99262
Predicted Endpoint (-log C)	2.16134	3.37438	3.36172
Distance	0.600	0.717	0.735
Reference	NCI/NTP TR-330	NCI/NTP TR-401	NCI/NTP TR-403

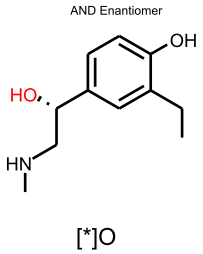
Model Applicability

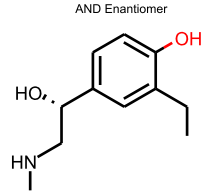
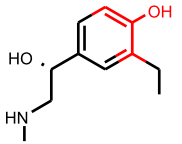
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 136686699: [*]NC

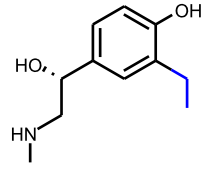
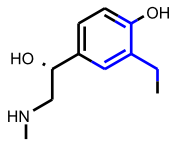
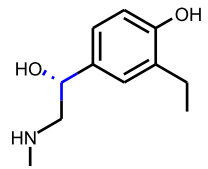
Feature Contribution

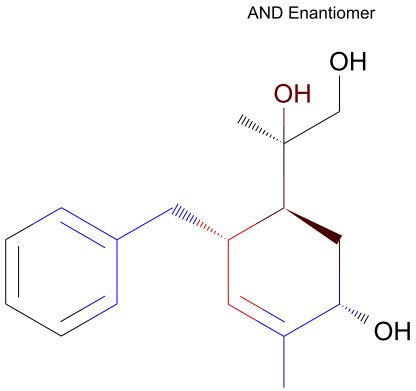
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3		0.104

FCFP_2	7	AND Enantiomer  [*]O	0.0144
FCFP_2	74595001	AND Enantiomer  [*]:[cH]:[c](O):[cH]: [*]	0.000246

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantiomer  [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	0	AND Enantiomer  [*]C([*])[*]	-0.29



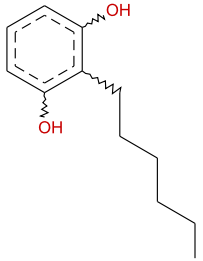
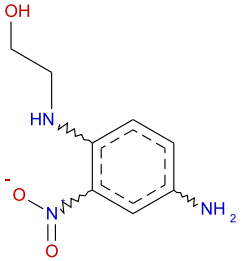
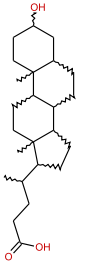
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 0.59
Unit: g/kg_body_weight
Mahalanobis Distance: 9.11
Mahalanobis Distance p-value: 8.71e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	4-HEXYLRESORCINOL	HC RED 3	LITHOCHLOLIC ACID
Structure			
Actual Endpoint (-log C)	3.1915	2.59592	2.87689
Predicted Endpoint (-log C)	2.16134	3.285	3.8262
Distance	0.707	0.808	0.847
Reference	NCI/NTP TR-330	NCI/NTP TR-281	NCI/NTP TR-175

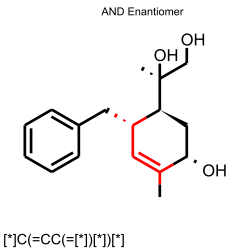
Model Applicability

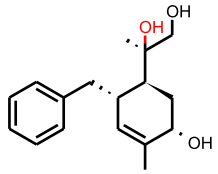
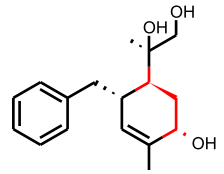
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -415156552: [*]C[C@](C)(O)C([*])[*]

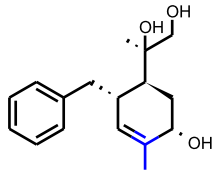
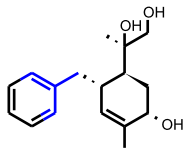
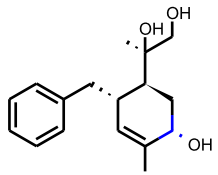
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225

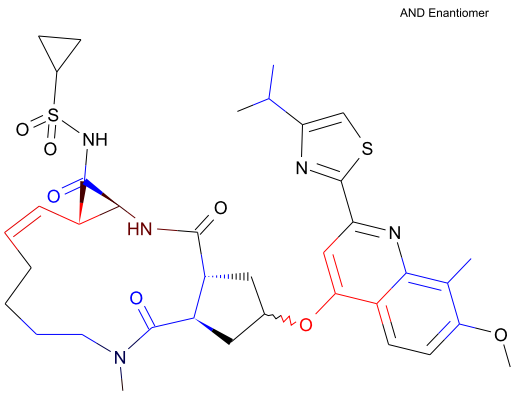
FCFP_2	3	AND Enantiomer  [*]O	0.104
FCFP_2	-1272798659	AND Enantiomer  [*]C([*])CC(=[*])[*]	0.0703

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantiomer  [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantiomer  [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	0	AND Enantiomer  [*]C([*])[*]	-0.29

Simeprevir

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Model Prediction

Prediction: 0.000856

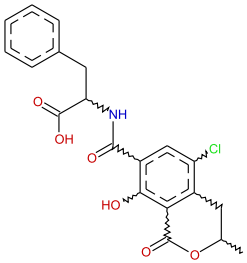
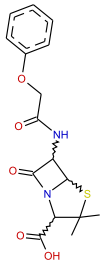
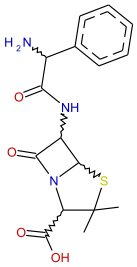
Unit: g/kg_body_weight

Mahalanobis Distance: 16.8

Mahalanobis Distance p-value: 1.77e-015

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	OCHRATOXIN	PENICILLIN VK	AMPICILLIN TRIHYDRATE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.36724
Predicted Endpoint (-log C)	5.12358	3.9702	2.27651
Distance	1.425	1.545	1.716
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-318

Model Applicability

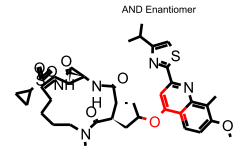
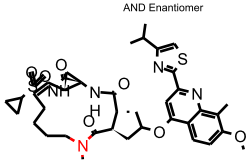
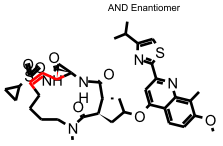
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 749.94. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 9. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. Molecular_PolarSASA out of range. Value: 272.1. Training min, max, mean, SD: 0, 223.97, 50.816, 55.15.
5. Molecular_PolarSurfaceArea out of range. Value: 193.5. Training min, max, mean, SD: 0, 138.03, 28.978, 32.1.
6. OPS PC1 out of range. Value: 9.6287. Training min, max, SD, explained variance: -4.0008, 7.9165, 2.861, 0.2531.
7. OPS PC6 out of range. Value: -3.4118. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
8. Unknown FCFP_2 feature: -415156552: [*]C[C@](C)(O)C([*])[*]
9. Unknown FCFP_2 feature: 690511177: [*]:[cH]:[c](:n:[*])[c](:[*]):[*]
10. Unknown FCFP_2 feature: 690481386: [*]:[c](:[*])[c]1:n:[*]:[*]:s:1
11. Unknown FCFP_2 feature: 1018942292: [*]NS(=O)(=O)C1[*][*]1

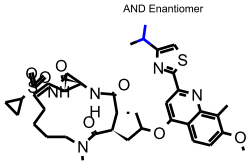
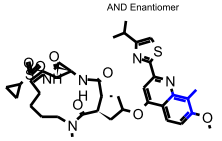
Feature Contribution

