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: lbrahimeissa@azhar.edu.eg

Content

Method	Fingerprint studies
	Molecular Similarity
	Docking studies
	ADMET studies
	Toxicity studies
	Molecular dynamics and MMPBSA
Toxicity re	eport

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-GUI¹⁻³ interface with the CHARMM36 force field⁴. All the simulations were done using the NAMD 2.13⁵ package. The TIP3P explicit solvation model was used⁶, 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 field⁷ 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 thermostat¹⁰. 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,} 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 algorithm¹³. 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 code¹⁶ 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 Å² were used for MM/GBSA calculations.

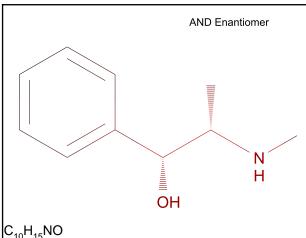
MMPBSA

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



101115110

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

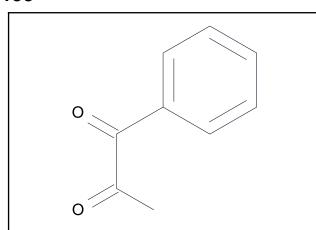
Structural Similar Compounds					
Name	Phenethyl Alcohol	Mexiletine .HCl (Free base form)	Clenbuterol		
Structure	OH	NH 2	H ₂ N ₁ , OH		
Actual Endpoint	Toxic	Non-Toxic	Toxic		
Predicted Endpoint	Toxic	Non-Toxic	Toxic		
Distance	0.514	0.540	0.621		
Reference	J Toxicol Environ Health 12(2-3):235-44; 1983	lyakuhin Kenkyu 14(4):550-570; 1983	lyakuhin 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.

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	1702724181	AND Enantiomer TOH H TOH H	0.558	9 out of 9	

SCFP_6	-561151481	AND Enantiomer OH H [*]C([*])[C@H](O)[c]1 :[cH]:[cH]: [t]t]1	0.523	6 out of 6
SCFP_6	1318513260	AND Enantiomer OH H [']N[C@@H](C) C@H](O) [c]([cH]['])[cH]; [']	0.478	4 out of 4
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1379591900	AND Enantiomer	-0.282	33 out of 84
SCFP_6	1653911926	AND Enantiomer	-0.188	12 out of 28
SCFP_6	-1631132401	AND Enantiomer OH [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-0.0973	8 out of 17



C₀H₈O₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Ethyl Acrylate	Methylisobutylketone	2;2'-Dipyridyl		
Structure		0	N N		
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		

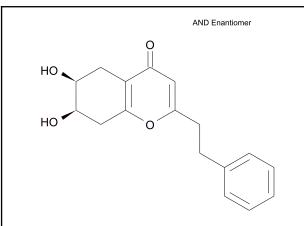
Model Applicability

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

Feature Contribution							
	Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set			
SCFP_6	-1980302127	[*]C(=[*])C(=O)[o](:[*]):[*]	0.25	5 out of 7			

SCFP_6	-2056718782		0.0786	33 out of 58
		(*]C(=[*])[c](:[cH]:[
]):[cH]:[]		
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1379591900		-0.282	33 out of 84
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1		
SCFP_6	1649310982		-0.212	10 out of 24
		[*]C(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1		
SCFP_6	1653911926		-0.188	12 out of 28
		[*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1		

V



C₁₇H₁₈O₄

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

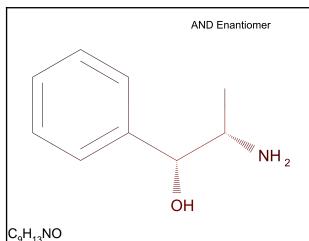
Name	Sulfonylurea Gliclazide	Cyclopiazonic Acid	N-(3-Piperidyl)3;4;5- trimethoxybenzamide	
Structure	NH ONH	N HO	NH HN	
Actual Endpoint	Toxic	Non-Toxic	Toxic	
Predicted Endpoint	Toxic	Non-Toxic	Toxic	
Distance	0.536	0.591	0.603	
Reference	Yakuri to Chiryo 9:3551- 3571; 1981	J Toxicol Environ Health 14:585-594; 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.

Feature Contribution Top features for positive contribution					
SCFP_6	-1971196727	AND Enantlomer HO HO (*)C(=CC(=[*])[*])[*]	0.293	13 out of 18	

			1	
SCFP_6	-424515134	AND Enantiomer	0.275	39 out of 56
		HO		
		[*]C([*])O		
SCFP_6	-1211866396	AND Enantiomer HO HO	0.21	8 out of 12
		[*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1		
	Top Feat	ures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	616547045	AND Enantiomer HO HO HO HO HO HO HO HO HO H	-0.438	1 out of 4
		[*]C\C(=C\[*])\O[*]		
SCFP_6	-1379591900	AND Enantiomer HO HO	-0.282	33 out of 84
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1		
SCFP_6	55464376	AND Enandiomer HO HO O O O O O O O O O O	-0.252	1 out of 3
		[*]CC(=C((*))(*))C(=[*])(*]		



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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

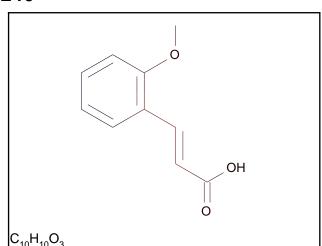
Structural Similar Compounds					
Name	Phenethyl Alcohol	2-Methylresorcinol	Mexiletine .HCI (Free base form)		
Structure	OH	OH OH	NH 2		
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic		
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic		
Distance	0.557	0.560	0.592		
Reference	J Toxicol Environ Health 12(2-3):235-44; 1983	Fundam Appl Toxicol 7:293-298; 1986	lyakuhin 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.

Feature Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	1702724181	AND Enantiomer NH 2 OH [7]C([7])(C@H](O)(c)(:(1))(7)	0.558	9 out of 9		

SCFP_6	-561151481	AND Enantiomer OH [*]C([*])[C@H](O)[c]1 :[cH][cH][*][cH]: [cH]1	0.523	6 out of 6
SCFP_6	1318513260	AND Enantiomer OH NH ₂ OH (*INIC@@HI(C) C@HI(O)	0.478	4 out of 4
	Top Foot	[c](:[cH]:[1]):[cH]: [*]		
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1379591900	AND Enantiomer , NH 2 OH [*][c]1:[*]:[cH]:[cH] :[cH]:1	-0.282	33 out of 84
SCFP_6	1653911926	AND Enantiomer OH [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.188	12 out of 28
SCFP_6	-1631132401	AND Enantiomer . , NH 2 OH [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-0.0973	8 out of 17



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

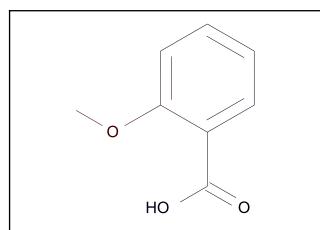
Structural Similar Compounds				
Name	Carbaryl	Mexiletine .HCI (Free base form)	Trichloroacetic Acid	
Structure	N H	NH 2	OH CI CI CI	
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic	
Predicted Endpoint	Non-Toxic	Non-Toxic	Non-Toxic	
Distance	0.540	0.562	0.570	
Reference	Toxicol Appl Pharmacol 26(4):621-38; 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.

Feature Contribution					
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	-1977229858	[*]C(=[*]))C=C\[o](:[cH]:[*]):[o]((*]):[*]	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 Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1652428703	[*]O[c]1:[cH]:[cH]:[c H]:[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



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

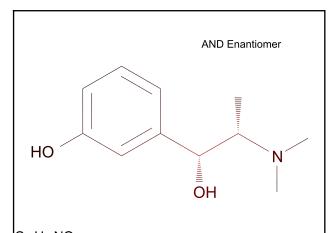
Structural Similar Compounds				
Name	Trichloroacetic Acid	3;6-Dichloropicolinic Acid	11-oxo-11H-Pyrido(2;1- b)quinazoline-2-carboxylic Acid	
Structure	OH CI CI CI	CI OH	N OH	
Actual Endpoint	Toxic	Non-Toxic	Toxic	
Predicted Endpoint	Non-Toxic	Non-Toxic	Toxic	
Distance	0.490	0.519	0.565	
Reference	Teratology 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.

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	HO (1) (cH):[*]:[cH	0.271	1 out of 1
]:[cH]:[c]:1C(=O)O		
SCFP_6	136239834	HOOO	0.242	23 out of 34
		[*]OC		
		es for negative c		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1652428703	HO O	-0.438	1 out of 4
		H]:[cH]:[c]:1[*]		
SCFP_6	-424485343	HOO	-0.328	19 out of 51
		[*]C(=[*])O		
SCFP_6	-1379591900	(*)[c]1:[*]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84



C₁₁H₁₇NO₂

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

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

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 nonnormal 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	NH 2	НО	H ₂ N ₁ , OH		
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic		
Predicted Endpoint	Non-Toxic	Non-Toxic	Toxic		
Distance	0.546	0.586	0.599		
Reference	lyakuhin Kenkyu 14(4):550-570; 1983	Fundam Appl Toxicol 8:571-582; 1987	lyakuhin 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.

Feature Contribution					
	Top fea	atures for positive c	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	1702724181	AND Enantiomer HO OH [*]C([*])[C@H](O)[c](::(*);[*]	0.558	9 out of 9	

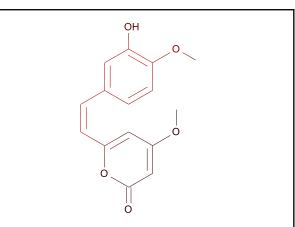
SCFP_6	-561151481	AND Enantiomer OH [']C(['))(C@H)(O)[e]1 :[cH:[cH:]'[:[cH:	0.523	6 out of 6
SCFP_6	1318513260	AND Enantiomer HO OH [']N[C@@H](C)[C@H](O) [c](:[cH];[']):[cH]: [']	0.478	4 out of 4
	Top Feat	ures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1434483901	AND Enantiomer : OH [*][c]1:[cH]:[cH]:[cH]:[c](O):[cH]:1	-0.422	0 out of 1
SCFP_6	-1379591900	AND Enantiomer HO OH [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.282	33 out of 84
SCFP_6	3	AND Enantiomer HO OH [*][c](:[*]):[*]	0	92 out of 181

Toxic

0.587

Arzneimittelforschung

23(4):504-8; 1973



C₁₅H₁₄O₅

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 nonnormal 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 .HCI (Free base form)	
Structure	NH HN ONTO	CI NO OH	CI N N N N N N N N N N N N N N N N N N N	
Actual Endpoint	Toxic	Non-Toxic	Toxic	

Model Applicability

Predicted Endpoint

Distance

Reference

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

Non-Toxic

Fundam Appl Toxicol

4:872-882: 1984

0.580

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

Kiso to Rinsho 18:91-101;

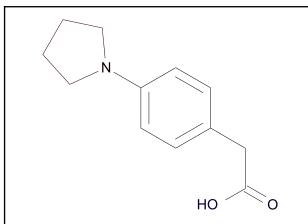
Toxic

0.573

1984

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	2116304939	[*]O[c]1:[cH]:[*]:[c] ([*]):[cH]:[c]:10	0.504	5 out of 5	

SCFP_6	-538866216	OH O	0.478	4 out of 4
		[*]\C=C/[c]1:[cH]:[cH]:[*]:[c]((*]):[cH]: 1		
SCFP_6	-1977229858	OH O	0.478	4 out of 4
		[*]C(=[*])\C=C\[o](:[cH]:[*]):[o]([*]):[*		
	Ton Feat	ures for negative of	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	OH O\ (*)O	0	97 out of 178
SCFP_6	1	[*]C(=[*])[*]	0	90 out of 173



C₁₂H₁₅NO₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Pirprofen	Miroprofen	Mexiletine .HCI (Free base form)	
Structure	OH OH	OH OH	NH 2	
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	lyakuhin Kenkyu 12:808- 826; 1981	lyakuhin 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 fea	ntures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1100167450	ho o ho o ho o ho o ho o ho o ho o ho o ho o o ho o o o	0.441	3 out of 3

SCFP_6	470101049	HOOO [*]C(=[*])C[c](:[*]): [*]	0.322	4 out of 5
SCFP_6	1155241219	HO O	0.302	11 out of 15
	Top Feat	ures for negative of	contribution	1
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	[*]:[eH]:[e](CC(=0)0) :[eH]:[*]	-0.422	0 out of 1
SCFP_6	1334669481	[*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.355	10 out of 28

AND Enantiomer OH N H O

C₁₂H₁₃NO₃ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Bropirimine	Sulfonylurea Gliclazide	Trichloroacetic Acid	
Structure	Br N NH 2	NH ONH	OH CI CI CI	
Actual Endpoint	Non-Toxic	Toxic	Toxic	
Predicted Endpoint	Non-Toxic	Toxic	Non-Toxic	
Distance	0.581	0.624	0.627	

Model Applicability

Reference

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

Yakuri to Chiryo 9:3551-

3571; 1981

Teratology 40:445-451;

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

Teratology 38(1):7-14;

1988

Feature Co	Feature Contribution				
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	1983581611	AND Enantiomer OH N (*)C((*))(C@@H)(C(=(*))(*)()(-(-(*)(-(-(*)()(-(-(-(*)()(-(-(-(-	0.271	1 out of 1	

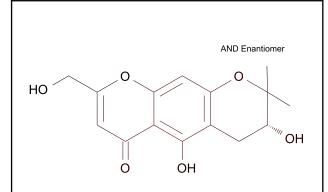
SCFP_6	9	AND Enantiomer OH N OH N Frinkfil	0.0928	45 out of 78
SCFP_6	-1272768868	[*]N[*] AND Enantiomer OH	0.0607	14 out of 25
		[*]C([*])CC(=[*])[*]		
		res for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1214451192	AND Enantiomer OH N O (*)C([*))(C@@H)(C(=[*))(*))(c] 1: (cH); (cH) :[*]:[cH]:[cH]:1	-0.849	1 out of 7
SCFP_6	413428563	AND Enantiomer OH NH O I*IC(I*))(C@@H)(C(=O) O)(c)(1*]1:1*	-0.729	1 out of 6
SCFP_6	-1047009138	AND Enantiomer OH NH O [*][C@@H]1CNC(=0)C1	-0.718	0 out of 2

Toxic

0.672

Toxicol Appl Pharmacol

36(2):227-37; 1976



C₁₅H₁₆O₆

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sillinar Compounds				
Name	Piroxicam	Dexamethasone	Caffeic Acid	
Structure	OH HNWW	HO start to the start of the st	HO W OH	
Actual Endpoint	Toxic	Non-Toxic	Toxic	

Model Applicability

Predicted Endpoint

Distance

Reference

Structural Similar Compounds

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

Toxic

0.667

Preclin Rep Cent Inst Exp

Anim 8:95-115; 1982

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

Yakuri to Chiryo 8:4655-

Toxic

0.628

4671; 1980

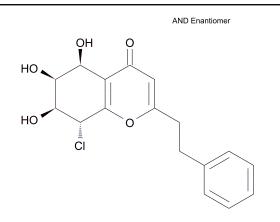
Feature Co	Feature Contribution				
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	-1849894309	AND Enantiomer HO OH [*][c](:[*]):[c]1C[C@ @H](O)C([*])([*])[*] [c]:1:[*]	0.381	2 out of 2	

SCFP_6	-617610981	AND Enantiomer HO OH OH [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*]	0.381	2 out of 2
SCFP_6	130348166	AND Enantiomer HO OH [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]: 1[*]	0.369	5 out of 6
	Top Fea	tures for negative of	ontribution	•
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	616547045	AND Enandomer HO OH OH [*]C\C(=C\[*])\O[*]	-0.438	1 out of 4
SCFP_6	-711656408	AND Enantiomer HO OH OH [*]C(=[*])CO	0	7 out of 13
SCFP_6	1311071855	AND Enantiomer HO OH OH [*]C(=O)[*]	0	72 out of 141

0.610

3571; 1981

Yakuri to Chiryo 9:3551-



C₁₇H₁₇CIO₅

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Dexamethasone	Diflorasone Diacetate	Sulfonylurea Gliclazide	
Structure	HO TO	HO state of the st	NH ONH	
Actual Endpoint	Non-Toxic	Toxic	Toxic	
Predicted Endpoint	Toxic	Toxic	Toxic	

Model Applicability

Distance

Reference

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

0.586

1984

Oyo Yakuri 28(2):207-224;

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

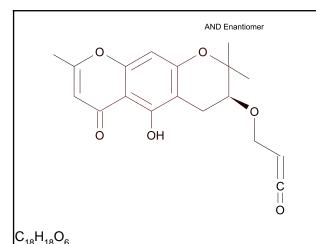
Preclin Rep Cent Inst Exp

Anim 8:95-115; 1982

0.564

Feature Co	Feature Contribution					
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	1702664599	AND Enantiomer HO CI (')C((')))C@@H)(O)C(= ('))(')	0.322	4 out of 5		

SCFP_6	-1971196727 -424515134	[']C(=CC(=['])['])[']		13 out of 18 39 out of 56
		HO CI ([*])O		
P*		es for negative c		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-52163885	AND Enantiomer HO OH OO CI [*]C([*])CI	-0.496	6 out of 20
SCFP_6	616547045	AND Enantiomer HO CI [*]C\C(=C\[*])\O[*]	-0.438	1 out of 4
SCFP_6	-1379591900	AND Enantiomer HO CI [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.282	33 out of 84



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Sulfonylurea Gliclazide	D&C Yellow 8	N-(3-Piperidyl)3;4;5- trimethoxybenzamide	
Structure	NH ONH	OH OH	NH HN ON THE STATE OF THE STATE	
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.

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	-617610981	AND Enantiomer O OH O OH C O [*][o](:[*]):[o]1C(=O)C=C([*])[*][o]:1:[*]	0.381	2 out of 2	

SCFP_6	-1849894309	AND Enantiomer OH [*][c](:[*]):[c]1C[C@ @H](O)C([*])([*])**] [c]:1:[*]	0.381	2 out of 2
SCFP_6	130348166	AND Enantiomer OH [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]: 1[*]	0.369	5 out of 6
		tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	616547045	AND Enantiomer OH C C C (*]C\C(=C\[*])\O[*]	-0.438	1 out of 4
SCFP_6	136627117	AND Enantiomer OH OH C C O O [*]C(=[*])C	-0.0885	10 out of 21
SCFP_6	136597326	AND Enantiomer OH C C C O [*]C([*])C	0	50 out of 97

AND Enantiomer	
ОН	
HO IIIIIII	
HN	
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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Clenbuterol	Procarbazine .HCI (Free base form)	Ritodrine .HCI (Free base form)
Structure	H ₂ N ₁ , OH	N H N NH	HO NH OH
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.491	0.539	0.567
Reference	lyakuhin 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.