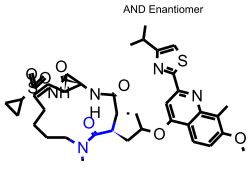
Top features for positive contribution

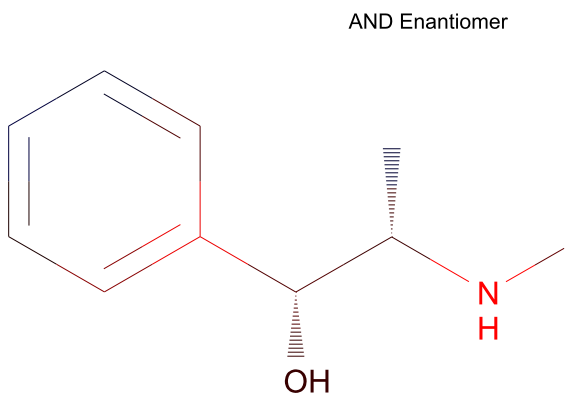
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.672
FCFP_2	1	 [*]=O	0.511
FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C([*])C	-0.489
FCFP_2	203677720	 [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	565998553	 [*]OC(=O)C=[*]	-0.348



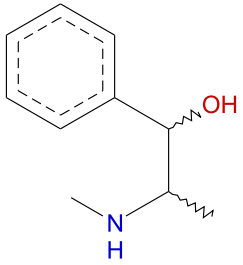
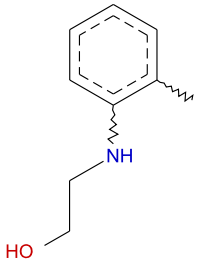
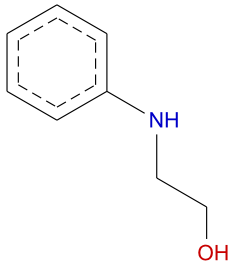
$C_{10}H_{15}NO$
Molecular Weight: 165.232
ALogP: 1.234
Rotatable Bonds: 3
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 0.679
Unit: g/kg_body_weight
Mahalanobis Distance: 15.3
Mahalanobis Distance p-value: 0.635

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE	ANILINE; N-HYDROXYETHYL-2-METHYL	p-AMINOPHENYLETHANOL
Structure			
Actual Endpoint (-log C)	2.399	1.837	1.789
Predicted Endpoint (-log C)	2.38651	2.16262	1.93606
Distance	0.000	0.405	0.416
Reference	JPMSAE 60;1523;71	JIHTAB 31;60;49	UCDS** 8/21/61

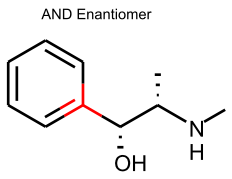
Model Applicability

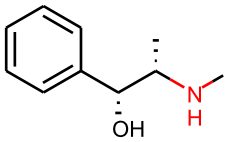
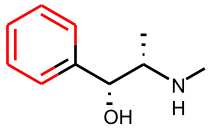
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1186393305: [*]C([*])[C@H](O)[c](:[*]):[*]
4. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
5. Unknown FCFP_6 feature: 136686699: [*]NC

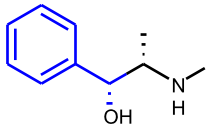
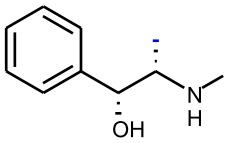
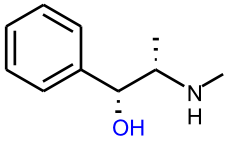
Feature Contribution

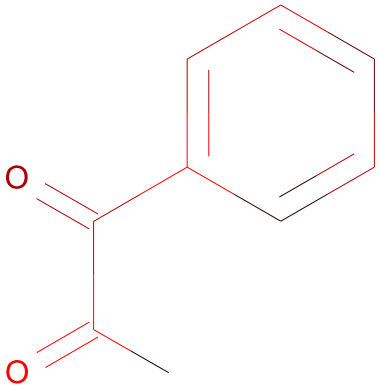
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

ECFP_6	-1897341097	AND Enantiomer  [*]N[*]	0.216
ECFP_6	1571214559	AND Enantiomer  [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2014710090	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.225
ECFP_6	734603939	AND Enantiomer  [*]C	-0.201
FCFP_6	3	AND Enantiomer  [*]O	-0.107



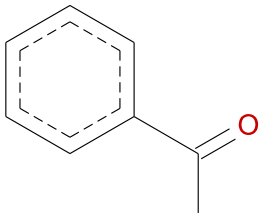
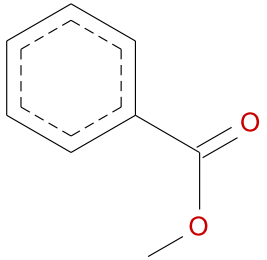
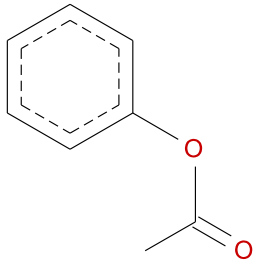
$C_9H_8O_2$
Molecular Weight: 148.159
ALogP: 1.446
Rotatable Bonds: 2
Acceptors: 2
Donors: 0

Model Prediction

Prediction: 0.804
Unit: g/kg_body_weight
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 1

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACETOPHENONE	METHYLBENZOATE	PHENYL ACETATE
Structure			
Actual Endpoint (-log C)	2.169	2.063	1.922
Predicted Endpoint (-log C)	2.12357	1.70515	2.03223
Distance	0.334	0.363	0.390
Reference	GTPZAB 26(8);53;82	FAVUAI 18;69;86	AIHAAP 30;470;69

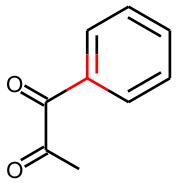
Model Applicability

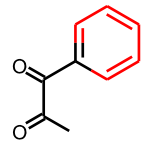
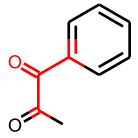
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

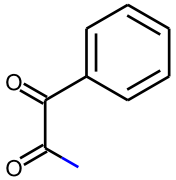
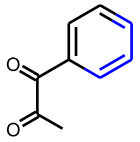
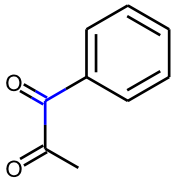
Feature Contribution

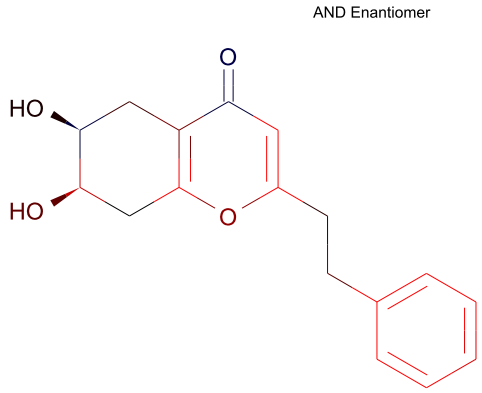
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19
FCFP_6	-1549192822	 [*]C(=[*])C(=O)[c](:[*]):[*]	0.168

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	734603939	 [*]C	-0.201
ECFP_6	1997021792	 [*]:[cH]:[cH]:[cH]:[*]	-0.0912
FCFP_6	0	 [*]C([*])[*]	-0.0791



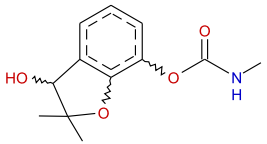
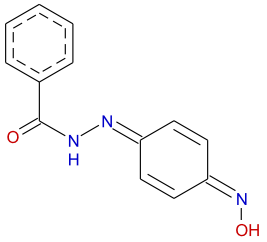
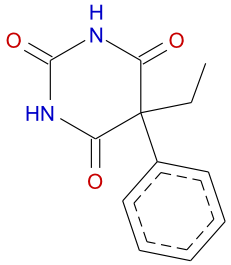
$C_{17}H_{18}O_4$
Molecular Weight: 286.322
ALogP: 1.568
Rotatable Bonds: 3
Acceptors: 4
Donors: 2

Model Prediction

Prediction: 0.366
Unit: g/kg_body_weight
Mahalanobis Distance: 21.6
Mahalanobis Distance p-value: 1.44e-014

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	CARBAMIC ACID; METHYL-; 2;3-DIHYDRO-2;2-DIMETHYL-3-HYDROXY-7-BENZOFURANYL ESTER	BENZOIC ACID; (4-OXO-2;5-CYCLOHEXADIEN-1-YLIDENE)HYDRAZIDE; OXIME	PHENOBARBITAL
Structure			
Actual Endpoint (-log C)	4.12	3.382	3.156
Predicted Endpoint (-log C)	3.7204	2.73172	2.85686
Distance	0.513	0.524	0.546
Reference	NTIS** PB85-143766	FMCHA2 -;C63;89	TXAPA9 18;185;71

Model Applicability

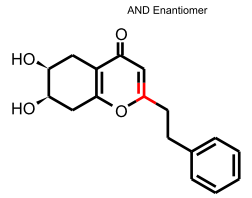
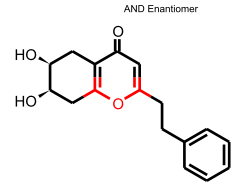
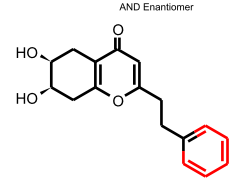
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1650944136: [*]CC(=C([*])[*])O[*]
3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
4. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

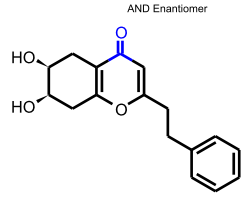
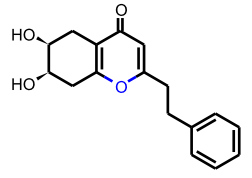
Feature Contribution

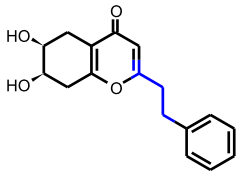
Top features for positive contribution

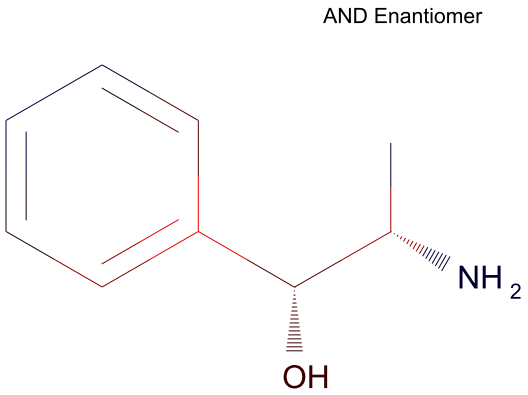
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-560785749	 [*]C(=[*])OC(=[*])[*]	0.259
ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	683445015	 [*]O[*]	-0.266

ECFP_6	-1795525632	AND Enantiomer  [*]CCC(=[*])[*]	-0.176



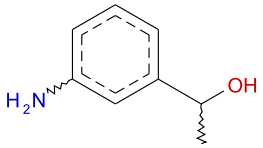
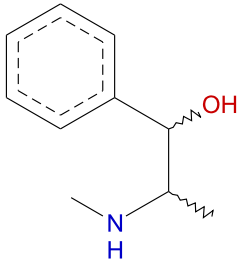
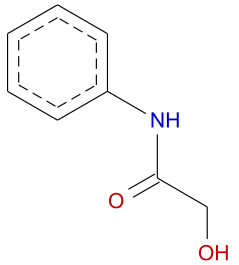
$C_9H_{13}NO$
Molecular Weight: 151.206
ALogP: 0.802
Rotatable Bonds: 2
Acceptors: 2
Donors: 2

Model Prediction

Prediction: 1.18
Unit: g/kg_body_weight
Mahalanobis Distance: 13.1
Mahalanobis Distance p-value: 1

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3-AMINO-2-METHYLBENZYL ALCOHOL	EPHEDRINE	GLYCOLANILIDE
Structure			
Actual Endpoint (-log C)	1.646	2.399	1.949
Predicted Endpoint (-log C)	2.29496	2.38651	1.84865
Distance	0.310	0.314	0.430
Reference	AMIHBC 10;61;54	JPMSAE 60;1523;71	JAPMA8 35;50;46

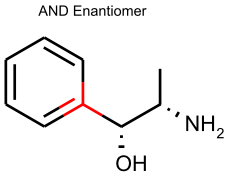
Model Applicability

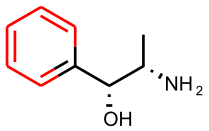
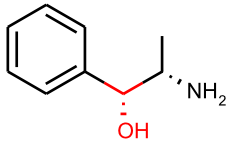
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1186393305: [*]C([*])[C@H](O)[c](:[*]):[*]
4. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

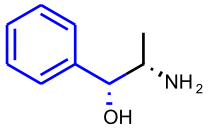
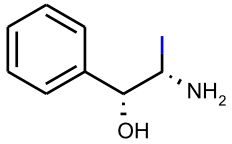
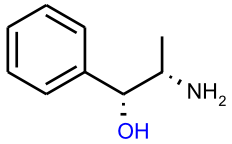
Feature Contribution

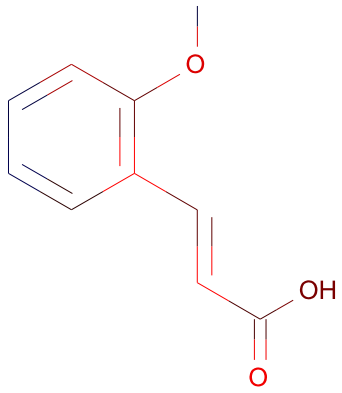
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

ECFP_6	1571214559	AND Enantiomer  [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19
FCFP_6	1070061035	AND Enantiomer  [*]C([*])O	0.0996

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2014710090	AND Enantiomer  [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.225
ECFP_6	734603939	AND Enantiomer  [*]C	-0.201
FCFP_6	3	AND Enantiomer  [*]O	-0.107



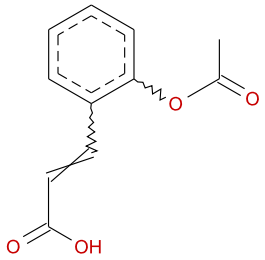
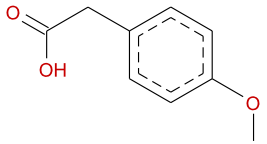
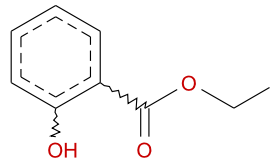
$C_{10}H_{10}O_3$
Molecular Weight: 178.185
ALogP: 1.911
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.976
Unit: g/kg_body_weight
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 1

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	CINNAMIC ACID; 2-ACETOXY-	p-METHOXYPHENYLACETIC ACID	ETHYL SALICYLATE
Structure			
Actual Endpoint (-log C)	1.816	2.03	2.1
Predicted Endpoint (-log C)	2.18026	2.31007	1.7354
Distance	0.290	0.325	0.337
Reference	JPETAB 99;450;50	FCTXAV 14;685;76	FCTXAV 16;751;78

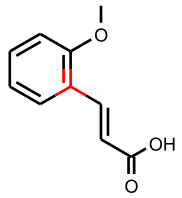
Model Applicability

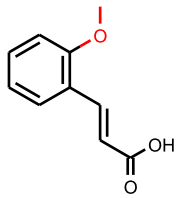
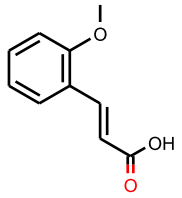
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 451371068: [*]\C=C\[c](:[*]):[*]
4. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