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	1702724181	AND Enantiomer HO	0.558	9 out of 9	

SCFP_6	-561151481	AND Enantomer OH HO I**C([*])[C@H](O)[c]1 I**Cell*[J**CHI**] TO CELL*[J**CHI**] AND Enantomer OH	0.523	6 out of 6
SCFP_6	-1059390504	[OH]:[OH]:[]:[OH]: [OH]:1 AND Enantiomer OH HN ['INC(C@H](O)(c)(c)H I('I):[OH]:[']	0.441	3 out of 3
	Top Feat	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1931277081	AND Enantiomer HO . HN [*]C[c]1:[cH]:[*]:[cH]:[cH]:[c]:10	0	1 out of 2
SCFP_6	136597326	AND Enantiomer OH HO,	0	50 out of 97
SCFP_6	-496409612	AND Enantiomer OH HO HN [*][c](:[*]):[cH]:[cH]:[*]	0	82 out of 163

AND Enantiomer OH ОН

 $C_{17}H_{24}O_{3}$

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

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

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

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 prediction. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Clenbuterol	Ambroxol	Procarbazine .HCl (Free base form)
Structure	H ₂ N ₁ , OH	Br to Br NH 2	N H N NH
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.468	0.608	0.610
Reference	lyakuhin Kenkyu 15(4):597-613; 1984	lyakuhin 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.

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	AND Enantiomer OH OH (*]C([*])[C@@H]((O)C(= [*])[*]	0.322	4 out of 5		

SCFP_6	-424515134	AND Enantiomer OH OH OH (*]C([*])O	0.275	39 out of 56
SCFP_6	1416196903	AND Enantiomer OH OH (*)OC(C)(C)(C)([*))(*)	0.255	3 out of 4
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1379591900	AND Enantiomer OH OH (*)[c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.282	33 out of 84
SCFP_6	1653911926	*][c]1:[cH]:[cH]:[cH]	-0.188	12 out of 28
SCFP_6	-1043339860	AND Enantiomer OH OH OH (*]C([*])C1C[*][*]C1	-0.114	13 out of 28

Simeprevir

TOPKAT_Developmental_Toxicity_Potential

	AND Enantiomer
O NH O HN O	N S N O
HN	mo

 $|C_{38}H_{47}N_5O_7S_2|$

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

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 nonnormal 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	No hard to have the second of	Br Nth OH has	CI C		
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic		
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic		
Distance	0.675	0.778	0.819		

Model Applicability

Reference

Structural Similar Compounds

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

1990

Toxicol Lett 50:189-194;

Oyo Yakuri 20(6):1219-

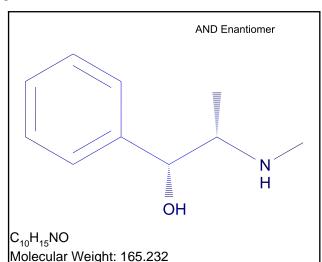
1236; 1980

Oyo Yakuri 18:105-124;

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

Fingerprint Bit/Smiles Feature Structure Score Toxic in training set SCFP_6 591469355 0.411 10 out of 12

SCFP_6	-395254381	AND Enantiomer	0.271	1 out of 1
		[*]:[cH]:[c](OC1C[*][*]C1):[c](:[*]):[*]		
SCFP_6	-109092631	AND Enantiomer AND Enantiomer NS N S N (1) CCN(C)C(=0)C([1])[0.271	1 out of 1
		*1	4 11 41	
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-98332825	AND Enantiomer NS	-0.718	0 out of 2
		[*]CCCCC=[*]		
SCFP_6	-1476112164	AND Enantiomer AND Enantiomer NS N S N (*]CC\C=C/C1[*][*]1	-0.718	0 out of 2
SCFP_6	1260369147	AND Enantiomer AND Enantiomer NS N S S	-0.718	0 out of 2



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

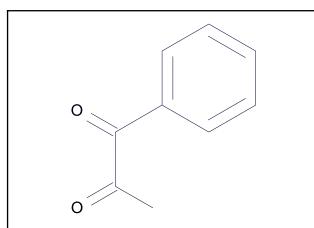
Name	Ephedrine	Phenelzine	Tocainide
Structure	OH N H	NH NH NH 2	NH NH NH 2
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.000	0.523	0.550
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (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.

ntribution						
Top features for positive contribution						
Bit/Smiles	Feature Structure	Score	Carcinogen in training set			
864287155	AND Enantiomer OH [*]NC	0.2	4 out of 10			
Top Features for negative contribution						
Bit/Smiles	Feature Structure	Score	Carcinogen in training set			
	Top fe Bit/Smiles 864287155 Top Fe	Bit/Smiles Feature Structure 864287155 AND Enantiomer OH [*]NC Top Features for negative of	Top features for positive contribution Bit/Smiles Feature Structure Score 864287155 AND Enantiomer 0.2 Top Features for negative contribution			

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



C₀H₈O₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Coumarin	Amphetamine	Phenacetin
Structure		NH ₂	O H
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 fea	atures for positive o	ontribution		
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		-0.459	1 out of 7
		[*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1		
ECFP_6	169261700		-0.27	0 out of 1
		[*]C(=[*])C(=O)[c]1:[cH]:[cH]:[*]:[cH]:[c H]:1		
ECFP_6	1432101658		-0.27	0 out of 1
		[*]C(=[*])C(=O)[c](:[*]):[*]		

AND Enantiomer HO HO O

C₁₇H₁₈O₄ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sim	ilar Compounds Tolazamide	Naltrexone	Phenobarbital
Structure		HO MAN TO THE PART OF THE PART	HN O
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

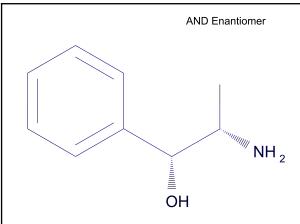
Model Applicability

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 fe	atures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	464808839	AND Enartiomer HO O (")C(=CC(=[")([")[")]")	0.524	8 out of 14	

ECFP_6	2106656448	AND Enantiomer HO HO [*]C(=O)[*]	0.254	31 out of 77
ECFP_6	-560785749	AND Enantiomer HO HO (*)C(=[*])OC(=[*])[*]	0.212	1 out of 2
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	571598839	AND Enantiomer HO HO ("]C1=["]C[C@@H](O)[C @@H](O)C1	-0.27	0 out of 1
ECFP_6	1299558496	AND Enantiomer HO HO [*]C(=[*])C(=0)C=[*]	-0.27	0 out of 1
ECFP_6	1875238785	AND Enantiomer HO HO [*]C\C(=C\[*])\O[*]	-0.27	0 out of 1



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	Ephedrine	Phenelzine	Tocainide
Structure	OH N H	NH NH NH 2	NH NH 2
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

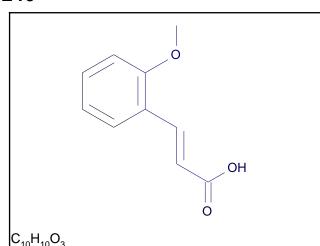
Model Applicability

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 Co	ntribution				
	Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	-933808133	AND Enantiomer NH ₂ OH [*]C([*])N	-1.33	0 out of 9	

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



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	O OH	HX	OH OH
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen

Model Applicability

Distance

Reference

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

0.555

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

0.560

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

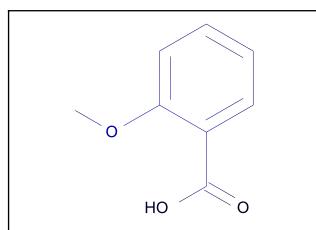
0.527

2. Unknown ECFP_2 feature: 1335702447: [*][c](:[*]):[c](C=[*]):c:[*]

Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set ECFP_6 -1925046727 0.391 11 out of 23

Top Features for negative contribution

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



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

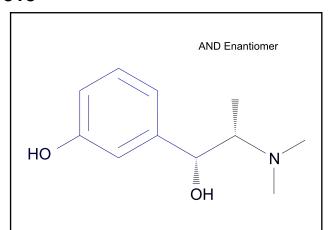
Name	Aspirin	Nicotinic acid	Phenacetin
Structure	O OH	N OH	O H
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.348	0.569	0.592
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for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

Feature Co	ntribution			
	Top fe	eatures for positive c	ontribution	
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 Fe	eatures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	1307307440	HO O	-0.558	4 out of 25
ECFP_6	2025485523	HO O	-0.506	8 out of 45
ECFP_6	1635415905	HO O [*]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 Sim	ilar Compounds		
Name	Ephedrine	Phenylephrine	Tocainide
Structure	OH N H	HO NH	NH NH 2
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

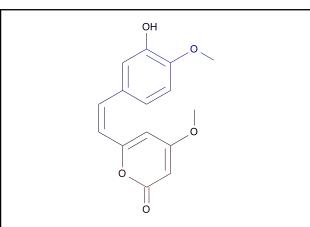
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution				
	Top fea	atures for positive c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-177786161	AND Enantiomer OH [*]:[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	AND Enantiomer HO OH	-0.857	1 out of 12
		[*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1		
ECFP_6	-783815036	AND Enantiomer : OH	-0.657	0 out of 3
		O[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1		
ECFP_6	2007300961	AND Enantiomer HO OH	-0.652	5 out of 34
		[*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1		



C₁₅H₁₄O₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Cytembena	Scopolamine	Fluconazole
Structure	O OH Br	O THE STATE OF THE	OH NINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

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

Non-Carcinogen

Non-Carcinogen

Model Applicability

Actual Endpoint

Structural Similar Compounds

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:[*]

Non-Carcinogen

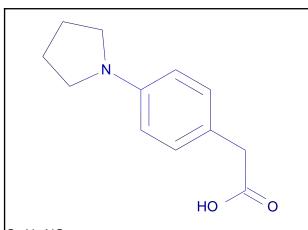
- 3. Unknown ECFP_2 feature: -1053980253: [*]O\C(=C/[*])\C=[*]
- 4. Unknown ECFP_2 feature: -444332269: [*]O\C(=C\[*])\C=[*]

Feature Contribution

Top features for positive contribution					
ECFP_6	464808839	[,]c(=cc(=[,])[,])[,] OH	0.524	8 out of 14	

ECFP_6	-1885846789	[*]OC(-O)C-[*]	0.424	1 out of 1
ECFP_6	-1925046727	[*]OC(=O)C=[*]	0.391	11 out of 23
	Top Feat	tures for negative of	contribution	1
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

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



 $C_{12}H_{15}NO_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Methylphenidate	Phenacetin	Tolmetin		
Structure	N H O	O H N	O OH		
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.

Feature Contribution						
Top fe	atures for positive o	ontribution				
Bit/Smiles	Feature Structure	Score	Carcinogen in training set			
341335091	HO O	0.212	1 out of 2			
	Top fe Bit/Smiles	Bit/Smiles Feature Structure 341335091	Top features for positive contribution Bit/Smiles Feature Structure Score 341335091 0.212			

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 Feat	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1731135544	N HO [*]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 OH N H O

 $C_{12}H_{13}NO_3$

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	HN O	O NH 2 N H	NH 2	
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

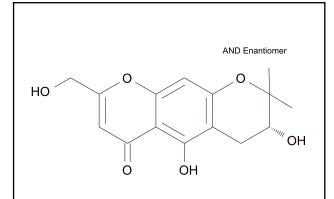
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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	AND Enantiomer OH N (*)C((*))(C@@H)(C(=[*))(*)()(-1:(-1)(-1)(-1)(-1)(-1)(-1)(-1)(-1)(-1)(-1)	0.424	1 out of 1	

ECFP_6	1603312431	AND Enantiomer OH NH ("]C((")) C@@H) C(=["))(") C (-[")](")	0.424	1 out of 1
ECFP_6	-1135409258	AND Enantiomer OH OH N H O [*]C([*)](C@@H)(C(=[*))[*)](c]1:[cH];[cH] :[cH]:[cH]:[cH]:[1	0.424	1 out of 1
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-388186450	AND Enantiomer OH OH N O [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.857	1 out of 12
ECFP_6	2014710090	AND Enantiomer OH OH O [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[1	-0.716	1 out of 10
ECFP_6	-1905455774	AND Enantiomer OH N N (*]C([*])C(=O)O	-0.552	1 out of 8



C₁₅H₁₆O₆

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Piroxicam	Triamcinolone	Sulfamethazine
Structure	OH HN N	OH OH OH	HN NH 2
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.

Feature Contribution						
	Top fe	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_6	464808839	AND Enantiomer HO OH OH [*]C(=CC(=[*])[*])[*]	0.524	8 out of 14		

ECFP_6	-200406221	AND Enantomer HO OH OH [*]C([*])([*])O[c](:[*]):[*]	0.451	3 out of 5
ECFP_6	-1051556861	AND Enantiomer HO OH OH [*]C[C@@H](O)C([*])([*])(*]	0.442	2 out of 3
		res for negative c		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-813997308	AND Enantiomer HO OH OH [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.452	2 out of 12
ECFP_6	1778376725	AND Enantiomer HO OH OH [*]OC(C)(C)C((*))[*]	-0.27	0 out of 1
ECFP_6	-1679469954	AND Enandomer HO OH OH [*][C@@H]1[*][c](:[*]):[c](OC1(C)C):[cH]: [*]	-0.27	0 out of 1

AND Enantiomer HO HO CI

C₁₇H₁₇CIO₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Naltrexone	Piroxicam	Tolazamide
Structure	HO M	OH HN P	T Z T T Z T T Z T T T T T T T T T T T T

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

Non-Carcinogen

Non-Carcinogen

Model Applicability

Actual Endpoint

Structural Similar Compounds

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

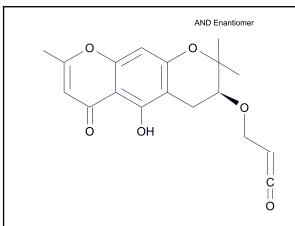
1. All properties and OPS components are within expected ranges.

Non-Carcinogen

- 2. Unknown ECFP_2 feature: -7106223: [*]C([*])[C@@H](CI)C(=[*])[*]
- 3. Unknown ECFP_2 feature: 1652274794: [*]OC(=C([*])[*])C([*])[*]

Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set ECFP_6 464808839 0.524 8 out of 14

ECFP_6	105634199	AND Enantiomer HO CI [*]C([*])CI	0.424	1 out of 1
ECFP_6	2106656448	AND Enantiomer HO CI [*]C(=O)[*]	0.254	31 out of 77
	Top Featu	res for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1795792463	AND Enantlomer HO CI [*]C([*])C(=C([*])[*])C(=[*])[*]	-0.657	0 out of 3
ECFP_6	305695353	AND Enantomer HO CI [*]C([*])C(O)C([*])[*]	-0.638	1 out of 9
ECFP_6	-1567907747	AND Enantiomer HO CI CI [*]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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sim	ilar Compounds		
Name	Naltrexone	Tolazamide	Indapamide
Structure	HO the state of th	T N N N N N N N N N N N N N N N N N N N	HN POOL
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=[*]

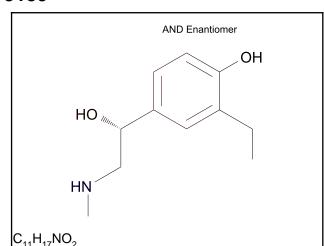
Feature Contribution

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-1114776580	AND Enantiomer OH C OH C	0.755	11 out of 15
ECFP_6	464808839	AND Enantiomer OH C:	0.524	8 out of 14
ECFP_6	-200406221	AND Enantiomer OH [*]C([*])([*])O[c](:[*]):[*]	0.451	3 out of 5
	Top Feat	ures for negative of	ontribution	1

ingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
CFP_6	-1250019913	AND Enantiomer OH C C C	-0.482	0 out of 2
CFP_6	-813997308	[*]COC([*])[*] AND Enantiomer OH OH [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.452	2 out of 12

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



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Phenylephrine	Albuterol	Procarbazine
Structure	HO	HO TO OH	N H N NH
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.

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-566124999	AND Enantiomer HO. OH HO. IT C[C@H](O)[c]1:[cH] :[cH]:[c](O):[c]([T]):[cH]:1	0.442	2 out of 3

ECFP_6	-1438409867	AND Enantiomer OH HO I*]C[C@H](O)[c]1:[cH] :[cH]:[*]:[c]([*)]:[cH]:1	0.337	3 out of 6
ECFP_6	628100036	AND Enantiomer HO (")NC(C@H)(O)(c)(:cH);");(eH);")	0.279	4 out of 9
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1334400011	AND Enantiomer HO . HN OH [*][c](:[*]):[c](O):[cH]:[*]	-0.496	3 out of 18
ECFP_6	493154328	AND Enantiomer OH HO, HN [*]CNC	-0.482	0 out of 2
ECFP_6	-1910270391	AND Enantiomer OH HO, [*]C([*])[*]	-0.307	20 out of 89

AND Enantiomer
OН
OH Imma
MilmOH

 $C_{17}H_{24}O_3$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Pindolol	Carteolol	Procarbazine
Structure	HN OH H	HN OH H	N H N NH
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	AND Enantiomer 0.297 OH OH (SH) OH (SH) OH (SH) (SH) OH (SH) (SH)	12 out of 28
ECFP_6	470495651	AND Enantiomer 0.2 OH OH OH [*]C([*])C=C([*])[*]	4 out of 10
ECFP_6	2022454958	AND Enantiomer 0.135	9 out of 25

Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	2025123907	AND Enantiomer OH OH OH (C) OH (C) OH (C) (C) (C) (C) (C) (C) (C) (C	-1.05	0 out of 6
ECFP_6	-1085223908	AND Enantiomer OH OH (a) (b) (c) (c) (d) (e) (d) (e) (e) (f) (f) (f) (f) (f) (f	-0.168	9 out of 35

ECFP_6	1559650422		-0.164	50 out of 191
		OH OH 1. L		
		ОН		
		[*]C[*]		

Simeprevir

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

AND Enantiomer
O NH O N N O N O O O O O O O O O O O O O
, N

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Reserpine	Bromocriptine	Itraconazole		
Structure	O The Control of the	Br What O O O O O O O O O O O O O O O O O O O			
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	AND Enantiomer AND Enantiomer N S N S N S I'IIC@@HJ1['I]'C]C[C@H	0.891	12 out of 14
ECFP_6	2082767335]1C(=[*])[*] AND Enantiomer	0.617	2 out of 2
		[*][c](:[*]):[c]1:[cH]:[cH]:[c]([*]):[*]: [c]:1:[*]		
ECFP_6	-1331088410	AND Enantiomer AND Enantiomer N S N N N N N N N N N N N N N N N N N	0.442	2 out of 3
	Ton For	[*]CCC=[*]	contribution	
		atures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	2077607946	AND Enantiomer AND Enantiomer (*)O[o]1:[cH]:[cH]:[c]((*)):[*]:[o]:1[*]	-1.15	0 out of 7	

ECFP_6	1526862590	AND Enantiomer AND Enantiomer N S N S N (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	-0.638	1 out of 9
ECFP_6	1307307440	AND Enantiomer AND Enantiomer NS NS N (*]:[c](:[*])OC	-0.558	4 out of 25

US FDA (Centre for Drug

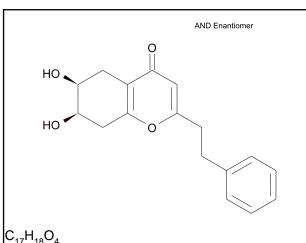
Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Phenobarbital	Oxazepam	Phenytoin		
Structure	HN O	CI NOH NOH	O N H		
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen		
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen		
Distance	0.601	0.605	0.635		

Model Applicability

Reference

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.