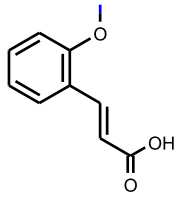
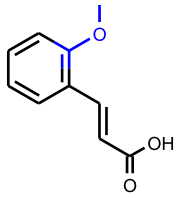
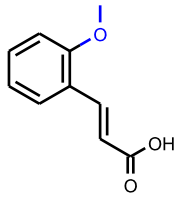
Feature Contribution

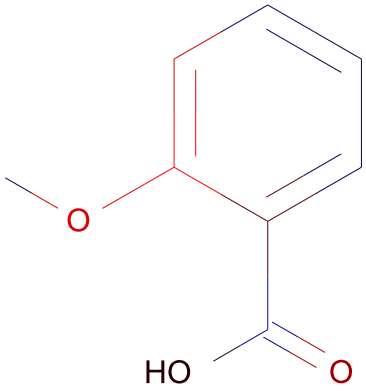
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

FCFP_6	136627117	 [*]OC	0.17
ECFP_6	-1074141656	 [*]=O	0.142

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	734603939	 [*]C	-0.201
FCFP_6	1036089772	 [*]:[c](:[*])OC	-0.136
ECFP_6	864909220	 [*]OC	-0.119



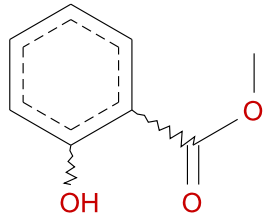
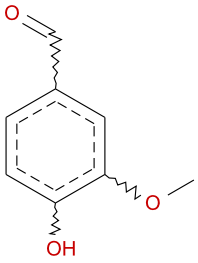
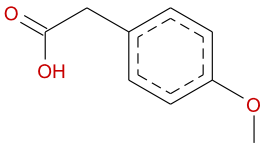
$C_8H_8O_3$
Molecular Weight: 152.147
ALogP: 1.443
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 1.1
Unit: g/kg_body_weight
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 1

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	METHYL SALICYLATE	VANILLIN	p-METHOXYPHENYLACETIC ACID
Structure			
Actual Endpoint (-log C)	2.234	1.984	2.03
Predicted Endpoint (-log C)	1.90826	1.65955	2.31007
Distance	0.226	0.278	0.321
Reference	FCTXAV 2;327;64	FCTXAV 2;327;64	FCTXAV 14;685;76

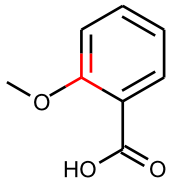
Model Applicability

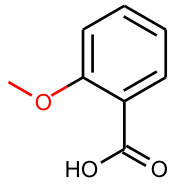
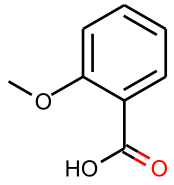
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

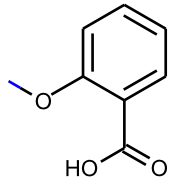
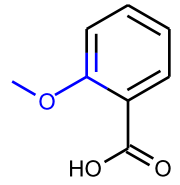
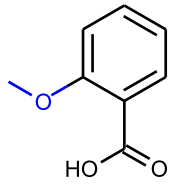
1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
4. Unknown FCFP_6 feature: -1549222613: [*]:[c](:[*])C(=O)O

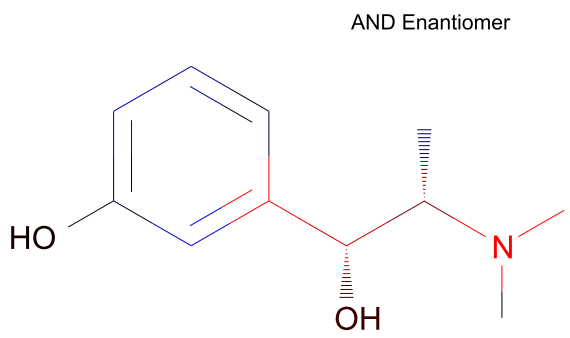
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

FCFP_6	136627117	 [*]OC	0.17
ECFP_6	-1074141656	 [*]=O	0.142

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
ECFP_6	734603939	 [*]C	-0.201
FCFP_6	1036089772	 [*]:[c](:[*])OC	-0.136
ECFP_6	864909220	 [*]OC	-0.119



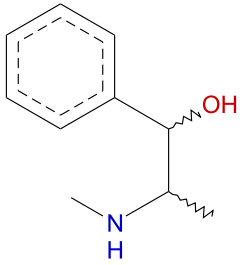
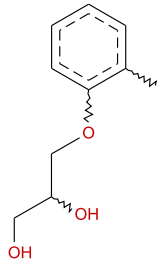
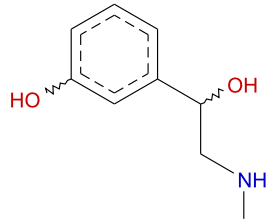
$C_{11}H_{17}NO_2$
Molecular Weight: 195.258
ALogP: 1.528
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 0.712
Unit: g/kg_body_weight
Mahalanobis Distance: 18.1
Mahalanobis Distance p-value: 0.000499

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	EPHEDRINE	3-(2-TOLYLOXY)1;2-PROPANEDIOL	PHENYLEPHRINE
Structure			
Actual Endpoint (-log C)	2.399	2.465	2.679
Predicted Endpoint (-log C)	2.38651	2.03752	2.59876
Distance	0.449	0.469	0.475
Reference	JPMSAE 60;1523;71	AIPTAK 130;280;61	AIPTAK 180;155;69

Model Applicability

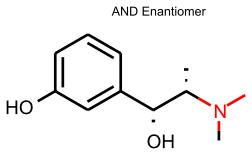
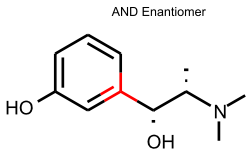
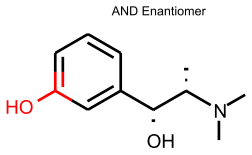
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1280034494: [*]C([*])[C@H](C)N([*])[*]
3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
4. Unknown FCFP_6 feature: 1186393305: [*]C([*])[C@H](O)[c](:[*]):[*]
5. Unknown FCFP_6 feature: -1946918893: [*]C([*])[C@H](C)N([*])[*]
6. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
7. Unknown FCFP_6 feature: 74595001: [*]:[cH]:[c](O):[cH]:[*]
8. Unknown FCFP_6 feature: -549108873: [*]:[c](:[*])O

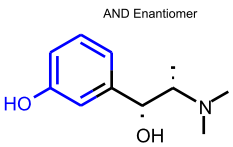
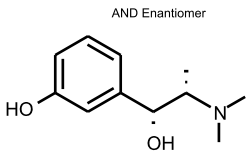
Feature Contribution

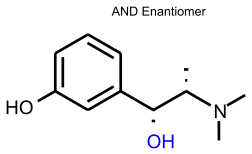
Top features for positive contribution

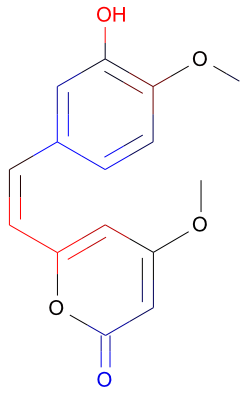
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	865379614	 [*]N([*])C	0.322
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	2019062761	 [*]:[c](:[*])O	0.138

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	946589555	 O[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.204
ECFP_6	734603939	 [*]C	-0.201