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

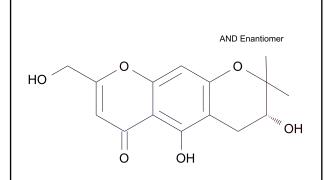
- 2. Unknown ECFP_2 feature: 1875238785: [*]C\C(=C\[*])\O[*]
- 3. Unknown ECFP_2 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
- 4. Unknown ECFP_2 feature: 1794461805: [*]CC(=C([*])[*])C(=[*])[*]
- 5. Unknown ECFP_2 feature: 1650944136: [*]CC(=C([*])[*])O[*]

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set

ECFP_4	2024749573	AND Enantiomer	0.494	8 out of 11
		HO		
		[*]C([*])O		
ECFP_4	-560785749	AND Enantiomer	0.351	1 out of 1
		[*]C(=[*])OC(=[*])[*]		
ECFP_4	1571214559	AND Enantiomer HO HO	0.283	9 out of 16
		[*]1:[cH]:[cH]:[cH]:1 cH]:[cH]:1		
	Top Fea	atures for negative	contribution	n
Fingerprint	Bit/Smiles	Feature Structure		Multiple- Carcinogen in training set
ECFP_4	1205550831	AND Entantiomer	-0.545	1 out of 6

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

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



 $C_{15}H_{16}O_{6}$

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Triamcinolone	Triamcinolone Zidovudine			
Structure	OH OH OH	HO TZZZ N N N N N N N N N N N N N N N N N	HN NH 2		
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 &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &		

Model Applicability

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

Res.) Sept. 1997

Res.) Sept. 1997

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

Res.) Sept. 1997

- 3. Unknown ECFP_2 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
- 4. Unknown ECFP_2 feature: 1778376725: [*]OC(C)(C)C([*])[*]

eature	Contri	bution	
		Ta	_

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	2024749573	AND Enantiomer HO OH OH [*]C([*])O	0.494	8 out of 11

ECFP_4	-560785749	AND Enantiomer HO OH OH [*]C(=[*])OC(=[*])[*]	0.351	1 out of 1
ECFP_4	-570915357	AND Enantlomer HO OH [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.351	1 out of 1
	Top Fea	atures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-813997308	AND Enantiomer HO OH OH [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]	-0.597	0 out of 2
ECFP_4	464808839	AND Enantiomer HO OH OH [*]C(=CC(=[*])(*))[*]	-0.352	2 out of 8
ECFP_4	769217534	AND Enantiomer HO OH OH [*]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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Alizapride	Phenolphthalein	Sulfamethazine		
Structure	N = N N N N N N N N N N N N N N N N N N	НО	HN Market No.		
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.

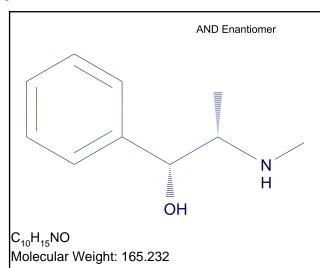
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -331149802: [*]O\C(=C/[*])\C
- 3. Unknown ECFP_2 feature: 1299558496: [*]C(=[*])C(=O)C=[*]
- 4. Unknown ECFP_2 feature: 1778376725: [*]OC(C)(C)C([*])[*]
- 5. Unknown ECFP_2 feature: -1591590376: [*]C=C=O
- 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 MultipleCarcinogen in training set

Fingernrint	Rit/Smiles	Feature Structure Score	Multiple-
	Top Feat	ures for negative contr	ibution
ECFP_4	-2124995946	AND Enantiomer 0.351	1 out of 1
ECFP_4	-560785749	AND Enantiomer 0.351	
		[*]O[c](:[cH]:[*]):[c]([*]):[*]	
ECFP_4	-570915357	AND Enantiomer 0.351	1 out of 1

lop Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-1925046727	AND Enantiomer OH C C C C O C	-0.605	2 out of 11
ECFP_4	-813997308	AND Enantiomer OH [*]C(=[*])[c](:[c]([*]):[*]):[c]([*)]:[*]	-0.597	0 out of 2

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



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

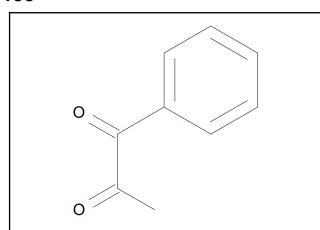
Structural Simi	lar Compounds		
Name	Ephedrine	Phenelzine	Amphetamine
Structure	OH N H	NH NH 2	NH ₂
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.000	0.484	0.487
Reference	US FDA (Centre for Drug Eval.& Res./Off. 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.

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

FCFP_6	0	AND Enantiomer OH [*]C([*])[*]	0.114	90 out of 305
FCFP_6	3	AND Enantiomer OH [*]O	0.105	62 out of 212
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1083860676	AND Enantiomer OH ("IC("))(C@H)(O)(c)1 :(cH):(cH);(cH);(cH) :(cH):1	-0.719	0 out of 4
FCFP_6	-239801958	AND Enantiomer : OH H ["N[C@@H](C)[C@H](O) [c]([cH];"])[cH]; ["]	-0.423	0 out of 2
FCFP_6	223071155	AND Enantiomer OH I*]C([*])[C@H](C)NC	-0.423	0 out of 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Coumarin	Amphetamine	Benzyl alcohol
Structure		NH ₂	OH
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.

Feature Co	ntribution			
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	565968762		0.245	7 out of 20
		[*]C(=[*])C(=O)C		

FCFP_6	1872154524		0.205	69 out of 213
FCFP_6	136597326	[*]C(=O)[*]	0.155	49 out of 159
		[*]C([*])C		
Fingerprint	Bit/Smiles	ures for negative of Feature Structure	Score	
		- Gararo Giragiano		Carcinogen in training set
FCFP_6	-1549192822		-0.489	3 out of 21
		[*]C(=[*])C(=O)[c](:[*]):[*]		
FCFP_6	975909016		-0.233	0 out of 1
		[*]C(=[*])C(=O)[o]1:[cH]:[cH]:[cH]:[cH]: cH]:1		
FCFP_6	-1698724694		-0.22	15 out of 72
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1		

AND Enantiomer O HO O C₁₇H₁₈O₄

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

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.

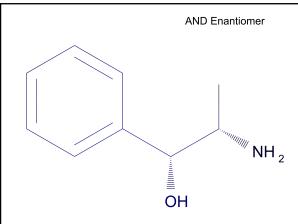
Structural Sim	ilar Compounds		
Name	Tolazamide	Naltrexone	Phenobarbital
Structure	T Z D D D D D D D D D D D D D D D D D D	HO 11	HN O
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.524	0.581	0.584
Reference	US FDA (Centre 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.

FCFP_6	859241794	AND Enantiomer HO	0.46	1 out of 1
FCFP_6	436886043	AND Enantiomer HO	0.306	15 out of 41
	Top Fea	atures for negative of	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1981711554	AND Enantiomer HO	-1.42	0 out of 12
FCFP_6	1388176727	AND Enantiomer HO	-1.21	0 out of 9
FCFP_6	-497728148	AND Enandomer HO	-0.96	2 out of 26



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

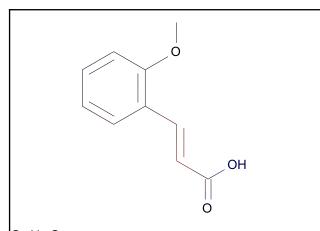
Name	Ephedrine	Phenelzine	Tocainide
Structure	N H	NH NH NH 2	NH NH NH 2
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.308	0.429	0.458
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US 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.

Feature Contribution				
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	136597326	AND Enantiomer NH2 OH [*]C([*])C	0.155	49 out of 159

FCFP_6	0	AND Enantiomer	0.114	90 out of 305
		· NH ₂		
		ОН		
		[*]C([*])[*]		
FCFP_6	3	AND Enantiomer	0.105	62 out of 212
		NH ₂		
		OH		
		[*]O		
		es for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1083860676	AND Enantiomer	-0.719	0 out of 4
		. NH2		
		ОН		
		[*]C([*))[C@H](O)[c]1 :[cH]:[cH]:[cH] :[cH]:1		
FCFP_6	-239801958	AND Enantiomer	-0.423	0 out of 2
		· NH ₂		
		OH		
		[*]N[C@@H](C)[C@H](O) [c](:[cH]:[*]):[cH]: [*]		
FCFP_6	-1931573337	AND Enantiomer	-0.387	2 out of 13
		· NH ₂		
		[*]C([*])(C@H](O)(c]1 :[cH]:[cH]:*]:[cH]: [cH]:1		



C₁₀H₁₀O₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

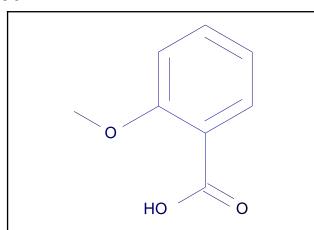
Name	Aspirin	Coumarin	Eugenol
Structure	O OH		OH OH
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.466	0.498	0.529
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.&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

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-146015125	[*]C(=[*]))C=C\[o](:[oH]:{(")}:[e]([")]:[*]	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]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH	0.46	1 out of 1
	Top Fea	tures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1458856986	[*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[o]:10C	-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



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

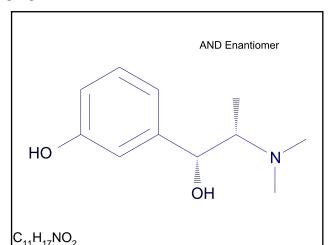
Name	Aspirin	Nicotinic acid	Coumarin
Structure	O OH	O OH	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.271	0.528	0.551
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.&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

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1872154524	НО	0.205	69 out of 213
		[*]C(=O)[*]		

FCFP_6	0	HO O	0.114	90 out of 305
FCFP_6	1	HO (*)=O	0.0783	76 out of 267
		res for negative of		
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	HO O	-0.383	9 out of 52
FCFP_6	-1549222613	[*]:[c](:[*])C(=O)O	-0.351	1 out of 7



ALogP: 1.528
Rotatable Bonds: 3

Molecular Weight: 195.258

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

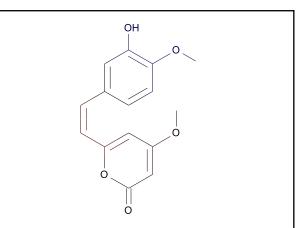
Name	Phenylephrine	Ephedrine	Tocainide
Structure	HO	OH N H	NH NH NH 2
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.458	0.463	0.534
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (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.

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-158888774	AND Enantiomer OH [*][c]1:[*]:[cH]:[cH] :[c](O):[cH]:1	0.367	5 out of 12

FCFP_6	136597326	AND Enantiomer	0.155	49 out of 159
		HO OH		
		[*]C([*])C		
FCFP_6	0	AND Enantiomer	0.114	90 out of 305
		HO OH		
		[*]C([*])[*]		
		es for negative c		
Fingerprint		Feature Structure		Carcinogen in training set
FCFP_6	-1083860676	AND Enantiomer	-0.719	0 out of 4
		но		
		[*]C((*))(C@H)(O)(c)1 :[cH]:[cH];[cH] :[cH]:1		
FCFP_6	-1946918893	AND Enantiomer	-0.692	2 out of 19
		но		
		[*]C([*))[C@H](C)N([*])[*]		
FCFP_6	-1343180157	AND Enantiomer	-0.533	9 out of 61
		HO OH N		
		[*]C([*])N(C)C		



C₁₅H₁₄O₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

on dotarar c	minar Compounds		
Name	Cytembena	Scopolamine	Omeprazole
Structure			
	OH Br	on N) H

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

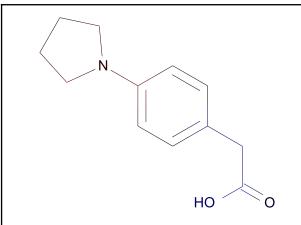
Structural Similar Compounds

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

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-146015125	[*]C(=[*])/C=C\[o](:[oH]:[*]):[o]([*]):[*]	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 Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	523826990	[*]O[c]1:[cH]:[*]:[c] ([*]):[cH]:[c]:10	-0.423	0 out of 2
FCFP_6	7	OH O\ (*)O	-0.308	15 out of 79
FCFP_6	136627117	[*]OC	-0.252	10 out of 50

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



C₁₂H₁₅NO₂

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

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Mexiletine	Methylphenidate	Eugenol		
Structure	NH 2	N H O	OH OH		
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.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1033057683	[*]C[e]1:[cH]:[cH]:[c](:[cH]:[cH]:1)N2C[*][*]C2	0.805	3 out of 3

FCFP_6	-822674211	HO 0	0.517	2 out of 3
FCFP_6	243588731	[*]1:[cH]:[cH]:[c](:[cH]:[cH]:1)N2CCCC2	0.46	1 out of 1
	Top Fea	tures for negative	contributio	n
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	HO O	-0.793	1 out of 13
FCFP_6	2039196715	[*]:[c](:[*])CC(=0)O	-0.582	0 out of 3

0.561

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

AND Enantiomer OH N H O

C₁₂H₁₃NO₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Phenobarbital	Carbromal	Phenytoin	
Structure	HN O	O NH 2 N H	H Z H	
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen	

Model Applicability

Distance

Reference

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

0.548

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

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

US FDA (Centre for Drug

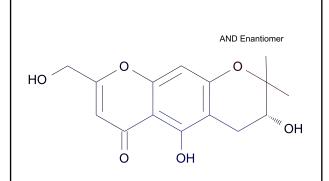
Eval.& Res./Off. Testing &

Res.) Sept. 1997

0.452

Feature Contribution				
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	566058135	AND Enantiomer OH N N OH N OH N OH OH OH OH	0.447	17 out of 40
		O=C1C[*][*]N1		

FCFP_6	-1043339860	AND Enantiomer OH NH NH O [*]C([*])C1C[*][*]C1	0.383	24 out of 61
FCFP_6	1872154524	AND Enantiomer OH N H O [*]C(=O)[*]	0.205	69 out of 213
		res for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-548632217	AND Enantiomer OH N N H O [*]C(=[*])O	-0.383	9 out of 52
FCFP_6	-1716639150	AND Enantiomer OH OH O (*)C((*))(C@@H)(C(=[*])(*))(d]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH	-0.351	1 out of 7
FCFP_6	7	AND Enentiomer OH N N H O [*]O	-0.308	15 out of 79



C₁₅H₁₆O₆

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Piroxicam	Sulfamethazine	Naltrexone
Structure	OH HN PAR N	HN NH 2	HO the last of the
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

i catale ool	catale contribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
FCFP_6	-1979984779	AND Enantiomer HO OH OH [*][C@@H]1[*][c](:[*]):[c](OC1(C)C):[cH]: [*]	0.668	3 out of 4		

FCFP_6	451847724		0.479	21 out of 48
		AND Enandiomer HO O OH OH		
		[*]C(=CC(=[*])[*])[*]		
FCFP_6	-415216134	AND Enantiomer HO OH	0.306	5 out of 13
		T T O OH		
	Top Featu	res for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1601875224	AND Enantiomer HO O OH OH	-0.582	0 out of 3
		[*]C[c]1:[c]([*]):[*] :[c]([*]):[c](C(=[*])[*]):[c]:1O		
FCFP_6	-2006448698	AND Enantiomer HO O OH OH	-0.582	0 out of 3
		[*][c](:[*]):[c]1C[C@ @H](O)C([*])([*])[*] [c]:1:[*]		
FCFP_6	-1549192822	AND Enantiomer HO O OH OH	-0.489	3 out of 21
		[*]C(=[*])C(=O)[c](:[*]):[*]		

AND Enantiomer HO HO CI

C₁₇H₁₇CIO₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Naltrexone	Tolazamide	Piroxicam
Structure	HO MAN NO	T Z T Z T Z T Z T Z T Z T Z T Z T Z T Z	OH HN N
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.

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set FCFP_6 71953198 AND Enanthomer 0.612 12 out of 23

FCFP_6	451847724	AND Enantiomer HO	0.479	21 out of 48
FCFP_6	859241794	AND Enantiomer HO CI [*]CC1=CC(=0)C(=[*]O1)[*]	0.46	1 out of 1
	Top Fea	atures for negative of	contribution	on
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1981711554	AND Enantiomer HO	-1.42	0 out of 12
FCFP_6	1388176727	AND Enantiomer HO OH O HO CI [*]C(=[*])CC[c]1:[cH] :[cH]:[cH]:[cH]:[:1	-1.21	0 out of 9
FCFP_6	-497728148	AND Enantiomer HO OH O	-0.96	2 out of 26

C₁₈H₁₈O₆

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Naltrexone	Tolazamide	Indapamide
Structure	HO the state of th	H Z H Z N N N N N N N N N N N N N N N N	HN H ₂ N O

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

Structural Similar Compounds

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

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set FCFP_6 -1979984779 0.668 3 out of 4

FCFP_6	451847724	AND Enantiomer OH C C C C C C	0.479	21 out of 48
FCFP_6	-415216134	AND Enantiomer OH (*)OC(C)(C)C([*])[*]	0.306	5 out of 13
		ures for negative of	ontribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1601875224	AND Enantiomer [*]C[c]1:[c]([*]):[*] :[c]([*]):[c](C(=[*])[*]):[c]:10	-0.582	0 out of 3
FCFP_6	-1549192822	AND Enantomer OH C C C O (*)C(=[*])C(=O)[c](:[*)):[*]	-0.489	3 out of 21
FCFP_6	-2000646149	AND Enantiomer OH (*)O[C@H]1C[c](:[c]([1]):[1]):[0:[1])!*	-0.423	0 out of 2

AND Enantiomer OH HO MM C₁₁H₁₇NO₂

ALogP: 1.557
Rotatable Bonds: 4

Acceptors: 3
Donors: 3

Model Prediction

Molecular Weight: 195.258

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Phenylephrine	Albuterol	Terbutaline
Structure	HO	HO TO OH	HO NO OH
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.

Feature Co	ntribution				
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	136597326	AND Enantiomer OH HO, HN [*]C([*])C	0.155	49 out of 159	

FCFP_6	0	AND Enantiomer OH HO, [*]C([*])[*]	0.114	90 out of 305
FCFP_6	3	AND Enantiomer HO	0.105	62 out of 212
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-451251206	AND Enantomer OH HO , HN [*]C[c]1:[cH]:[c](:[c H]:[*]:[c]:1[*])C([*])[*]	-0.731	1 out of 12
FCFP_6	-306804326	AND Enantiomer HO. HN [*]C[c]1:[cH]:[c](:[c H]:[cH]:[c]:10)C([*])[*]	-0.582	0 out of 3
FCFP_6	-616387365	AND Enartioner OH HO [*]C[c]1:[cH]:[c](:[c H]:[cH]:[c]:1[*])[C@ @H](O)CN[*]	-0.582	0 out of 3

AND Enantiomer
OH
OH Minne
Minne.
ОН

Molecular Weight: 276.371

ALogP: 2.165 Rotatable Bonds: 4

Acceptors: 3
Donors: 3

C₁₇H₂₄O₃

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 Simi	ilar Compounds		
Name	Pindolol	Carteolol	Procarbazine
Structure	HN OH H	HN OH H	N H N NH
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.

Feature Co	ntribution				
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	451847724	AND Enantiomer OH OH (*)C(=CC(=[*])[*)](*)	0.479	21 out of 48	

FCFP_6	-1043339860	AND Enantiomer OH OH OH (*]C([*])C1C[*][*]C1	0.383	24 out of 61
FCFP_6	1000913921	AND Enantiomer OH OH (*)C[C@H]1C=C(C)(C@@H)((*))(*)[C@@H]1(*)	0.38	2 out of 4
	Top Feat	ures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1981711554	AND Enantiomer OH OH OH (*)CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.42	0 out of 12
FCFP_6	-497728148	AND Enantiomer OH OH (*)CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.96	2 out of 26
FCFP_6	907007053	AND Enentiomer OH OH OH OH (*]CC[c](:[*]):[*]	-0.366	11 out of 62

Simeprevir

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen

AND Enantiomer
O NH N S N N O N N O O O O O O O O O O O O O
, N

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Reserpine	Bromocriptine	Itraconazole
Structure	O SALVE MAN O SALV	Br What Name of the state of th	
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

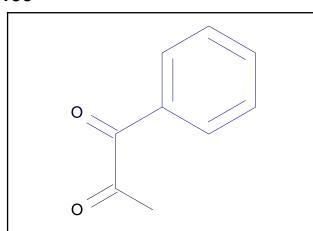
Model Applicability

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 Co	ntribution			
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-55265897	AND Enerdiomer AND Enerdiomer NS N O O (*]S(=[*])(=[*])C1CC1	0.594	17 out of 34

FCFP_6	-1289661876	AND Enantiomer NS N N O N O N O N O N O N O N O N O N	0.517	2 out of 3
		[*]CCC\C=C/[*]		
FCFP_6	451847724	AND Enandomer AND Enandomer	0.479	21 out of 48
		[*]C(=CC(=[*])[*])[*]		
		tures for negative o		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-98332825	AND Enantiomer AND Enantiomer NS N S N S N S N S N S N S N S N S N	-0.793	1 out of 13
FCFP_6	-1972798083	AND Enantiomer AND Enantiomer N S N O O O O O O O O O O O O O O O O O	-0.582	0 out of 3
FCFP_6	-1553874037	AND Enantiomer AND Enantiomer NS N S N O N O N O N O N O N O N O N O	-0.45	5 out of 32



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

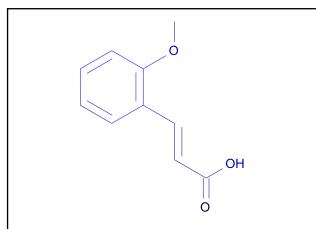
Name	Coumarin	Toluidine; o-	Phenacetin
Structure		NH 2	O H
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.528	0.637	0.642
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 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.

ntribution			
Top fe	atures for positive o	ontribution	
Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
565968762	[*]C(=[*])C(=O)C	0.168	3 out of 7
	Top fe Bit/Smiles	Top features for positive of Bit/Smiles Feature Structure	Top features for positive contribution Bit/Smiles Feature Structure Score 565968762 0.168

FCFP_12	136597326		0.0722	18 out of 49
		[*]C(=[*])C		
	Top Featur	es for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1549192822		-0.704	0 out of 3
FCFP_12	991735244	["]C(=["])C(=O)]e](:["]):["]	-0.562	5 out of 28
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1		
FCFP_12	-1698724694		-0.423	3 out of 15
		[*]C(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1		



 $|C_{10}H_{10}O_3|$

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Coumarin	Phenacetin	Methylphenidate	
Structure		O H	N H O O	
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen	

Single-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.548

Model Applicability

Predicted Endpoint

Distance

Reference

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

Single-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.559

Single-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.580

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

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	451847724	[*]\C=C\C(=[*])[*]	0.3	10 out of 21

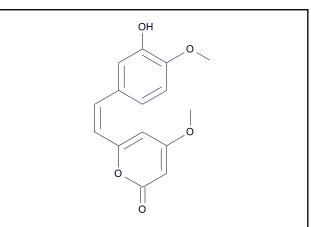
FCFP_12	-1977641857		0.105	2 out of 5
		ОН		
		[*][c](:[*]):[c](OC):		
		[cH]:[*]		
	Top Featur	es for negative o	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-548632217	1 0	-1.4	0 out of 9
		ОН		
		0		
		[*]C(=[*])O		
FCFP_12	-1176841573	_ I	-1.31	0 out of 8
		ОН		
		0		
		[*]=CC(=O)O		
FCFP_12	-1099193755		-1.11	0 out of 6
		ОН		
) Tr		
		[*]=C[c]1:[cH]:[*]:[c		
		H]:[cH]:[c]:1OC		

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Eval.& Res./Off. Testing &

Res.) Sept. 1997



C₁₅H₁₄O₅

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sim	Structural Similar Compounds					
Name	Nitroacetophenetide	Methylphenidate	Tramadol			
Structure	o No	N H O O	N OH			
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	US FDA (Centre for Drug	US FDA (Centre for Drug			

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.

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Feature Contribution				
	Top fe	atures for positive o	ontribution	
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]:10	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	OH 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]:[*	-0.409	13 out of 59	

AND Enantiomer OH

 $|C_{12}H_{13}NO_3|$

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

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Phenobarbital	Acetaminophen	Tranexamic acid
Structure	HN O	HONH	OH ONH ₂
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

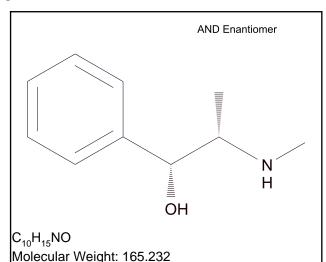
Model Applicability

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

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

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	1186303932	AND Enantiomer OH OH O ("]C(("))(C@@H)(C(=[" D)(")[C]-(-[")["]	0.497	4 out of 6

FCFP_12	-1043339860	AND Enantiomer	0.349	12 out of 24
		OH OH		
		N—O		
		[*]C([*])C1C[*][*]C1		
	Top Feat	tures for negative	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-548632217	AND Enantiomer	-1.4	0 out of 9
		OH OH		
		N—NO		
		[*]C(=[*])O		
FCFP_12	-1176841573	AND Enantiomer	-1.31	0 out of 8
		OH OH		
		N—O		
		[*]=CC(=O)O		
FCFP_12	7	AND Enantiomer	-0.71	2 out of 15
		OH 		
		N—O		
		[*]0		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

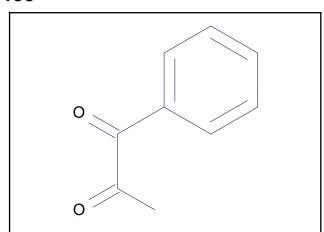
Structural Similar Compounds				
Name	ETHANOL;2- (PHENETHYLAMINO)-		Ethanol; 2-anilino-	
Structure	NH	NH NH	NH	
Actual Endpoint	Moderate_Severe	Moderate_Severe	Moderate_Severe	
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild	
Distance	0.430	0.439	0.486	
Reference	AMIHBC 10;61;54	JIHTAB 31;60;49	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.

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-885550502	AND Enantiomer in the second of the second	0.239	54 out of 64

FCFP_10	1070061035	AND Enantiomer	0.239	284 out of 338
		[*]C([*])O		
FCFP_10	-1043250487	AND Enantiomer OH (")N C@@H (C)C((")) * J	0.22	62 out of 75
	Top Feat	tures for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1698724694	AND Enantiomer OH H [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	136686699	AND Enantiomer OH [*]NC	-0.243	6 out of 12
FCFP_10	991735244	AND Enantiomer OH [*][c]1:[*]:[cH]:[cH]	-0.185	130 out of 237
		:[cH]:[cH]:1		



C₉H₈O₂

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

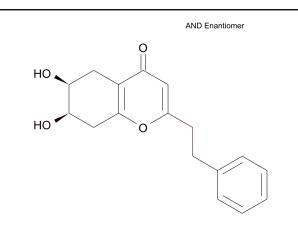
Name	ACETOPHENONE	METHYL BENZOATE	Propiophenone
Structure	0	0	0
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.312	0.382	0.394
Reference	AJOPAA 29;1363;46	AMIHBC 10;61;54	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;291;86

Model Applicability

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

Feature Contribution				
	Top Fe	atures 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
			1	L

FCFP_10	-1698724694	0 [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	975909016	["]C(=["])C(=0)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.216	3 out of 6



C₁₇H₁₈O₄

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 nonnormal 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	OH OH	OH NO H	HN MH 2	
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	

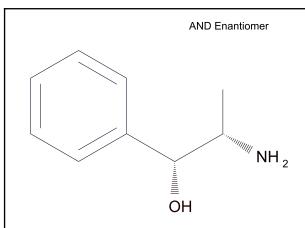
Model Applicability

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

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	1388176727	AND Enantiomer HO HO 'O (eH) (eH)	0.389	19 out of 19	

FCFP_10	-497728148	AND Enandomer HO HO O [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.356	24 out of 25
FCFP_10	1070061035	HO HO ([*])O	0.239	284 out of 338
	Top Fea	tures for negative	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	565968762	AND Enantiomer HO HO O [*]C(=[*])C(=O)C	-0.372	17 out of 38
FCFP_10	-1698724694	AND Enantiomer HO HO O [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	436915834	AND Enantiomer HO HO O	-0.194	2 out of 4

[*]C\C(=C\[*])\O[*]



C₉H₁₃NO

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

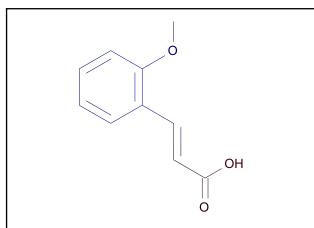
Structural Similar Compounds				
Name	3-AMINO-2-METHYL BENZYL ALCOHOL	1;2-DIMETHYL;1;3- BUTANEDIOL	METHYLAMINE;M- PHENYLENEBIS-	
Structure	H ₂ N ^M OH	НО	NH 2 NH 2	
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe	
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe	
Distance	0.341	0.459	0.461	
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.

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1070061035	AND Enantiomer NH ₂ [*]C([*])O	0.239	284 out of 338

FCFP_10	-1043250487	AND Enantiomer	0.22	62 out of 75
		NH ₂		
		[*]N[C@@H](C)C([*])[*]		
FCFP_10	3	AND Enantiomer	0.165	383 out of 491
		OH NH ₂		
		[*]0		
	Top Feat	cures for negative (contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1698724694	AND Enantiomer	-0.284	53 out of 107
		.,NH ₂		
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1		
FCFP_10	991735244	AND Enantiomer NH 2	-0.185	130 out of 237
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1		
FCFP_10	-453677277	AND Enantiomer NH 2	-0.13	153 out of 264
		[*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1		



 $|C_{10}H_{10}O_3|$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

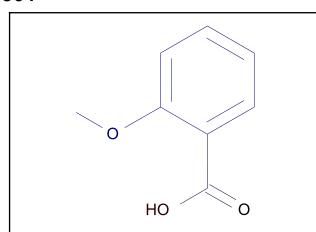
Structural Similar Compounds					
Name	MANDELIC ACID; ISOPROPYL ESTER	MANDELIC ACID; ETHYL ESTER	SALICYCLIC ACID; METHYL ESTER		
Structure	O OH	ОН	OH O		
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild		
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild		
Distance	0.469	0.474	0.475		
Reference	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.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1176841573	 OH OH ST=CC(=O)O	0.356	41 out of 43

FCFP_10	-2107131107	OH OOH	0.344	6 out of 6
FCFP_10	-548632217	[*]C=C\C(=O)O	0.319	54 out of 59
	Top Feat	ures for negative of	contribution	1
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	CH]:[1]):[c]([1]):[*	-0.507	0 out of 1



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

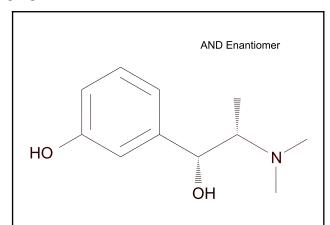
Structural Similar Compounds				
Name	SALICYCLIC ACID; METHYL ESTER	BENZOIC ACID	RESORCINOL; MONOACETATE	
Structure	OH O	OH	HON	
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe	
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe	
Distance	0.355	0.370	0.386	
Reference	28ZPAK-;106;72	BIOFX* 28-4/73	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.

Feature Co	ntribution			
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-548632217	HO O [*]C(=[*])O	0.319	54 out of 59

FCFP_10	7	HO (*)O	0.219	117 out of 142
FCFP_10	-1549222613	HOO [*]:[c](:[*])C(=O)O	0.157	11 out of 14
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1458856986	HOOO [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:10C	-0.842	0 out of 2
FCFP_10	-1977641857	[*][c](:[*]):[c](OC): [cH]:[*]	-0.78	4 out of 15
FCFP_10	-1099193755	HOOO [*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:10C	-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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

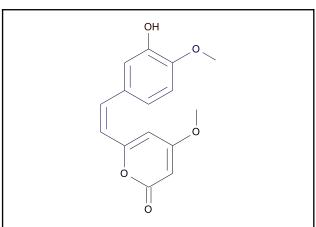
Structural Similar Compounds				
Name	o-Toluidine; N;N-bis(2- hydroxyethyl)-	ETHANOL; 2;2'- CYCLOHEXYLIMINODI-	BENZOIN; OXIME	
Structure	HO OH	OH OH	OH OH	
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild	
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild	
Distance	0.482	0.513	0.526	
Reference	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.

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-158888774	AND Enantiomer	0.356	24 out of 25

FCFP_10	-1849700223	AND Enantiomer OH [*][c]1:[cH]:[cH]:[cH]:[c](O):[cH]:1	0.344	6 out of 6
FCFP_10	136388789	AND Enantiomer HO OH [*]N([*])C		28 out of 30
Eingarnrint	Top Feat	tures for negative of Feature Structure	Score	
Fingerprint	Bit/Sillies	reature Structure		Moderate_Severe in training set
FCFP_10	-1694796561	AND Enantiomer HO OH [*]C([*])[c]1:[cH]:[c H]:[cH]:[c](O):[cH]:	-0.6	1 out of 4
FCFP_10	991735244	AND Enantiomer : HO : OH [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.185	130 out of 237
FCFP_10	-453677277	AND Enantiomer : HO :: OH [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:1	-0.13	153 out of 264



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

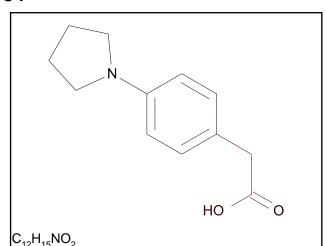
Structural Similar Compounds				
Name	2;8;9-TRIOXA-5-AZA-1- SILABICYCLO(3.3.3)UNDE CANE;1-(3- AMINOPROPYL)-	2;5-DICHLORO-4(3'- METHYL-5' PYRAZOLON- 1'-YL)BENZENE SULFONIC ACID	BENZOIC ACID; 5- (CHLOROSULFONYL)-2;4- DICHLORO-	
Structure	NH 2	O I I N I N I N I N I N I N I N I N I N	CI O OH	
Actual Endpoint	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.

Feature Contribution				
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	7	OH O\ \(\circ\)	0.219	117 out of 142

FCFP_10	523826990	[*]O[c]1:[cH]:[*]:[c] ([*]):[cH]:[c]:10	0.186	1 out of 1
FCFP_10	74595001	[*]:[cH]:[c](O):[cH]: [*]	0.16	63 out of 81
	Top Fea	tures for negative of	contribution	1
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\[o](:[cH]:[*]):[o]([*]):[*	-0.507	0 out of 1
FCFP_10	-747977902	[*]O[c]1:[cH]:[c]](\C=C/[*]):[cH]:[c] :1[*]	-0.507	0 out of 1



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

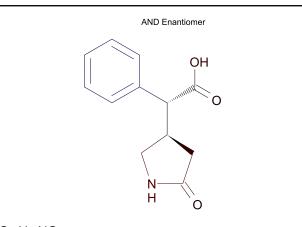
Structural Similar Compounds				
Name	1-NAPHTHALENEACETIC ACID	Propionitrile; 3-(N-(2- hydroxyethyl)anilino)-	MANDELIC ACID; ISOPROPYL ESTER	
Structure	OH OH	N N OH	OH OH	
Actual Endpoint	Moderate_Severe	Mild	Moderate_Severe	
Predicted Endpoint	Moderate_Severe	Mild	Moderate_Severe	
Distance	0.492	0.492	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.