FCFP_6	3	AND Enantiomer  [*]O	-0.107



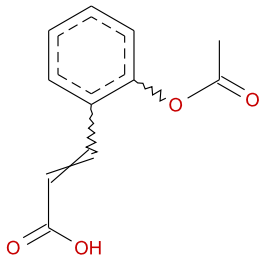
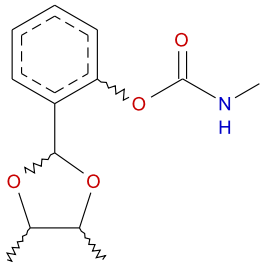
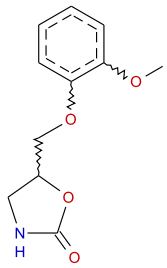
$C_{15}H_{14}O_5$
Molecular Weight: 274.269
ALogP: 1.969
Rotatable Bonds: 4
Acceptors: 5
Donors: 1

Model Prediction

Prediction: 0.765
Unit: g/kg_body_weight
Mahalanobis Distance: 17.1
Mahalanobis Distance p-value: 0.02

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	CINNAMIC ACID; 2-ACETOXY-	CARBAMIC ACID; METHYL-; 2-(4;5-DIMETHYL-1;3-DIOXOLAN-2-YL)PHENYL ESTER	5-[(o-METHOXYPHENOXY)-METHYL]-2-OXAZOLIDINONE
Structure			
Actual Endpoint (-log C)	1.816	3.359	1.767
Predicted Endpoint (-log C)	2.18026	3.85288	2.07191
Distance	0.493	0.494	0.532
Reference	JPETAB 99;450;50	FMCHA2 -;C256;89	TXAPA9 6;642;64

Model Applicability

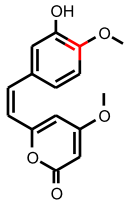
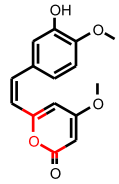
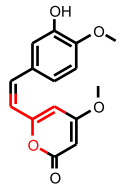
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1053980253: [*]O\C(=C/[*])\C=[*]
3. Unknown ECFP_2 feature: -444332269: [*]O\C(=C\[*])\C=[*]
4. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
5. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
6. Unknown FCFP_6 feature: 74595001: [*]:[cH]:[c](O):[cH]:[*]
7. Unknown FCFP_6 feature: -549108873: [*]:[c](:[*])O
8. Unknown FCFP_6 feature: 451371068: [*]\C=C\[c](:[*]):[*]

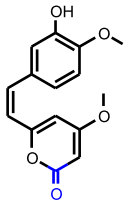
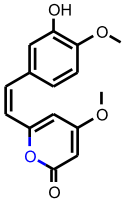
Feature Contribution

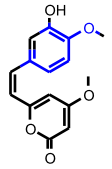
Top features for positive contribution

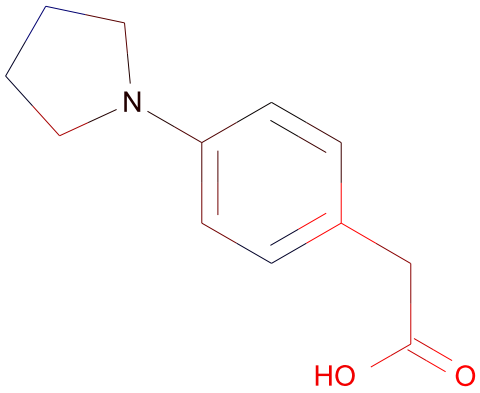
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-560785749	 [*]C(=[*])OC(=[*])[*]	0.259
FCFP_6	436915834	 [*]C\C(=C\[*])\O[*]	0.184

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	683445015	 [*]O[*]	-0.266

ECFP_6	2077607946	 [*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-0.252



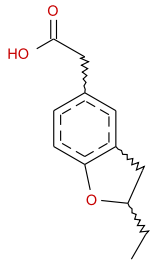
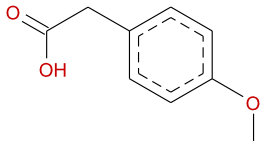
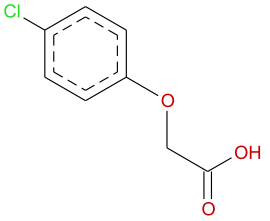
$C_{12}H_{15}NO_2$
Molecular Weight: 205.253
ALogP: 2.117
Rotatable Bonds: 3
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.652
Unit: g/kg_body_weight
Mahalanobis Distance: 16.3
Mahalanobis Distance p-value: 0.187

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	FUROFENAC	p-METHOXYPHENYLACETIC ACID	PHENOXYACETIC ACID; p-CHLORO
Structure			
Actual Endpoint (-log C)	2.202	2.03	2.341
Predicted Endpoint (-log C)	2.50202	2.31007	2.38776
Distance	0.388	0.403	0.438
Reference	DRFUD4 3;586;78	FCTXAV 14;685;76	RREVAH 10;97;65

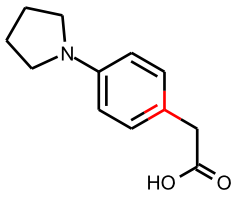
Model Applicability

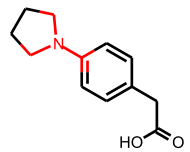
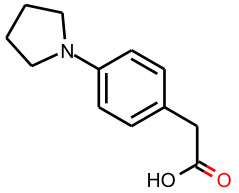
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

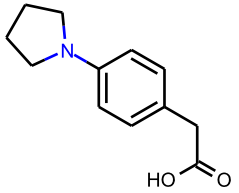
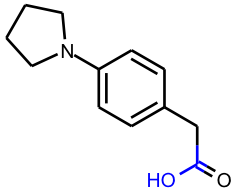
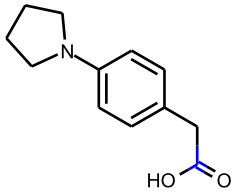
Feature Contribution

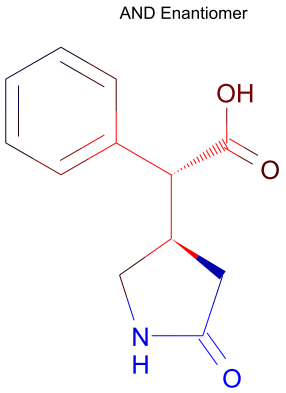
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

FCFP_6	675769755	 [*]:[c](:[*])N1C[*][*]C1	0.155
ECFP_6	-1074141656	 [*]=O	0.142

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	670515721	 [*]N([*])[*]	-0.108
ECFP_6	2025485523	 [*]C(=[*])O	-0.0801
FCFP_6	0	 [*]C([*])[*]	-0.0791



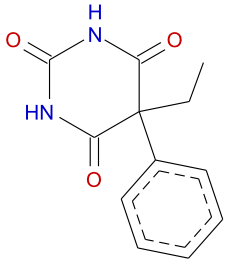
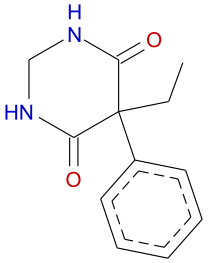
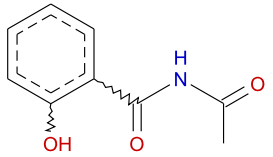
$C_{12}H_{13}NO_3$
Molecular Weight: 219.237
ALogP: 0.621
Rotatable Bonds: 3
Acceptors: 3
Donors: 2

Model Prediction

Prediction: 0.566
Unit: g/kg_body_weight
Mahalanobis Distance: 18.5
Mahalanobis Distance p-value: 5.43e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	PHENOBARBITAL	PRIMIDONE	SALACETAMIDE
Structure			
Actual Endpoint (-log C)	3.156	2.163	1.952
Predicted Endpoint (-log C)	2.85686	2.5813	1.75044
Distance	0.464	0.483	0.485
Reference	TXAPA9 18;185;71	NIIRDN 6;691;82	JPETAB 89;205;47

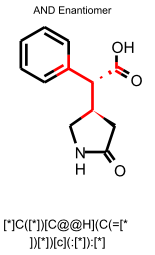
Model Applicability

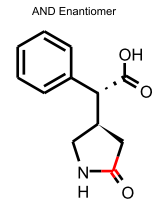
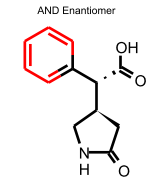
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1603312431: [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]
3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
4. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

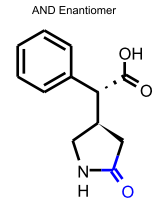
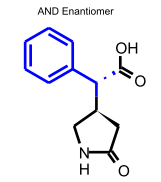
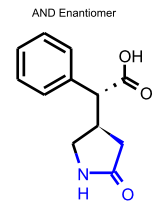
Feature Contribution

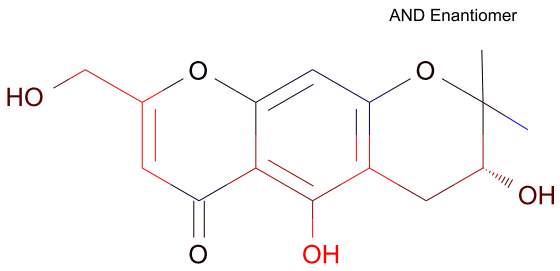
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1186303932	 [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]	0.375

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	2014710090	 [*]C([*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.225
FCFP_6	566058135	 O=C1C[*][*]N1	-0.216



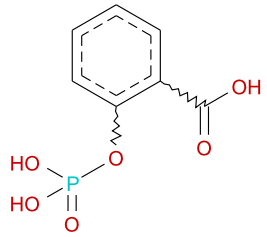
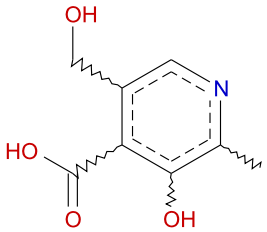
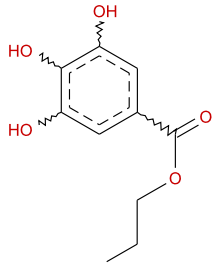
$C_{15}H_{16}O_6$
Molecular Weight: 292.284
ALogP: 0.901
Rotatable Bonds: 1
Acceptors: 6
Donors: 3

Model Prediction

Prediction: 0.113
Unit: g/kg_body_weight
Mahalanobis Distance: 21.8
Mahalanobis Distance p-value: 1.43e-015

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	SALICYLIC ACID; DIHYDROGEN PHOSPHATE	PYRIDOXIC ACID	PROPYL GALLATE
Structure			
Actual Endpoint (-log C)	2.296	1.388	2.005
Predicted Endpoint (-log C)	2.14696	2.13432	2.00438
Distance	0.600	0.624	0.625
Reference	ARZNAD 30;1098;80	ARZNAD 11;922;61	NTIS** PB245-441

Model Applicability

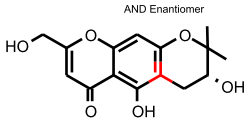
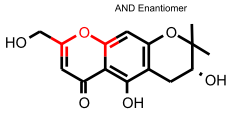
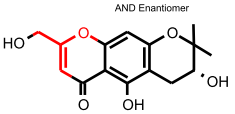
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
4. Unknown FCFP_6 feature: 74595001: [*]:[cH]:[c](O):[cH]:[*]
5. Unknown FCFP_6 feature: -549108873: [*]:[c](:[*])O
6. Unknown FCFP_6 feature: -415216134: [*]OC(C)(C)C([*])[*]

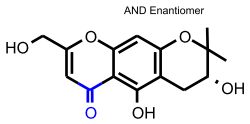
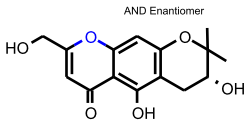
Feature Contribution

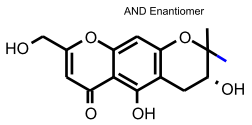
Top features for positive contribution

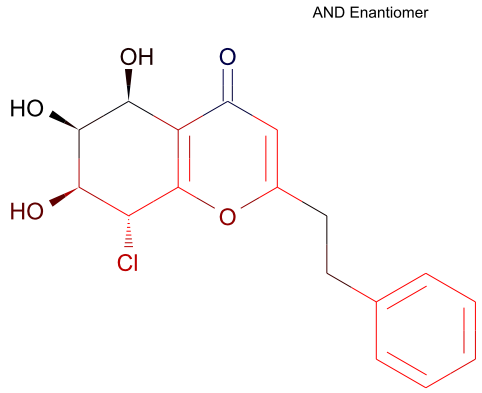
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-560785749	 [*]C(=[*])OC(=[*])[*]	0.259
FCFP_6	436915834	 [*]C\C(=C\[*])\O[*]	0.184

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	683445015	 [*]O[*]	-0.266

ECFP_6	734603939	 [*]C	-0.201



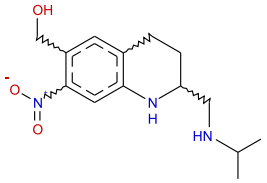
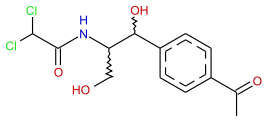
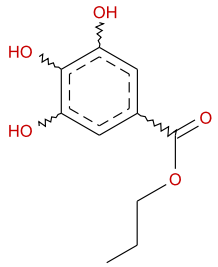
$C_{17}H_{17}ClO_5$
Molecular Weight: 336.767
ALogP: 1.368
Rotatable Bonds: 3
Acceptors: 5
Donors: 3

Model Prediction

Prediction: 0.428
Unit: g/kg_body_weight
Mahalanobis Distance: 22
Mahalanobis Distance p-value: 4.33e-016

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	OXAMNIQUINE	CETOPHENICOL	PROPYL GALLATE
Structure			
Actual Endpoint (-log C)	3.969	2.465	2.005
Predicted Endpoint (-log C)	2.33525	2.80207	2.00438
Distance	0.576	0.585	0.622
Reference	DDREDK 4;229;84	TXAPA9 18;185;71	NTIS** PB245-441

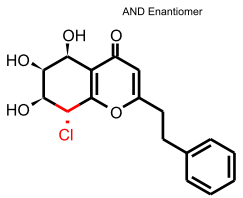
Model Applicability

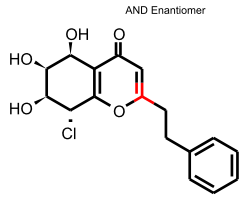
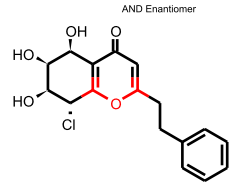
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1652274794: [*]OC(=C([*])[*])C([*])[*]
3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
4. Unknown FCFP_6 feature: -879019572: [*]C([*])[C@@H](Cl)C(=[*])[*]
5. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

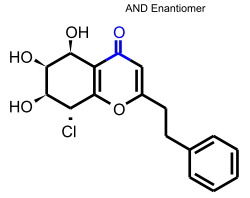
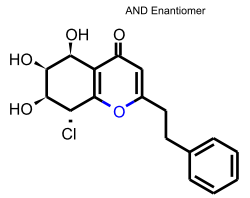
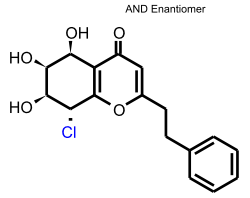
Feature Contribution