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]:[c H]:[c](CC(=[*])[*]): cH]:[cH]:1	0.385	16 out of 16	

FCFP_10	-1176841573	HO OO OO OO	0.356	41 out of 43
FCFP_10	-497728148	[*]=CC(=O)O *]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.356	24 out of 25
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-822674211	Ho o	-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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	ACETANILIDE; 3'-AMINO- 4'-ETHOXY-	ACETIC ACID; CYCLOHEXYLAMINO-	Anthranilic acid; N-methyl-	
Structure	O No	NH OH		
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;8	

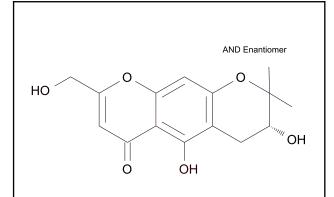
Model Applicability

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

Feature Contribution					
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-1176841573	AND Enantiomer OH N N H O [*]=CC(=O)O	0.356	41 out of 43	

FCFP_10	-416918913	AND Enantiomer OH OH ("1C("))(C@@H)1CNC(= ["))C1	0.332	5 out of 5	
FCFP_10	-548632217	AND Enantiomer OH N OH (*]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	AND Enantiomer OH N O ("C("))(C@@H)(C(= " D("))(G)(1:[GH);[GH] :["]:[GH];[GH]:1	-0.507	0 out of 1	
FCFP_10	-1716639150	AND Enantiomer OH N O (*]C((*))(C@@H)(C(=1* D(*))(G1:[cH]:[cH] :[cH]:[cH]:[cH]:	-0.507	0 out of 1	
FCFP_10	-1698724694	AND Enantiomer	-0.284	53 out of 107	

[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1



 $C_{15}H_{16}O_{6}$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

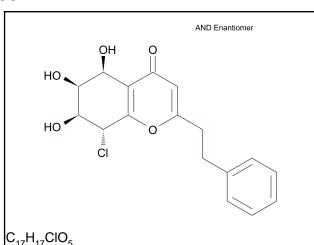
Structural Similar Compounds				
Name	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	ANTHRAQUINONE; 1-((2- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE	
Structure	OH NH	OH HN Th	O OH O OH O OH	
Actual Endpoint	Mild	Mild	Mild	
Predicted Endpoint	Mild	Mild	Mild	
Distance	0.614	0.647	0.650	
Reference	287PAK 190:72	287PAK 245:72	287PAK-:103:7	

Model Applicability

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

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-1601875224	HO OH OH OH	0.352	7 out of 7	

FCFP_10	-1272709286	AND Enantiomer HO O O	0.285	234 out of 266
		0 OH OH		
		[*][C@H]1[*][*]NC1		
FCFP_10	1679744180	AND Enantiomer HO OH OH	0.256	2 out of 2
		[*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]: 1[*]		
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-2006448698	AND Enantomer HO OH OH	-0.507	0 out of 1
		[*][o](:[*]):[o]10[0@ @H](0)0([*])([*])[*] [o]:1:[*]		
FCFP_10	-1099193755	AND Enantiomer HO OH OH	-0.361	2 out of 5
		[*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:1OC		
FCFP_10	436915834	AND Enantiomer HO OH OH	-0.194	2 out of 4
		[*]C\C(=C\[*])\O[*]		



ALogP: 1.368
Rotatable Bonds: 3

Molecular Weight: 336.767

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- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	
Structure	OH HN TAN H	HO OH	OH NH	
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	

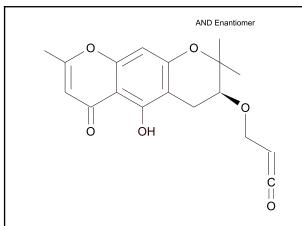
Model Applicability

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

Feature Contribution				
	Top fea	atures for positive o	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1388176727	AND Enantiomer HO CI [*]C(=[*])CC[c]1:[cH] :[cH]:[cH]:[cH]:[cH]:[cH]	0.389	19 out of 19

FCFP_10	-497728148	AND Enantiomer HO CI [*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.356	24 out of 25
FCFP_10	1070061035	AND Enantiomer HO CI [*]C([*])O	0.239	284 out of 338
	Top Feat	tures for negative of	ontribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	565968762	AND Enantiomer HO CI	-0.372	17 out of 38
FCFP_10	-1698724694	[*]C(=[*])C(=O)C AND Enantiomer HO CI [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-0.284	53 out of 107
FCFP_10	436915834	AND Enantiomer HO CI	-0.194	2 out of 4

[*]C\C(=C\[*])\O[*]



C₁₈H₁₈O₆

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 nonnormal 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	HN MH 2	HO to the hold of	OHCI CI CI OH	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance	0.608	0.610	0.625	

Model Applicability

Reference

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

28ZPAK 239;72

28ZPAK-:92:72

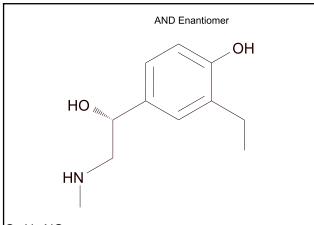
1. All properties and OPS components are within expected ranges.

28ZPAK-:124:72

2. Unknown FCFP_2 feature: 444624378: [*]C=C=O

Feature Contribution				
	Top fea	tures for positive o	ontribution	
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	AND Enantiomer OH C C O [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]: 1[*]	0.256	2 out of 2
FCFP_10	7	AND Enantiomer OH C C C O (*) OH	0.219	117 out of 142
	Top Feat	ures for negative of	ontribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1099193755	AND Enantomer O OH C C C C O [*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:10C	-0.361	2 out of 5
FCFP_10	436915834	AND Enantiomer OH C C C C C C C C C C C C C C C C C C C	-0.194	2 out of 4
FCFP_10	-1549192822	AND Enantiomer OH C C OH [*]C(=[*])C(=O)[c](:[*]):[*]	-0.11	7 out of 12



C₁₁H₁₇NO₂

Molecular Weight: 195.258

ALogP: 1.557
Rotatable Bonds: 4

Acceptors: 3
Donors: 3

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

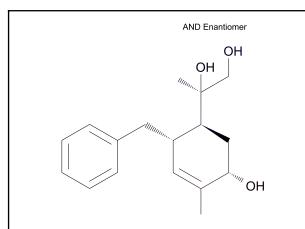
Structural Similar Compounds				
Name	2;3-DIMETHYL;1;2;6- HEXANETRIOL	p-Phenylenediamine; N;N- bis(2-hydroxyethyl)-; sulfate (1:1)	METHANE;TRIS(4- AMINOPHENYL)-	
Structure	HO OH OH	OH HO NH ₂	NH ₂ NH ₂ NH ₂	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance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

Model Applicability

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

Feature Co	Feature Contribution			
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	226619748	AND Enantiomer OH HO, HN [*]C([*])CNC	0.294	3 out of 3

FCFP_10	-306804326	AND Enantiomer HO. HIN [*]C[c]1:[cH]:[c](:[c H]:[cH]:[c]:10)C([*])[*]	0.294	3 out of 3
FCFP_10	-1272709286	AND Enantiomer OH HO, HN [*][C@H]1[*][*]NC1	0.285	234 out of 266
	Top Feat	tures for negative of	ontributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	136686699	AND Enantiomer OH HO, [*]NC	-0.243	6 out of 12
FCFP_10	-1604301295	AND Enantiomer OH HN [*]C[c]1:[cH]:[*]:[cH]:[cH]:[c]:10	-0.164	15 out of 27
FCFP_10	-453677277	AND Enantiomer OH HN [*]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 nonnormal 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	NH ₂ NH ₂ NH ₂	OH HN 11 N H	OH NOH	
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.

Feature Contribution				
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-497728148	AND Enantiomer OH OH (*)CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.356	24 out of 25

FCFP_10	-1272709286	AND Enantiomer OH	0.285	234 out of 266		
		OH				
		[*][C@H]1[*][*]NC1				
FCFP_10	1070061035	AND Enantiomer OH	0.239	284 out of 338		
		OH				
		[*]C([*])O				
Top Features for negative contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set		
FCFP_10	-2097085266	AND Enantiomer OH OH OH	-0.507	0 out of 1		
		[*]C([*])[C@@](C)(O)C O				
FCFP_10	-255810400	AND Enantiomer OH OH OH	-0.507	0 out of 1		
		[c@@](c)(O)CO [*]C[C@@H](C([*])[*])				
FCFP_10	-1698724694	AND Enantlomer OH	-0.284	53 out of 107		
		ОН.				

[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1

Simeprevir

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe

AND Enantiomer
O NH N S N O N O O O O O O O O O O O O O O O
1

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal 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	H H H		NH 2 HN 4th		
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

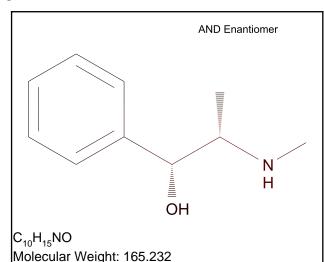
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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	AND Enantiomer NS NS ("JCCN(C)C(=0)C(["])[0.294	3 out of 3	

FCFP_10	-1474971978	AND Enantiomer AND Enantiomer NS N N O O O O O O O O O O O O O O O O	0.259	14 out of 16
FCFP_10	-124655670	AND Enantiomer AND Enantiomer (*)[c]1:[*]:[*]:s:[cH]:1	0.259	14 out of 16
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1977641857	AND Enantiomer AND Enantiomer [*][c](:[*]):[c](OC): [cH]:[*]	-0.78	4 out of 15
FCFP_10	-528918648	AND Enantiomer NS NN N N N N N N N N N N N N N N N N	-0.651	4 out of 13
FCFP_10	690481386	AND Enantiomer AND Enantiomer AND Enantiomer (*):[c](:[*])[c]1:n:[*]:[*]:s:1	-0.6	1 out of 4



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 nonnormal 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	NH OH	NH NH	NH		
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.

tructure Score	Severe in training set
	I .
0.400	
D Enantiomer 0.469	21 out of 21
C)C([*])[*]	
1	OH H

SCFP_12	5	AND Enantiomer	0.337	129 out of 151
		OH H		
		On On		
2055 42	4700704404	[*]N[*]	0.000	h
SCFP_12	1702724181	AND Enantiomer	0.303	2 out of 2
		N/N/		
		он Н		
		["]C(["])(C@H](O)(c](["]:("]:		
	Ton Foot	tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training
				set
SCFP_12	1653911926	AND Enantiomer	-0.0644	33 out of 58
		OH H		
		[*][c]1:[cH]:[cH]:[cH		
]:[cH]:[cH]:1	<u> </u>	
SCFP_12	0	AND Enantiomer	0	463 out of 727
		OH H		
		OII		
2055 10		[*]C([*])[*]		100 1 1000
SCFP_12	3	AND Enantiomer	0	162 out of 280
		OH H		
		[*][c](:[*]):[*]		
		[*][c](:[*]):[*]		

C₁₇H₁₈O₄

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

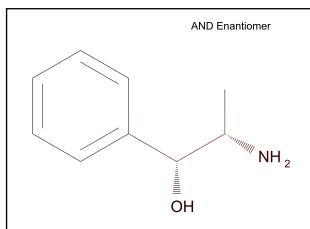
Structural Similar Compounds					
Name	2- BIPHENYLCARBOXYLIC ACID; 2'- HYDROXYMETHYL-	1;2-ETHANEDIOL;2;4- DICHLOROPHENOXY-	o-Toluidine; N;N-bis(2- hydroxyethyl)-		
Structure	OH OH	CI OH OH	HOOH		
Actual Endpoint	Moderate	Severe	Severe		
Predicted Endpoint	Moderate	Severe	Moderate		
Distance	0.594	0.633	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.

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set	
SCFP_12	-1640858361	AND Enantiomer HO	0.376	4 out of 4	

SCFP_12	2019034095	AND Enantiomer	0.345	15 out of 17		
		но				
		[*]C(=[*])OC(=[*])[*]				
SCFP_12	-1272709286	AND Enantiomer	0.231	24 out of 31		
		HO				
		[*]CC[c](:[*]):[*]				
Top Features for negative contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set		
SCFP_12	13	AND Enantiomer	-0.105	185 out of 338		
		HO				
		[*]=0				
SCFP_12	-1980361709	AND Enantiomer	-0.103	1 out of 2		
		HO				
		[*]C(=[*])C(=O)C=[*]				
SCFP_12	1311071855	AND Enantiomer	-0.0866	146 out of 262		
		HO				
		[*]C(=O)[*]				



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

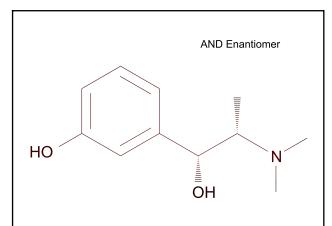
Structural Similar Compounds				
Name	METHYLAMINE;M- PHENYLENEBIS-	METHYLAMINE;P- PHENYLENEBIS-	1;2-DIMETHYL;1;3- BUTANEDIOL	
Structure	NH 2 NH 2	H ₂ N NH ₂	НО	
Actual Endpoint	Severe	Severe	Moderate	
Predicted Endpoint	Severe	Severe	Severe	
Distance	0.476	0.484	0.494	
Reference	28ZPAK-;64;72	28ZPAK-;64;72	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.

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1430588017	AND Enantiomer NH 2 OH [*]NIC@@HI(C)C([*])[*	0.469	21 out of 21

SCFP_12	260714409	AND Enantiomer	0.409	49 out of 53
		·NH ₂		
		OH -		
		[*10/[*1\N]		
SCFP_12	5	[*]C([*])N AND Enantiomer	0.337	129 out of 151
JOI 1 _ 1.2				120 000 01 10 1
		· NH ₂		
		[*]N[*]		
		es for negative c		
ingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1653911926	AND Enantiomer	-0.0644	33 out of 58
		., _{NH2}		
		он		
		[*][c]1:[cH]:[cH]:[cH		
]:[cH]:[cH]:1		
SCFP_12	-496409612	AND Enantiomer	o I	158 out of 272
		NH ₂		
		[*][c](:[*]):[cH]:[cH		
]:[*]		
SCFP_12	136597326	AND Enantiomer	0	236 out of 373
		OH NH ₂		
		[*]C([*])C		



C₁₁H₁₇NO₂

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

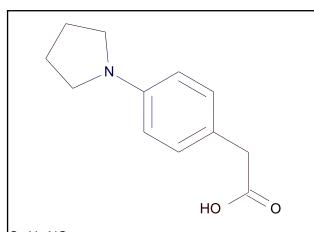
Structural Similar Compounds					
Name	o-Toluidine; N;N-bis(2- hydroxyethyl)-	ETHANOL; 2;2'- CYCLOHEXYLIMINODI-	ETHANOL; 2;2'- (PHENYLIMINO)DI-		
Structure	HOOH	OH OH	OH		
Actual Endpoint	Severe	Severe	Severe		
Predicted Endpoint	Moderate	Severe	Moderate		
Distance	0.476	0.509	0.531		
Reference	Prehled 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.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1430588017	AND Enantiomer HO OH	0.469	21 out of 21
		[*]N[C@@H](C)C([*])[*]		

SCFP_12	125999298	AND Enantiomer OH [*]C([*])[c]1:[cH]:[*]:[cH]:[c](O):[cH]:1	0.42	7 out of 7
SCFP_12	5	AND Enantiomer HO OH [*]N[*]	0.337	129 out of 151
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1379591900	AND Enantiomer : HO :: OH :: [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	0	77 out of 130
SCFP_12	-424425761	AND Enantiomer HO OH [*]:[c](:[*])O	0	40 out of 63
SCFP_12	560173167	AND Enantiomer OH [*]:[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 nonnormal 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	OH	OH O	O O O O O O O O O O O O O O O O O O O	
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.

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-658389839	*][c]1:[cH]:[cH]:[c] (:[cH]:[cH]:1)N2CCCC 2	0.303	2 out of 2

SCFP_12	-2056996303	\text{\text{N}}	0.303	2 out of 2
		но		
		[*]N1CCCC1		
SCFP_12	-424485343		0.263	43 out of 54
		но		
		[*]C(=[*])O		
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1334669481	HOOO	-0.685	28 out of 93
		[*]N([*])[o](:[cH]:[*]):[cH]:[*]		
SCFP_12	-1380909229	HOOO	-0.506	26 out of 72
		[*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH] :1		
SCFP_12	10	HOO	-0.451	43 out of 112
		[*]N([*])[*]		

AND Enantiomer OH N H O

 $|C_{12}H_{13}NO_3|$

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

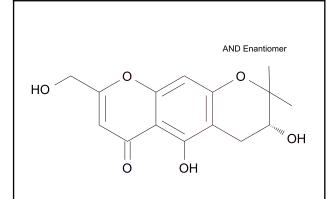
Name	Anthranilic acid; N-methyl-	ACETANILIDE. 4'-AMINO-	ETHANOL; 2;2'- (PHENYLIMINO)DI-
Structure	O OH	H ₂ N NH	OH
Actual Endpoint	Moderate	Moderate	Severe
Predicted Endpoint	Moderate	Moderate	Moderate
Distance	0.557	0.596	0.609
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;741;86	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.

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-424485343	AND Enantiomer OH N O [*]C(=[*])O	0.263	43 out of 54

SCFP_12	-1631132401	AND Enantiomer OH OH N N (*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	0.104	19 out of 28
SCFP_12	2109165795	AND Enantiomer OH N N (*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.0868	79 out of 119
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	2005026407	AND Enantiomer OH N N H O [*][C@@H]1CNC(=0)C1	-0.796	0 out of 2
SCFP_12	-587569116	AND Enantiomer OH NH O [*][C@H]1[*][*]NC1	-0.619	11 out of 35
SCFP_12	1256995004	AND Enantiomer OH N 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

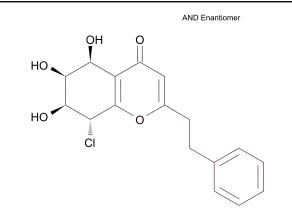
Structural Similar Compounds					
Name	p-Acetophenetidide; 3'- (bis(2- hydroxyethyl)amino)-		2- NAPHTHALENESULFONI C ACID;5-AMINO-6- ETHOXY-		
Structure	O N N O O O O O O O O O O O O O O O O O	H ₂ N _{th}	HO S NH ₂		
Actual Endpoint	Moderate	Moderate	Moderate		
Predicted Endpoint	Moderate	Moderate	Moderate		
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.