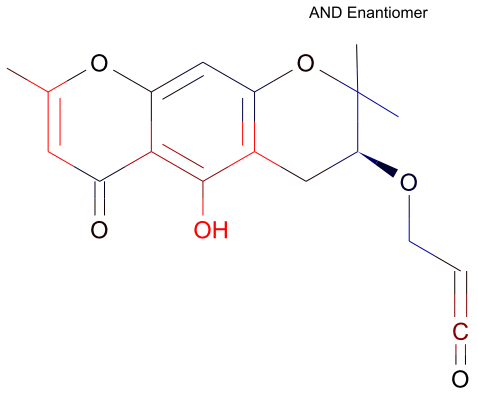
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	71953198	 [*]C([*])Cl	0.392

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-560785749	 [*]C(=[*])OC(=[*])[*]	0.259

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	683445015	 [*]O[*]	-0.266
ECFP_6	-817402818	 [*]Cl	-0.263



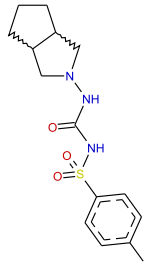
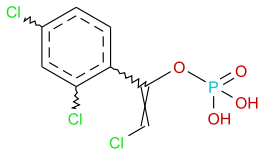
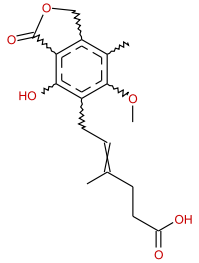
$C_{18}H_{18}O_6$
Molecular Weight: 333.356
ALogP: 2.007
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.129
Unit: g/kg_body_weight
Mahalanobis Distance: 25.1
Mahalanobis Distance p-value: 1.24e-031

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	N-[[(HEXAHYDROCYCLOPENTA[c]PYRROL-2(1H)-YL)AMINO]CARBONYL]-4-METHYL-BENZENESULFONAMIDE	PHOSPHORIC ACID; 2-CHLORO-1-(2;4-DICHLOROPHENYL)VINYL ESTER	MYCOPHENOLIC ACID
Structure			
Actual Endpoint (-log C)	1.811	4.368	2.959
Predicted Endpoint (-log C)	1.89924	2.70639	2.66045
Distance	0.546	0.553	0.582
Reference	YACHDS 8;2661;80	TXAPA9 21;315;72	TOIZAG 29;400;82

Model Applicability

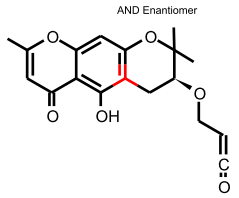
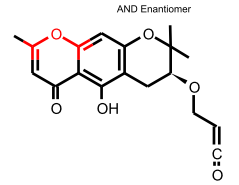
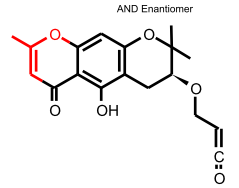
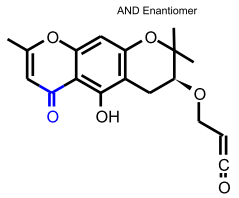
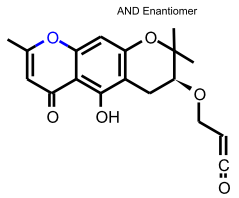
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

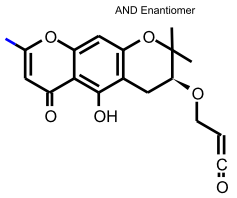
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1591590376: [*]C=C=O
3. Unknown ECFP_2 feature: -91536905: [*]CC=C=[*]
4. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
5. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
6. Unknown FCFP_6 feature: 74595001: [*]:[cH]:[c](O):[cH]:[*]
7. Unknown FCFP_6 feature: -415216134: [*]OC(C)(C)C([*])[*]
8. Unknown FCFP_6 feature: -549108873: [*]:[c](:[*])O
9. Unknown FCFP_6 feature: 444624378: [*]C=C=O

Feature Contribution

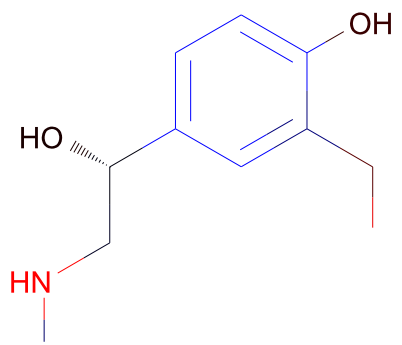
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-560785749	 [*]C(=[*])OC(=[*])[*]	0.259
FCFP_6	436915834	 [*]C\C(=C\[*])\O[*]	0.184
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	683445015	 [*]O[*]	-0.266

ECFP_6	734603939	AND Enantiomer  [*]C	-0.201

AND Enantiomer



$C_{11}H_{17}NO_2$

Molecular Weight: 195.258

ALogP: 1.557

Rotatable Bonds: 4

Acceptors: 3

Donors: 3

Model Prediction

Prediction: 0.527

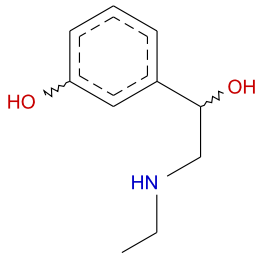
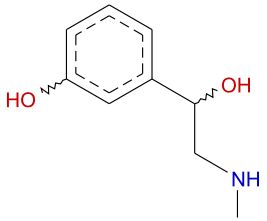
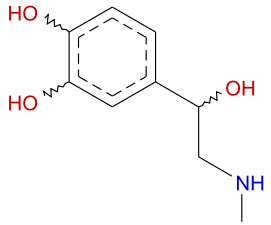
Unit: g/kg_body_weight

Mahalanobis Distance: 17.8

Mahalanobis Distance p-value: 0.00173

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ETILEFRINE	PHENYLEPHRINE	BENZYL ALCOHOL; 3;4-DIHYDROXY-.alpha.-[(METHYLAMINO)METHYL]-; .HCl (HCl STRIPPED)
Structure			
Actual Endpoint (-log C)	3.201	2.679	3.883
Predicted Endpoint (-log C)	2.7374	2.59876	2.22971
Distance	0.316	0.318	0.445
Reference	AIPTAK 180;155;69	AIPTAK 180;155;69	AIPTAK 180;155;69

Model Applicability

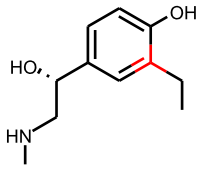
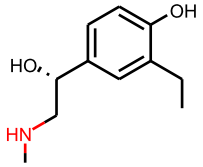
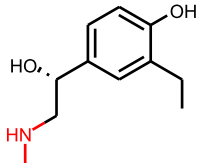
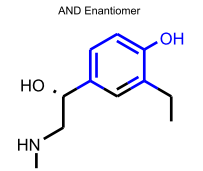
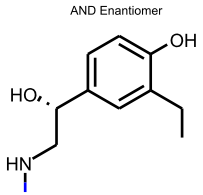
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

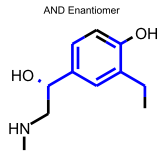
1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 74595001: [*]:[cH]:[c](O):[cH]:[*]
4. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
5. Unknown FCFP_6 feature: -549108873: [*]:[c](:[*])O
6. Unknown FCFP_6 feature: 1186393305: [*]C([*])[C@H](O)[c](:[*]):[*]
7. Unknown FCFP_6 feature: 136686699: [*]NC

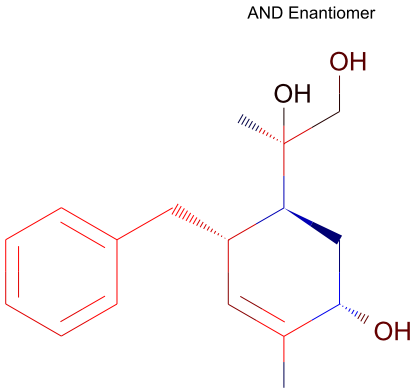
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	AND Enantiomer  [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	AND Enantiomer  [*]N[*]	0.216
ECFP_6	864287155	AND Enantiomer  [*]NC	0.188
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	946589555	AND Enantiomer  O[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.204
ECFP_6	734603939	AND Enantiomer  [*]C	-0.201

FCFP_6	-451251206	AND Enantiomer  [*]C[c]1:[cH]:[c](:[cH]:[*]:[c]:1[*])C([*])[*]	-0.143



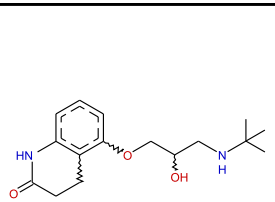
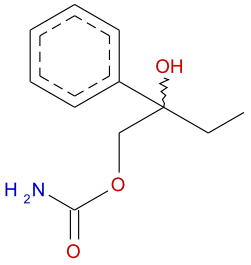
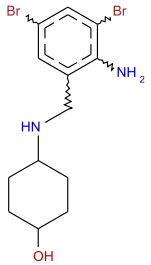
$C_{17}H_{24}O_3$
Molecular Weight: 276.371
ALogP: 2.165
Rotatable Bonds: 4
Acceptors: 3
Donors: 3

Model Prediction

Prediction: 2.1
Unit: g/kg_body_weight
Mahalanobis Distance: 20.3
Mahalanobis Distance p-value: 4.92e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	CARTEOLOL .HCl (HCl STRIPPED)	2-PHENYL-1;2-BUTANEDIOL-1-CARBAMATE	AMBROXOL
Structure			
Actual Endpoint (-log C)	2.342	2.538	1.451
Predicted Endpoint (-log C)	1.91383	2.27956	2.08503
Distance	0.551	0.604	0.609
Reference	OYYAA2 11;159;76	27ZQAG -;394;72	MDACAP 15;523;79

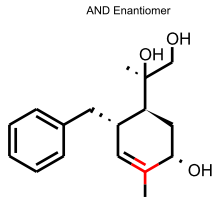
Model Applicability

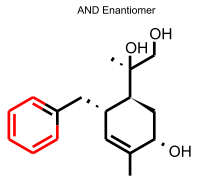
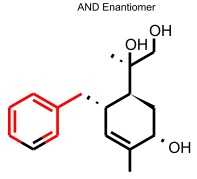
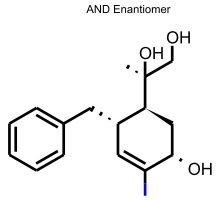
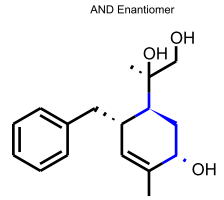
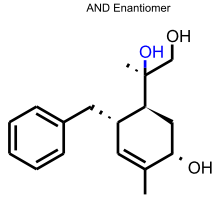
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1263967621: [*]C[C@@H](C=[*])C([*])[*]
3. Unknown ECFP_2 feature: 1280892564: [*]C[C@](C)(O)C([*])[*]
4. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
5. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

Feature Contribution

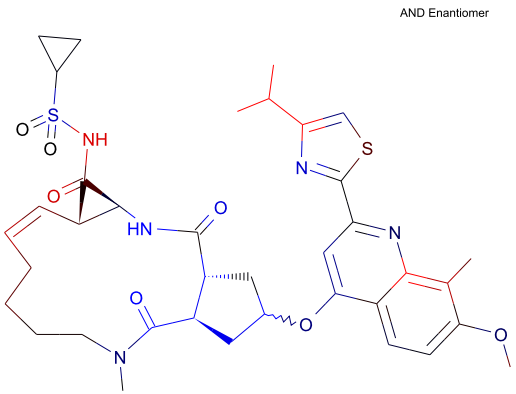
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19
ECFP_6	1095683433	 [*]C[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0.123
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
ECFP_6	734603939	 [*]C	-0.201
ECFP_6	-801490360	 [*]C([*])CC([*])[*]	-0.189
FCFP_6	3	 [*]O	-0.107

Simeprevir

TOPKAT_Rat_Oral_LD50



$C_{38}H_{47}N_5O_7S_2$

Molecular Weight: 749.93908

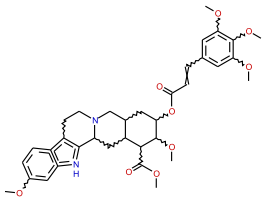
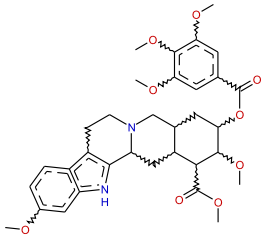
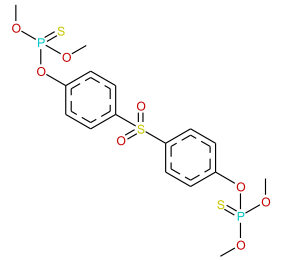
ALogP: 4.796

Rotatable Bonds: 8

Acceptors: 9

Donors: 2

Structural Similar Compounds

Name	ANAPREL	RESERPINE	PHOSPHOROTHIOIC ACID; O;O'-(SULFONYLDI-p-PHENYLENE) O;O;O';O'-TETRAMETHYL ESTER
Structure			
Actual Endpoint (-log C)	2.803	3.161	2.397
Predicted Endpoint (-log C)	2.99154	2.72801	3.8515
Distance	1.032	1.042	1.057
Reference	NIIRDN 6;898;82	PSSCBG 11;555;80	TXAPA9 21;315;72

Model Prediction

Prediction: 0.209

Unit: g/kg_body_weight

Mahalanobis Distance: 37.6

Mahalanobis Distance p-value: 1.1e-129

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Model Applicability

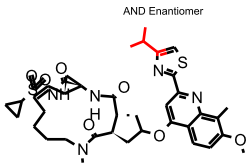
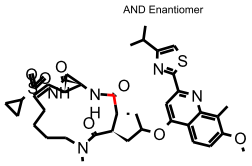
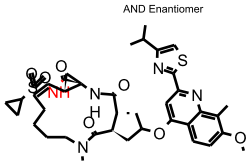
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -309676563: [*]N[C@]1(C[C@H]1[*])C(=[*])[*]
3. Unknown ECFP_2 feature: -1049290660: [*]C1([*])C[C@H]1C=[*]
4. Unknown ECFP_2 feature: 2121941848: [*]NS(=O)(=O)C1[*][*]1
5. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
6. Unknown FCFP_6 feature: 1747237384: [*][c](:[*]):n:[c](:[*]):[*]
7. Unknown FCFP_6 feature: 690511177: [*]:[cH]:[c](:n:[*])[c](:[*]):[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 690481386: [*]:[c](:[*])[c]1:n:[*]:[*]:s:1
10. Unknown FCFP_6 feature: 1018942292: [*]NS(=O)(=O)C1[*][*]1

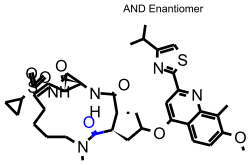
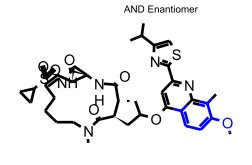
Feature Contribution

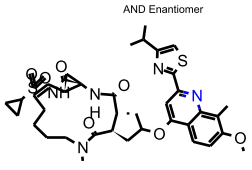
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	1186303932	 [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]	0.375
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	2077607946	 [*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-0.252

ECFP_6	655739385	AND Enantiomer  [*]:n:[*]	-0.239