Feature Contribution				
Top fea	atures for positive o	ontribution		
Bit/Smiles	Feature Structure	Score	Severe in training set	
130348166	AND Enantioner HO OH OH [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]: 1[*]	0.376	4 out of 4	
	Top fea	Top features for positive of Bit/Smiles Feature Structure 130348166 AND Enarritomer HO OH [*]O[c]1:[cH]:[c](O[*]):[c](([*)):[*]:[c]:	Top features for positive contribution Bit/Smiles Feature Structure Score 130348166 AND Enantiomer HO	

SCFP_12	-1272709286		0.231	24 out of 31
		AND Enantiomer HO OH OH		
		[*]CC[c](:[*]):[*]		
SCFP_12	616547045		0.218	1 out of 1
		HO OH OH		
		[*]C\C(=C\[*])\O[*]		
	Top Feat	tures for negative (contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1416196903	AND Enantiomer	-0.383	3 out of 8
		HO OH OH		
		[*]OC(C)(C)C([*])[*]		
SCFP_12	-1980302127	AND Enantiomer HO OH OH	-0.345	1 out of 3
		[*]=CC(=O)[e](:[*]):[*]		
SCFP_12	-2056718782	AND Enantioner HO OH OH	-0.163	20 out of 39
		[*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]		



C₁₇H₁₇CIO₅

Molecular Weight: 336.767

ALogP: 1.368
Rotatable Bonds: 3

Acceptors: 5
Donors: 3

Model Prediction

Mahalanobis Distance: 11.6

Prediction: Severe Probability: 0.728 Enrichment: 1.17 Bayesian Score: 1.23

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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	HO OH	NH ₂ NH ₂ NH ₂	OHCI CI CI OH	
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.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1640858361	AND Enantiomer HO CI [*]CC[c]1:[cH]:[cH]:[0.376	4 out of 4

SCFP_12	2019034095	AND Enantiomer HO CI [*]C(=[*])OC(=[*])[*]	0.345	15 out of 17
SCFP_12	540920128	AND Enantiomer HO	0.303	2 out of 2
		res for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	13	AND Enantiomer HO CI [*]=O	-0.105	185 out of 338
SCFP_12	-1980361709	AND Enantiomer HO CI [*]C(=[*])C(=O)C=[*]	-0.103	1 out of 2
SCFP_12	1311071855	AND Enantiomer HO CI [*]C(=O)[*]	-0.0866	146 out of 262

C₁₈H₁₈O₆ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

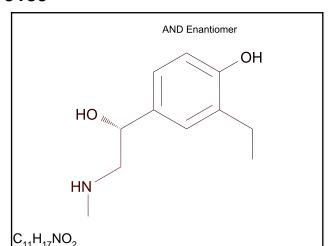
Structural Similar Compounds					
Name	me 5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-		PHENOL; 4;4'- SULFONYLDI-		
Structure	OHCI CI CI OH	HO NH ₂	HO OF SOME		
Actual Endpoint	Severe	Moderate	Moderate		
Predicted Endpoint	Severe	Moderate	Moderate		
Distance	0.622	0.676	0.697		
Reference	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.

Feature Contribution					
	Top fe	atures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set	
SCFP_12	130348166	AND Enantiomer OH OH C C O [*]O[c]1:[cH]:[c](O[*]):[c]([*]):[*]:[c]: 1[*]	0.376	4 out of 4	

SCFP_12	1311101646	AND Enantiomer O OH C C C C C C	0.311	7 out of 8
SCFP_12	-1272709286	AND Enantomer OH [*]CC[c](:[*]):[*]	0.231	24 out of 31
	Top Feat	ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1416196903	AND Enantiomer OH C C C C (*)OC(C)(C)C([*))[*]	-0.383	3 out of 8
SCFP_12	-1980302127	AND Enantiomer OH C: OH (*]=CC(=O)[c](:[*]):[*]	-0.345	1 out of 3
SCFP_12	-2056718782	AND Entantiomer OH C C C [*]C(=[*])[c](:[c]([*])]):[*]:[c]([*]):[*]	-0.163	20 out of 39



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

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 nonnormal 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- TRIMETHYLCYCLOHEXA NOL	METHANE;TRIS(4- AMINOPHENYL)-	
Structure	OH NOH	HO NH 2	NH ₂ NH ₂ NH ₂ NH ₂	
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.

Feature Contribution				
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	696607952	AND Enantiomer OH HO, HN [*]C([*])CNC	0.348	3 out of 3

SCFP_12	2088704928	AND Enantiomer OH HO, HN [*]NCC([*])[*]	0.342	110 out of 128
SCFP_12	Š	AND Enantiomer OH HO, [*]N[*]	0.337	129 Out 01 151
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1848446841	AND Enantiomer HO. HO. OH HO. IT IN THE PROPERTY OF THE PROP	-0.345	1 out of 3
SCFP_12	-350503170	AND Enantiomer OH HO. HO. (*)C[c]1:[cH]:[c](:[c H]:[*]:[c]:1[*])C([*])[*]	-0.0702	9 out of 16
SCFP_12	-496409612	AND Enantiomer HO	0	158 out of 272

AND Enantiomer OH OH Ulling

Molecular Weight: 276.371

ALogP: 2.165 Rotatable Bonds: 4

Acceptors: 3
Donors: 3

 $C_{17}H_{24}O_{3}$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

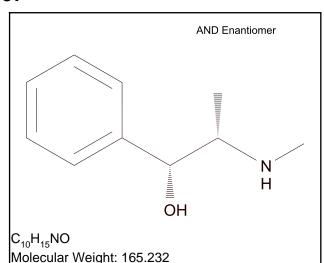
Structural Similar Compounds				
Name	METHANE;TRIS(4- AMINOPHENYL)-	2- BIPHENYLCARBOXYLIC ACID; 2'- HYDROXYMETHYL-	o-Toluidine; N;N-bis(2- hydroxyethyl)-	
Structure	NH ₂ NH ₂ NH ₂	OH OH	HOOH	
Actual Endpoint	Moderate	Moderate	Severe	
Predicted Endpoint	Moderate	Moderate	Moderate	
Distance	0.579	0.696	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.

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set		
SCFP_12	-1640858361	AND Enantiomer OH OH (*)CC[c]1:[cH]:[cH]:[cH]:[cH]:1	0.376	4 out of 4		

SCFP_12	-1272709286	AND Enantiomer OH OH OH OH (*]CC[c](:[*]):[*]	0.231	24 out of 31
SCFP_12	1702664599	AND Enantiomer OH OH OH (*)C((*))(C@@H)(O)C(= [*))(*)	0.212	7 out of 9
		res for negative of		n
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1847692671	AND Enantiomer OH OH ("]C(("))(C@@)(C)(O)C	-0.475	0 out of 1
SCFP_12	1416196903	AND Enantiomer OH OH (*)OC(C)(C)(C((*))(*)	-0.383	3 out of 8
SCFP_12	-1043310069	AND Enantiomer OH OH OH (*)C(C@@H](C=[*])C([*))[*]	-0.246	15 out of 32



Model Prediction

Prediction: Irritant

ALogP: 1.234
Rotatable Bonds: 3

Acceptors: 2 Donors: 2

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

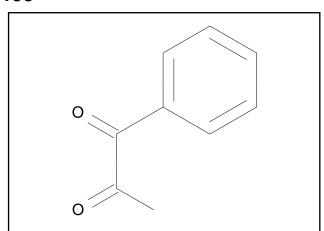
Structural Similar Compounds				
Name	ETHANOL;2- (PHENETHYLAMINO)-	0-TOLUIDINO-ETHANOL	Ethanol; 2-anilino-	
Structure	NH OH	HO NH	NH	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.425	0.436	0.484	
Reference	AMIHBC 10;61;54	JIHTAB 31;60;49	Prehled 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.

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	1186393305	AND Enantiomer OH H [']C(['])[C@H](O)[c](:[']):[']	0.197	13 out of 13		

FCFP_12	-1931573337	AND Enantiomer OH (*)C((*))(C@H)(O)(c)1	0.195	12 out of 12
FCFP_12	-1083860676	:[cH]:[cH]: [cH]:1	0.192	10 out of 10
	Ton Foot	["]C(["])(C@H)(O)(c)† :[cH]:[cH]:[cH]: :[cH]:1		
Fingerprint	Bit/Smiles	tures for negative (Feature Structure	Score	Irritant in training
FCFP_12	-1698724694	AND Enantiomer	-0.0964	set 107 out of 146
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1		
FCFP_12	-453677277	AND Enantiomer	0	264 out of 323
FCFP_12	203677720	AND Enantiomer OH H [*]C([*])[c](:[cH]:[*	0	319 out of 382
]):[cH]:[*]		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

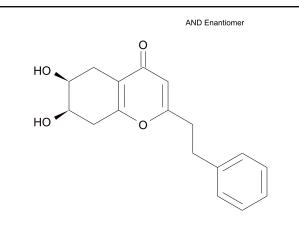
Name	ACETOPHENONE	METHYL BENZOATE	Propiophenone
Structure	0	0	0
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.310	0.380	0.389
Reference	AJOPAA 29;1363;46	AMIHBC 10;61;54	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;291;86

Model Applicability

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

Feature Contribution Top features for positive contribution					
FCFP_12	565968762		0.11	38 out of 42	
		[*]C(=[*])C(=O)C			
	l	<u> </u>		<u> </u>	

FCFP_12	16		0.0586	442 out of 516
FCFP_12	-1549192822	[*][c](:[*]):[*] [*]C(=[*])C(=0)[c](:[*]):[*]	0.0547	12 out of 14
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	[*]C([*])[c]1:[cH]:[c	-0.0964	107 out of 146
		H]:[cH]:[cH]:[cH]:1		
FCFP_12	975909016	[']C(=['])C(=O)[c]1:[cH]:[cH]:[cH]:[-0.0639	6 out of 8
FOED 40		cH]:1	<u></u>	4404 (4007
FCFP_12	0	[*]C([*])[*]	0	1184 out of 1397



C₁₇H₁₈O₄

Molecular Weight: 286.322

ALogP: 1.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

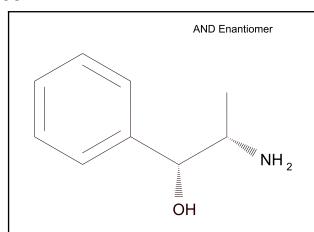
Structural Similar Compounds				
Name	2- BIPHENYLCARBOXYLIC ACID; 2'- HYDROXYMETHYL-	2;2'-Dihydroxy-4- methoxybenzophenone	BENZOIN; OXIME	
Structure	OH OH	OH O HO NO	N OH	
Actual Endpoint	Irritant	Non-Irritant	Irritant	
Predicted Endpoint	Irritant	Non-Irritant	Irritant	
Distance	0.585	0.585	0.602	
Reference	IHFCAY 6;1;67	J. Am. Coll. Toxicol. 2(5):35;1983	28ZPAK-;111;72	

Model Applicability

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

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	436915834	AND Enantiomer HO HO [*]C\C(=C\[*])\O[*]	0.167	4 out of 4		

FCFP_12	565968762	AND Enantiomer HO HO	0.11	38 out of 42
		[*]C(=[*])C(=O)C		
FCFP_12	1388176727	AND Enantiomer HO O O O O O O O O O O O O O O O O O O	0.107	19 out of 21
		[*]C(=[*])CC(e]1:[cH] :[cH]:[cH]:[cH]:[cH] :1		
	Top Feat	ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1981711554	AND Enantiemer HO HO	-0.103	5 out of 7
		[*]CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1		
FCFP_12	-1698724694	AND Enantiomer HO HO	-0.0964	107 out of 146
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1		
FCFP_12	991735244	AND Ensentiomer HO HO [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	0	237 out of 291



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

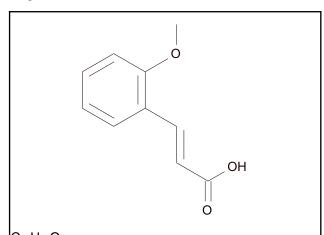
Structural Similar Compounds				
Name	3-AMINO-2-METHYL BENZYL ALCOHOL	1;2-DIMETHYL;1;3- BUTANEDIOL	METHYLAMINE;M- PHENYLENEBIS-	
Structure	H ₂ N ⁿ OH	НО	NH 2 NH 2	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.339	0.453	0.455	
Reference	AMIHBC 10;61;54	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.

Feature Co	Feature Contribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	1186393305	AND Enantiomer NH ₂ OH [']C(['])[C@H](O)[c](:[']).[']	0.197	13 out of 13		

FCFP_12	-1931573337	AND Enantiomer NH ₂ OH	0.195	12 out of 12
FCFP_12	-1083860676	[*]C([*))[C@H](O)[c]1 :[cH]:[cH]: [:[cH]: [cH]:1 AND Enantiomer OH NH ₂	0.192	10 out of 10
	Ton Foot	িবিশ্যেকের্ড্রিনা ্লেন্যুক্রন্ত্র্যাত্ত্ব ব্যক্তির বিদ্যালয়	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	AND Enantiomer ., NH 2 OH [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	991735244	AND Enantiomer ., NH 2 OH [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	0	237 out of 291
FCFP_12	-453677277	AND Enantiomer AND Enantiomer NH 2 OH [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0	264 out of 323



C₁₀H₁₀O₃

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

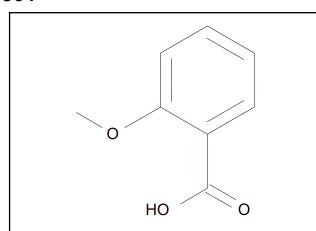
Structural Similar Compounds				
Name	Ethyl Paraben	MANDELIC ACID; ETHYL ESTER	MANDELIC ACID; ISOPROPYL ESTER	
Structure	НО	OH OH	OH	
Actual Endpoint	Non-Irritant	Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Irritant	Irritant	
Distance	0.458	0.467	0.467	
Reference	J. Am. Coll. Toxicol. 3(5):147;1984	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.

Top feat Bit/Smiles	tures for positive c	ontribution Score	
Bit/Smiles	Feature Structure	Score	luuit a sat isa tu aisaisa sa
		DCOIE	Irritant in training set
2107131107	[*]\C=C\C(=O)O	0.18	6 out of 6
	2107131107	ОН	ОН

FCFP_12	-548632217	[*]C(=[*])O	0.177	59 out of 61
FCFP_12	-1099193755	[*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:1OC	0.175	5 out of 5
	Top Feat	ures 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]:[c H]:[cH]:[cH]:1	-0.0964	107 out of 146



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

bayesian Score. 1.56

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

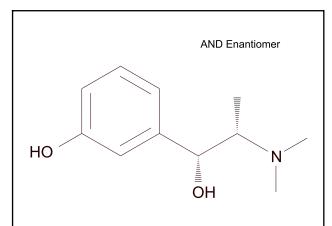
Structural Similar Compounds				
Name	SALICYCLIC ACID; METHYL ESTER	BENZOIC ACID	RESORCINOL; MONOACETATE	
Structure	OH O	OH	HONNO	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.355	0.365	0.385	
Reference	28ZPAK-;106;72	BIOFX* 28-4/73	JAPMA8 46;185;57	

Model Applicability

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

Feature Co	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 Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	136627117	HO O	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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

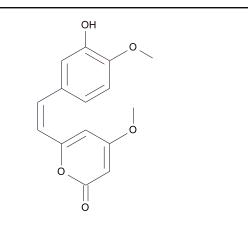
Structural Similar Compounds						
Name	o-Toluidine; N;N-bis(2- hydroxyethyl)-	ETHANOL; 2;2'- CYCLOHEXYLIMINODI-	BENZOIN; OXIME			
Structure	HO OH	OH OH	OH N-OH			
Actual Endpoint	Irritant	Irritant	Irritant			
Predicted Endpoint	Irritant	Irritant	Irritant			
Distance	0.476	0.505	0.512			
Reference	Prehled 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.

Feature Contribution						
	Top fe	atures for positive c	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	136388789	AND Enantiomer HO OH [*]N([*])C	0.206	30 out of 30		

FCFP_12	1186393305	AND Enantiomer HO OH	0.197	13 out of 13
FCFP_12	-1931573337	[*]C([*])[C@H](O)[c](:[*]);[*] AND Enantiomer HO OH [*]C([*])[C@H](O)[c]1	0.195	12 out of 12
		:[cH]:[cH]:[*]:[cH]: [cH]:1		
Fingerprint	Bit/Smiles	ures for negative of Feature Structure	Score	Irritant in training
				set
FCFP_12	991735244	AND Enantiomer : HO OH [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	0	237 out of 291
FCFP_12	-453677277	AND Enantiomer HO OH [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0	264 out of 323
FCFP_12	0	AND Enantiomer HO OH [*]C([*])[*]	0	1184 out of 1397



C₁₅H₁₄O₅

Molecular Weight: 274.269

ALogP: 1.969 Rotatable Bonds: 4

Acceptors: 5
Donors: 1

	•		
Name	2;8;9-TRIOXA-5-AZA-1- SILABICYCLO(3.3.3)UNDE CANE;1-(3- AMINOPROPYL)-	2-Hydroxy-4- methoxybenzophenone	2;5-DICHLORO-4(3'- METHYL-5' PYRAZOLON- 1'-YL)BENZENE SULFONIC ACID
Structure	NH 2	HO NATURAL OF THE OF TH	HO SM
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 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.

Model Applicability

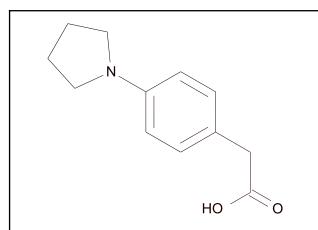
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

i catale 00	iid ibadioii							
	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	OH O (*)=CC1=C(*)=CC(=0)01	0.137	2 out of 2
FCFP_12	7	OH 0\	0.119	142 out of 156
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1078052987	[*]\C=C/[c]1:[cH]:[cH]:[:]-[:]:[c][(*]):[cH]:	-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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

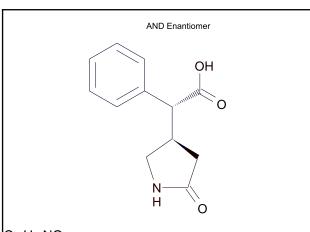
Structural Similar Compounds						
Name	Propionitrile; 3-(N-(2- hydroxyethyl)anilino)-	1-NAPHTHALENEACETIC ACID	MANDELIC ACID; ISOPROPYL ESTER			
Structure	OH OH	OH OH	OH OH			
Actual Endpoint	Irritant	Irritant	Irritant			
Predicted Endpoint	Irritant	Irritant	Irritant			
Distance	0.482	0.487	0.501			
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.

Feature Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	-548632217	HO O	0.177	59 out of 61		

FCFP_12	-1176841573	[*]=CC(=O)O	0.164	43 out of 45
FCFP_12	-822674211		0.137	2 out of 2
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	202105689	* c 1: cH : c 1 : c (: cH : cH : c 1) N2CCCC 2	-0.103	5 out of 7
FCFP_12	-98332825	HO O	-0.0663	105 out of 139
FCFP_12	203677720	[*]C([*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382



 $C_{12}H_{13}NO_3$

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

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 nonnormal 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	O NH NH 2	NH OH	O OH		
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.

Feature Contribution							
	Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set			
FCFP_12	-885550502	AND Enantiomer OH N N OF N N OF N N OF N N N N N N N N N N N N N	0.18	64 out of 66			

FCFP_12	-548632217	AND Enantiomer OH NH O [*]C(=[*])O	0.177	59 out of 61
FCFP_12	-416918913	AND Enantiomer OH OH (*)C((*))(C@@H)1CNC(= [*)C1	0.175	5 out of 5
		res for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1716639150	AND Enantiomer OH OH N O (*)C((*))(C@@H)(C(=[*))(*))(-)1;(cH);(cH) :[cH]:(cH);(cH);(cH);(cH)	-0.537	1 out of 3
FCFP_12	-1133295320	AND Enantiomer OH OH (*)C([*))C@@H)(C(=[*))(*))(=)1:[c+1]:[c+1	-0.537	1 out of 3
FCFP_12	-1698724694	AND Enantiomer OH N OH (*]C([*])[c]1:[cH]:[c H]:[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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

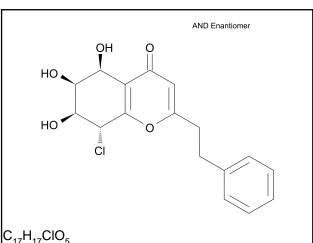
Structural Similar Compounds					
Name	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	ANTHRAQUINONE; 1-((2- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE		
Structure	OH NH NH	OH HN TH N N H	O OH OH		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Irritant	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.

Feature Contribution						
	Top fea	tures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	-1601875224	HO OH OH OH	0.184	7 out of 7		

FCFP_12	-1099193755	AND Enantiomer HO OH OH [*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:10C	0.175	5 out of 5
FCFP_12	436915834	AND Enantiomer HO OH OH [*]C\C(=C\[*])\O[*]	0.167	4 out of 4
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1244036906	AND Enantiomer HO OH OH [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*]	-0.592	0 out of 1
FCFP_12	-415216134	AND Enantiomer HO OH OH [*]OC(C)(C)C([*])[*]	-0.0889	11 out of 15
FCFP_12	0	AND Enantiomer HO O OH OH [*]C([*])[*]	0	1184 out of 1397



Molecular Weight: 336.767

ALogP: 1.368
Rotatable Bonds: 3

Acceptors: 5
Donors: 3

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

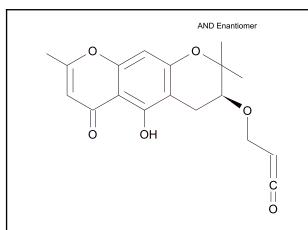
Structural Similar Compounds						
Name	ANTHRAQUINONE; 1-((2- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	HYDROXYETHYL)AMINO) (bis(2-				
Structure	OH HN TAN H	O HOOH	HO NH 2			
Actual Endpoint	Irritant	Irritant	Non-Irritant			
Predicted Endpoint	Irritant	Irritant	Non-Irritant			
Distance	0.537	0.654	0.657			
Reference	28ZPAK 245;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;645;86	J. Am. Coll. Toxicol. 5(3):205;1986			

Model Applicability

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

Feature Contribution						
Top features for positive contribution						
Bit/Smiles	Feature Structure	Score	Irritant in training set			
-879019572	AND Enantiomer HO	0.199	15 out of 15			
	Top fea	Bit/Smiles Feature Structure -879019572	Top features for positive contribution Bit/Smiles Feature Structure Score -879019572 O.199			

FCFP_12	436915834	AND Enantiomer HO CI	0.167	4 out of 4
FCFP_12	32	[*]C\C(=C\[*])\O[*] AND Enantiomer HO CI	0.15	202 out of 215
	Ton Feat	[*]CI tures for negative of	Contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1981711554	AND Enantiomer HO OH O	-0.103	5 out of 7
FCFP_12	-1698724694	AND Enantiomer HO CI (*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	451847724	AND Enantiomer HO OH O	0	161 out of 192



C₁₈H₁₈O₆

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 nonnormal 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	OH O	HO start to NH 2	O HN Att NH 2			
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			

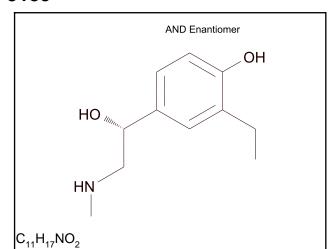
Model Applicability

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						
FCFP_12	-1601875224	[*]C[c]1:[c]([*]):[*] :[c]([*]):[c](C(=[*])[*]):[c]:10	0.184	7 out of 7		

FCFP_12	-1099193755	AND Enanticmer OH C C (*)=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:10C	0.175	5 out of 5
FCFP_12	436915834	AND Enentlomer O OH C O C O (*]C\C(=C\[*])\O[*]	0.167	4 out of 4
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1244036906	AND Enantomer O OH O OH C O [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*	-0.592	0 out of 1
FCFP_12	353886735	AND Enantiomer OH [*]C([*])OCC=C=[*]	-0.141	16 out of 23
FCFP_12	-415216134	AND Enanthomer OH C C O' OH C O' O' O' O' O' O' O' O' O'	-0.0889	11 out of 15



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

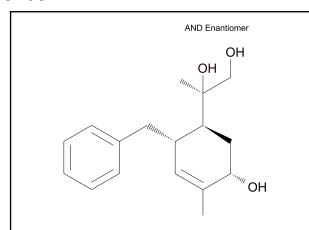
Structural Similar Compounds					
Name	2;3-DIMETHYL;1;2;6- HEXANETRIOL p-Phenylenediamine; N;N- bis(2-hydroxyethyl)-; sulfate (1:1)		1-AMINO-3- AMINOMETHYL-3;5;5- TRIMETHYLCYCLOHEXA NOL		
Structure	но ОН	HO NH 2	HO NH 2 NH 2		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Irritant	Irritant	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.

Feature Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	1186393305	AND Enantiomer HO HN ["]C(["])[C@H](O)[c](:["]:["]	0.197	13 out of 13		

FCFP_12	-1931573337	AND Enantiomer HO HN ["]C([")](C@H)(O)(c]1 :(cH):[cH]:[cH]: [cH]:	0.195	12 out of 12
FCFP_12	-885550502	AND Enantiomer OH HO,	0.18	64 out of 66
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	0	AND Enantiomer OH HO, HN [*]C([*])[*]	0	1184 out of 1397
FCFP_12	203677720	AND Enantiomer HO HN [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382
FCFP_12	136597326	AND Enantiomer OH HO,	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

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

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 nonnormal 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	NH ₂ NH ₂ NH ₂	OH HN 11 N H OH OH OH OH OH OH OH OH OH	OH NOH			
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.

Feature Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	-415156552	AND Enantiomer OH OH (*)C[C@)(C)(O)C((*))[1	0.184	7 out of 7		

FCFP_12	1070061035	AND Enantiomer OH OH OH [*]C([*])O	0.107	338 out of 376
FCFP_12	3	AND Enantiomer OH OH OH (*)OH	0.105	491 out of 547
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-59531427	AND Enantiomer OH OH OH (P)(C@H)1(P)=C((*))(C @@H)(O)C(C@@H)1C(P))((*))(*)	-0.156	4 out of 6
FCFP_12	1981711554	AND Enantiomer OH OH (*]CC[c]1:[cH]:[cH]:[cH]:[cH]:1	-0.103	5 out of 7
FCFP_12	-1698724694	AND Enantiomer OH OH OH (*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146

Non-Irritant

AND Enantiomer
O NH N S N N N N N N N N N N N N N N N N
Î

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Anthraquinone; 1;1'-	Anthraquinone; 1;1'-	ANILINE;N;N'-1;4-	
	(anthraquinon-1;5-	(anthraquinon-1;4-	ANTHRAQUINONYLENEB	
	ylenediimino)di-	ylenediimino)di-	S(4-PHENOXY-	
Structure	The state of the s	HI THE STATE OF TH		

Irritant

Predicted Endpoint Irritant Irritant Non-Irritant 0.837 0.838 1.053 Distance Prehled Prumyslove 28ZPAK-;114;72 Reference Prehled Prumyslove Toxikologie; Órganicke Toxikologie: Organicke Latky; Marhold; J. pp Latky; Marhold; J. pp 736:86 736:86

Model Applicability

Actual Endpoint

Structural Similar Compounds

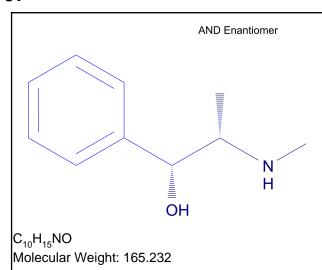
Irritant

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.

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Irritant in training set FCFP_12 1747237384 AND Enantiomer 0.208 44 out of 44

FCFP_12	178336375	AND Enantiomer	0.202	19 out of 19
		[*][c](:[*]):[c](:n:[*]):[c](:[*]):[*]		
FCFP_12	-124655670	AND Enantiomer	0.2	16 out of 16
		[*][c]1:[*]:[*]:s:[cH]:1		
	Top Feat	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1812846456	AND Enantiomer	-1.01	2 out of 9
		[*]CCCC\C=C/C1[*][*]1		
FCFP_12	690511177	AND Enantiomer	-0.268	1 out of 2
		[*]:[cH]:[c](:n:[*])[c](:[*]):[*]		
FCFP_12	1175638033	AND Enantiomer	-0.133	207 out of 293
		√ No.		
		[*]CCCCC=[*]		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

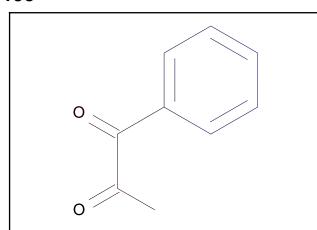
Name	Ephedrine	Tocainide	Amphetamine
Structure	OH N H	NH 2	NH ₂
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.000	0.554	0.573
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre 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.

Feature Contribution								
	Top features for positive contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set				
ECFP_12	734603939	AND Enantiomer OH [*]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 OH [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:1	-1.68	0 out of 14
ECFP_12	2014710090	AND Enantiomer OH H [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	-1.56	0 out of 12
ECFP_12	-221133010	AND Enantiomer OH ("C(")")C@H)(O)[e]1 :[eH]:[eH]:[eH]:	-0.661	0 out of 3



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

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Coumarin	Amphetamine	Phenacetin	
Structure		NH ₂		
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(=0)[c]1:[cH]:[cH]:[*]:[cH]:[c H]: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 Featur	es for negative o	ontribution	
Fingerprint		Feature Structure		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]:[*]	-0.296	36 out of 156

AND Enantiomer HO HO O

C₁₇H₁₈O₄ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Tolazamide	Naltrexone	Doxefazepam
Structure	N H N N N N N N N N N N N N N N N N N N	HO same	OH N O
		ОН	CI CI

Non-Carcinogen

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.611

Carcinogen

Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.616

Model Applicability

Actual Endpoint

Distance

Reference

Predicted Endpoint

Structural Similar Compounds

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.

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

- 2. Unknown ECFP_2 feature: 1794461805: [*]CC(=C([*])[*])C(=[*])[*]
- 3. Unknown ECFP_2 feature: 1650944136: [*]CC(=C([*])[*])O[*]

Non-Carcinogen

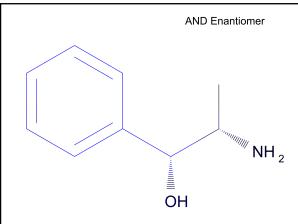
Non-Carcinogen

Res.) Sept. 1997

0.566

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	53207596	AND Enantiomer HO HO [*]C([*])CC(=[*])[*]	0.459	8 out of 15	

ECFP_12	683445015	AND Enantiomer HO HO [*]O[*]	0.294	28 out of 66
ECFP_12	-1650219925	AND Enantiomer HO HO (*)C[c]1:[cH]:[cH]:[c H]:[cH]:[cH]:1	0.208	6 out of 15
	Top Feat	ures for negative of	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-281505363	AND Enantiomer HO	-0.56	11 out of 64
ECFP_12	1571214559	AND Enantiomer HO HO ** 11:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	1997021792	AND Enantiomer HO HO (*):[cH]:[cH]:[t*]	-0.296	36 out of 156



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	OH N H	NH 2		
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

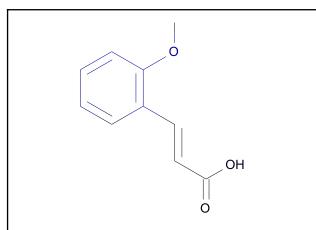
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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	AND Enantiomer NH ₂ OH	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 NH 2	-1.68	0 out of 14
		[*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1		
ECFP_12	2014710090	AND Enantiomer NH 2	-1.56	0 out of 12
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1		
ECFP_12	-221133010	AND Enantiomer NH2	-0.661	0 out of 3
		[*]C([*])[C@H](O)[c]1 :[cH]:[cH]:[*]: [cH]:1		



C₁₀H₁₀O₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Aspirin	Naproxen	Phenacetin
Structure			
			H
	0	О ОН	
			Ĭ
	OOH		

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

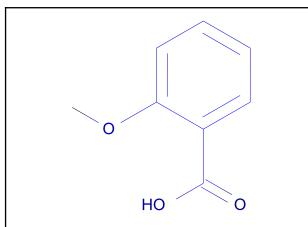
Structural Similar Compounds

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:[*]

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 Feat	ures for negative of	ontributio	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1634699529	[*]O[c]1:[cH]:[cH]:[c H]:[*]:[o]:1[*]	-1.25	0 out of 8
ECFP_12	-1961666573	[*]O[c]1:[cH]:[cH]:[c H]:[cH]:[c]:1[*]	-0.661	0 out of 3
ECFP_12	1408898974	[*]O[c](:[cH]:[*]):[c]([*]):[*]	-0.517	5 out of 29



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

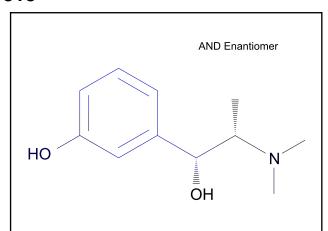
Name	Aspirin	Phenacetin	Ethionamide
Structure	O OH	H N O	S NH 2
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.351	0.601	0.606
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre 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.

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1074141656	HO (*)=0	0.103	86 out of 248

ECFP_12	734603939		0.0966	92 out of 267
		но		
		[*]C		
	Top Featur	es for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure		Carcinogen in training set
ECFP_12	1337040050	HOO	-1.84	0 out of 17
		[*]C(=[*])[c](:[cH]:[*]):[c]([*]):[*]		
ECFP_12	1634699529	но	-1.25	0 out of 8
		[*]O[c]1:[cH]:[cH]:[c H]:[*]:[c]:1[*]		
ECFP_12	989674687	HOO	-1.16	0 out of 7
		[*][c](:[*]):[c](:[cH]:[*])C(=O)O		



 $C_{11}H_{17}NO_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Ephedrine	Phenylephrine	Tocainide
Structure	N H	HO	NH NH NH 2
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

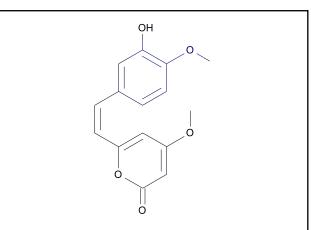
Model Applicability

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 Co	Feature Contribution				
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	-1235570483	AND Enantiomer HO OH	0.421	1 out of 1	
		[*]C([*]);C@H](O)[c]1 :[cH]:[cH]:[*]:[c]([*]):[cH]:1			

ECFP_12	-177786161	AND Enantiomer OH [*]:[cH]:[c](O):[cH]: [*]	0.341	7 out of 15
ECFP_12	734603939	AND Enantiomer HO OH [*]C	0.0966	92 out of 267
		tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-388186450	AND Enantiomer : HO OH [*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-1.68	0 out of 14
ECFP_12	1064495017	HO OH I*]C([*])N(C)C	-0.661	0 out of 3
ECFP_12	-783815036	AND Enantiomer .: OH 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar	Compounds
--------------------	-----------

Name	Cytembena	Scopolamine	Atropine	
Structure	OH Br	ON NO N	N OH	
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	

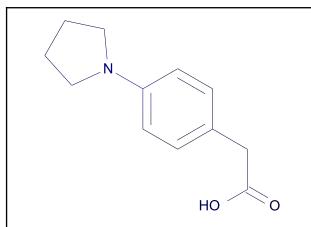
Model Applicability

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=[*]

Top features for positive contribution						
ECFP_12	-1925046727	OH O\ O	0.407	16 out of 33		

683445015	OH O	0.294	28 out of 66
	; [*]O[*]		
2106656448	OH O	0.141	30 out of 83
	[*]C(=O)[*]		
Top Fea	tures for negative	contribution	1
Bit/Smiles	Feature Structure	Score	Carcinogen in training set
2077607946	OH O	-1.25	0 out of 8
	[*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]		
1408898974	OH O	-0.517	5 out of 29
	[*]O[c](:[cH]:[*]):[c]([*]):[*]		
1680623188	[*Hol(:Filvid/OC):	-0.295	3 out of 14
	2106656448 Top Fea Bit/Smiles 2077607946 1408898974	2106656448 Comparison of the comparison of t	2106656448



 $C_{12}H_{15}NO_{2}$

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Naproxen	Methylphenidate	Phenacetin		
Structure	OH NO	N H O O	o H		
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.

Feature Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	1737023319	Ho o Ho o Ho o o	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 Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1251172932	HO O	-0.661	0 out of 3
ECFP_12	1731135544	[*]CC(=O)O	-0.583	2 out of 14
ECFP_12	1951894094	C	-0.505	3 out of 18

[*]:[c](:[*])N1C[*][*]C1

AND Enantiomer OH N H O

C₁₂H₁₃NO₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds	;
-------------------------------------	---

Name	Phenobarbital	Aminoglutethimide	Baclofen
Structure	HN	NH ₂	CI OH OH
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

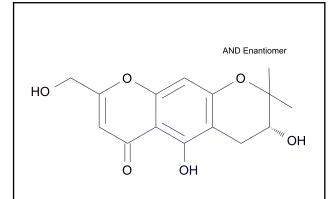
Model Applicability

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

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	53207596	AND Enantiomer OH N N (*]C([*])CC(=[*])[*]	0.459	8 out of 15	

ECFP_12	-742538367 -1905455774	O=C1C[*][*]N1 AND Enantiomer	0.445	3 out of 5 4 out of 8
		[*]C([*])C(=O)O		
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-388186450	AND Enantiomer OH N (*) (*) (*) (*) (cH): (-1.68	0 out of 14
ECFP_12	2014710090	AND Enantitomer OH OH N N (*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	-1.56	0 out of 12
ECFP_12	1571214559	AND Enantiomer OH N N (*]1:[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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Prednisolone	Hydrocortisone	Piroxicam
Structure	OH OH OH	OH OH	OH HN, N
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.

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	53207596	AND Enantiomer HO OH OH [*]C([*])CC(=[*])[*]	0.459	8 out of 15		

ECFP_12	-200406221	AND Enantiomer HO OH OH [*]C([*])([*])O[c](:[*]):[*]	0.33	3 out of 6
ECFP_12	683445015	AND Enantiomer HO OH OH [*]O[*]	0.294	28 out of 66
	Top Feat	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1660913849	AND Enantiomer HO OH OH [*][c](:[*]):[c](O):[-0.941	0 out of 5
ECFP_12	-1051556861	AND Enantiomer HO OH OH [*]C[C@@H](O)C([*])([*])[*]	-0.661	0 out of 3
ECFP_12	1875238785	AND Enantomer HO OH OH [*]C\C(=C\[*])\O[*]	-0.272	0 out of 1

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Eval.& Res./Off. Testing &

Res.) Sept. 1997

AND Enantiomer HO HO CI

 $C_{17}H_{17}CIO_5$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Prednisolone	Hydrocortisone	Beclomethasone		
Structure	OH OH OH	OH OH	HO et al. 12 to the total of th		
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	US FDA (Centre for Drug	US FDA (Centre for Drug		

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.

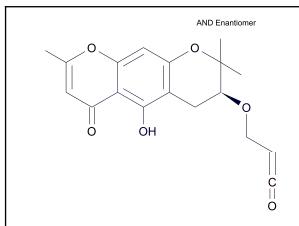
Eval.& Res./Off. Testing &

Res.) Sept. 1997

- 2. Unknown ECFP_2 feature: -1567907747: [*]C([*])[C@@H](O)C(=[*])[*]
- 3. Unknown ECFP_2 feature: -7106223: [*]C([*])[C@@H](CI)C(=[*])[*]
- 4. Unknown ECFP_2 feature: 1652274794: [*]OC(=C([*])[*])C([*])[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	105634199	AND Enantiomer HO CI [*]C([*])CI	0.421	1 out of 1	

ECFP_12	683445015	AND Enantiomer HO CI [*]O[*]	0.294	28 out of 66
ECFP_12	-1650219925	AND Enantiomer HO CI [*]C[c]1:[cH]:[cH]:[c H]:[cH]:[cH]:1	0.208	6 out of 15
	Top Feat	ures for negative of	ontributio	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1571214559	AND Enantomer HO CI [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	-281505363	AND Enantiomer HO CI [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	-817402818	AND Enantiomer HO CI [*]CI	-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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Naltrexone	Tolazamide	Doxefazepam
Structure			
	HO MAN TO THE ONLY OF THE ONLY	H N H N N N N N N N N N N N N N N N N N	OH OH
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.

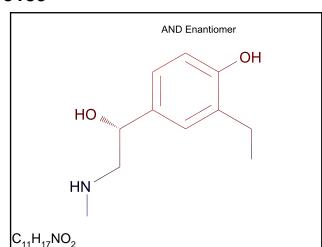
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -331149802: [*]O\C(=C/[*])\C
- 3. Unknown ECFP_2 feature: -1591590376: [*]C=C=O
- 4. Unknown ECFP_2 feature: 2106995136: [*]=C=O
 5. Unknown ECFP_2 feature: -91536905: [*]CC=C=[*]
- 6. Unknown ECFP 2 feature: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

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	

ECFP_12 53207596	[*]=C=[*]		
	AND Enantiomer	0.459	8 out of 15
ECFP_12 -1925046727	AND Enantiomer OH C C C C C C C	0.407	16 out of 33

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1660913849	AND Enantiomer OH [*][c](:[*]):[c](O):[c]([*]):[*]	-0.941	0 out of 5
ECFP_12	-1250019913	AND Enantiomer OH C C C C C C	-0.485	0 out of 2

ECFP_12	-2124995946	AND Enantiomer OH OH	-0.272	0 out of 1
		(1,1)(1,1) (1,1)(1,1) (1,1)(1,1) (1,1)(1,1)		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Phenylephrine	Albuterol	Procarbazine
Structure	HO	HO TANDOH	N H NN NH
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.

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	-1411846807	AND Enantiomer OH HO [*]C[c]1:[cH]:[c](:[c H]:[cH]:[c]:1[*])[C@ @H](O)CN[*]	0.613	2 out of 2	

ECFP_12	-283284905	AND Enantiomer OH HO HN [*]C[c]1:[cH]:[c](:[c H]:[*]:[c]:1[*])C([*])[*]	0.613	2 out of 2
ECFP_12	-188314265	AND Enantiomer HO. HO. (*)C[c]1:[cH]:[c](:[c H]:[cH]:[c]:10)C([*))[*]	0.613	2 out of 2
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	767488533	AND Enantiomer OH HO. [*]:[c](:[*])CC	-0.941	0 out of 5
ECFP_12	493154328	AND Enantiomer OH HO,	-0.661	0 out of 3
ECFP_12	-176846085	AND Enantiomer HO HN [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.374	7 out of 34

AND Enantiomer
он
OH mm,
Minning OH

C₁₇H₂₄O₃ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Pindolol	Carteolol	Procarbazine
Structure	HN OH H	HN OH H	N H N NH
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	AND Enantiomer OH OH [*]C([*])CC([*])[*]	0.339	14 out of 31
ECFP_12	-1650219925	AND Enantiomer OH OH OH (*]C[c]1:[cH]:[cH]:[c H]:[cH]:[cH]:1	0.208	6 out of 15
ECFP_12	1095683433	AND Enantiomer OH OH OH (*]C[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1).154	6 out of 16

ingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1571214559	AND Enantiomer OH OH OH OH OH OH OH OH OH O	-0.56	11 out of 64
ECFP_12	-281505363	AND Enantiomer OH OH OH (*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64

ECFP_12	1997021792	AND Enertiomer OH OH OH OH	-0.296	36 out of 156
]:[cH]:[cH]:[cH]:[

Simeprevir

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

AND Enantiomer
O NH N S
O N
o no
N \

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds Name Reserpine Deserpidine Bromocriptine						
Structure		Orthodor Market	Br N N N N N N N N N N N N N N N N N N N			
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.

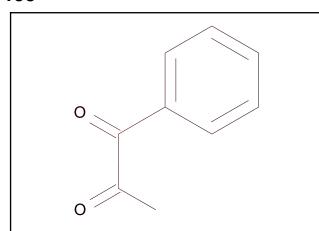
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -309676563: [*]N[C@]1(C[C@H]1[*])C(=[*])[*]
- 3. Unknown ECFP_2 feature: -1049290660: [*]C1([*])C[C@H]1C=[*]
- 4. Unknown ECFP_2 feature: 1616402542: [*]CN(C)C(=[*])[*]
- 5. Unknown ECFP_2 feature: -1818486371: [*]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

Top features for positive contribution						
Fingerprint	Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set					

ECFP_12	-1331088410	AND Enantiomer AND Enantiomer NS NS N (*]CCC=[*]	0.725	3 out of 3
ECFP_12	-2095963820	AND Enantiomer AND Enantiomer NS N I*I[C@@H]1[*][*]C[C@H]1C(=[*])[*]	0.722	12 out of 17
ECFP_12	914325265	AND Enantiomer AND Enantiomer NS N N N N N N N N N N N N N N N N N	0.516	8 out of 14

Top Features for negative contribution					
Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
2077607946	AND Enantiomer AND Enantiomer N S N O O O O O O O O O O O O O O O O O O	-1.25	0 out of 8		
-867777309	AND Enantiomer AND Enantiomer NS NS N (*) (*) (*) (*) (*) (*) (*)	-0.661	0 out of 3		
	Bit/Smiles 2077607946	Bit/Smiles Feature Structure 2077607946 AND Enantiomer [*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*] -867777309 AND Enantiomer	Bit/Smiles Feature Structure Score 2077607946 -1.25 [*]O[c]1:[cH]:[cH]:[c]:[1[*] -0.661		

ECFP_12	1408898974	AND Enantiomer	-0.517	5 out of 29
		Ness Ness		
		A SUBJECT OF N		
		~ 'v~~ o ~ o		
		[*]O[c](:[cH]:[*]):[c		
]([*]):[*]		



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

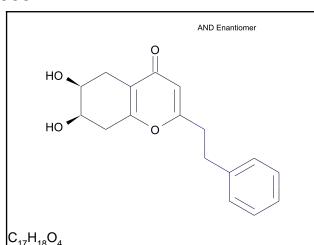
Name	Phenacetin	Methoxsalen; 8-	Nicotine
Structure	O THE STATE OF THE		N N
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.589	0.596	0.621
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & 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.

Feature Contribution Top features for positive contribution						
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]:[c	0.419	1 out of 1
SCFP_4	571795252	[*]C(=[*])C(=O)C	0.283	2 out of 4
	Top Feat	tures 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



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sim	Phenobarbital	Doxefazepam	
Structure	Aminoglutethimide NH 2	O HN O HN O	OH OOH
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

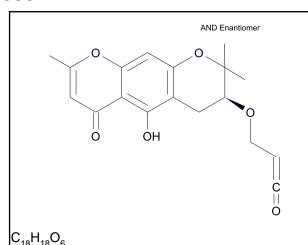
Model Applicability

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.

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
SCFP_4	-1971196727	AND Enantiomer HO	0.295	5 out of 11	

SCFP_4	0	AND Enantiomer HO (*):[cH]:[*] AND Enantiomer HO HO O HO O O O O O O O O	0.199	36 out of 93 41 out of 111
		[*]C	4 11 41	
F'		ures for negative of Feature Structure) haratetata
Fingerprint	Bit/Smiles	reature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantiomer HO HO O [*10CC[o](·[*1])·[*]	-1.16	1 out of 17
SCFP_4	-1211866396	[*]CC[c](:[*]):[*] AND Enantiomer HO HO () (*)CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.06	0 out of 6
SCFP_4	-1980361709	AND Enantiomer HO HO ["]C(=["])C(=0)C=["]	-0.489	0 out of 2



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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

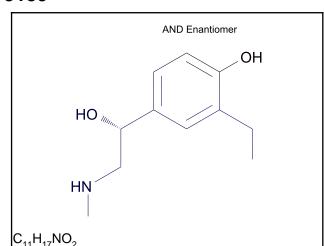
Structural Similar Compounds						
Name	Doxefazepam	Doxefazepam Phenolphthalein				
Structure	OH OH	НО	OH Br			
Actual Endpoint	Actual Endpoint Multiple-Carcinogen		Multiple-Carcinogen			
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen			
Distance	0.582	0.628	0.630			
Reference	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.

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
SCFP_4	-1971196727	AND Enantiomer OH [*]C(=CC(=[*])[*])[*]	0.295	5 out of 11	

SCFP_4	-1980302127	AND Enuntioner OH C O [*]C(=[*])C(=O)[c](:[*]):[*]	0.204	1 out of 2
SCFP_4	-424425761	AND Enantiomer OH [*]:[c](:[*])O	0.201	6 out of 15
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantomer OH C C C C C C C C C C C C C	-1.16	1 out of 17
SCFP_4	1238198777	AND Enantiomer OH C C [*][c]1:[*]:[c]([*]): [c]2C[*]C([*])([*])O [G]:2:[cH]:1	-0.489	0 out of 2
SCFP_4	616547045	AND Enantiomer OH C C C (*]C\C(=C\[*])\O[*]	-0.489	0 out of 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

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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Albuterol	Terbutaline	Procarbazine
Structure	HO TO OH	HO NOH	NH NH
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 &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &

Model Applicability

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

Res.) Sept. 1997

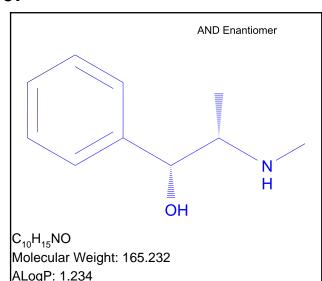
Res.) Sept. 1997

All properties and OPS components are within expected ranges.

Res.) Sept. 1997

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1374800107	AND Enantiomer HO	0.288	10 out of 23

SCFP_4	-424425761	AND Enantiomer OH HO, HN [*]:[c](:[*])O	0.201	6 out of 15
SCFP_4	3	AND Enantiomer OH HO,	0.199	36 out of 93
	Top Feat	ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantiomer OH HO	-1.16	1 out of 17
SCFP_4	-1931277081	AND Enantiomer HO . HO . OH HN [*]C[c]1:[cH]:[*]:[cH]:[cH]:[c]:10	-0.489	0 out of 2
SCFP_4	-350503170	AND Enantiomer OH HO. HO. [*]C[c]1:[cH]:[c](:[c H]:[*]:[c]:1[*])C([*])[*]	-0.489	0 out of 2



Model Prediction

Rotatable Bonds: 3

Acceptors: 2 Donors: 2

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

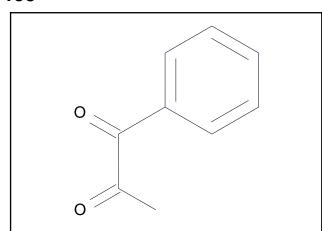
Structural SIM Name	ilar Compounds Ephedrine	Amphetamine	Pronetalol
Structure	OH N H	NH ₂	HN
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.000	0.515	0.522
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & 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.

Feature Contribution				
Top Fea	ntures for negative	contribution		
Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
1702724181	AND Enantiomer OH H [*]C([*])[C@H][(0)[c](:[*]]:[*]	-1.22	1 out of 18	
	Top Fea	Bit/Smiles Feature Structure 1702724181 AND Enantiomer OH (*ICI*) C@H O E	Top Features for negative contribution Bit/Smiles Feature Structure Score 1702724181 -1.22	

SCFP_6	-561151481	AND Enantiomer	-1.07	1 out of 15
SCED 6	1240542260	[]C([])[C@H](O)[c]1	0.057	O out of F
SCFP_6	1318513260	AND Enantiomer OH H	-0.957	0 out of 5
		[']N[C@@H](C)[C@H](O) [6](:[eH];'[']:[eH]: [1]		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Coumarin	Antipyrine	Phenacetin
Structure		N N N N N N N N N N N N N N N N N N N	O N N O O O O O O O O O O O O O O O O O
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.

Feature Contribution					
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	-541146916	[*]C(=[*])C(=0)[e]1:[cH]:[eH]:[*]:[eH]:[e H]:1	0.415	1 out of 1	

SCFP_6	136627117		0.167	18 out of 47
		[*]C(=[*])C		
SCFP_6	13		0.0717	90 out of 261
		0		
		[*]=O		
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1653911926		-0.504	12 out of 64
		[*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1		
SCFP_6	-2056718782	0 [*]C(=[*])[c](:[cH]:[-0.219	23 out of 90
]):[cH]:[]		
SCFP_6	387787917	[*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.164	14 out of 52

AND Enantiomer HO HO O

C₁₇H₁₈O₄ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Tolazamide	Naltrexone	Phenobarbital
Structure	H N H N N N N N N N N N N N N N N N N N	HO MAN ON THE PART OF THE PART	HN O
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

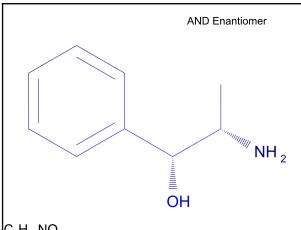
Model Applicability

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 fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971196727	AND Enantiomer HO HO [*]C(=CC(=[*])[*])[*]	0.361	17 out of 36

SCFP_6	55464376	AND Enantiomer HO	0.345	14 out of 30
SCFP_6	-1272768868	AND Enantiomer HO HO (*) CC(=[*])[*]	0.242	26 out of 63
	Top Feat	ures for negative of	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1211866396	AND Enantiomer HO	-1.1	2 out of 25
SCFP_6	-1640858361	AND Enantiomer HO	-0.817	1 out of 11
SCFP_6	-1852892018	AND Enardiomer HO	-0.674	0 out of 3



C₉H₁₃NO

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

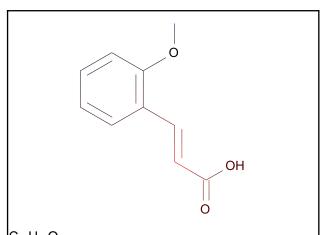
Structural Similar Compounds					
Name	Ephedrine	Tocainide	Amphetamine		
Structure	N H	NH 2	NH ₂		
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen		
Distance	0.314	0.489	0.514		
Reference	US FDA (Centre for Drug Eval.& Res./Off. 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.

Feature Contribution				
	Top Fea	atures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1702724181	AND Enantiomer NH ₂ OH [']C(['])(C@H)(O)(e)(:[']);[']	-1.22	1 out of 18

SCFP_6	-561151481	AND Enantiomer . NH ₂ OH	-1.07	1 out of 15
SCFP_6	1318513260	[cH]:(cH]:T]:[cH]: [cH]:1 AND Enantiomer NH ₂ OH ['NIC@@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

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

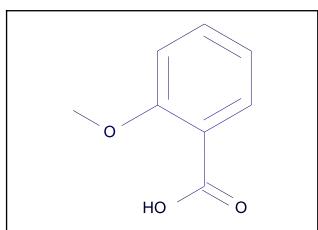
Name	Aspirin	Coumarin	Naproxen
Structure	O OH		OH NO
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.507	0.511	0.528
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& 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.

Feature Contribution				
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	966282057	[*]O[c]1:[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
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	503541685	[*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:10C	-0.484	1 out of 7
SCFP_6	1434056623	[*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:10C	-0.278	0 out of 1
SCFP_6	-2056718782	[*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.219	23 out of 90



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

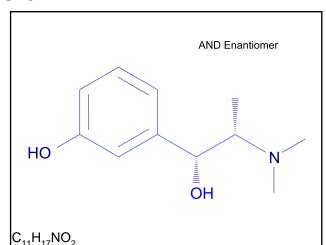
Name	Aspirin	Phenacetin	Acetaminophen
Structure	O OH	H N O	HONH
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.369	0.559	0.580
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug 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.

Feature Co	ntribution				
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	13	HO [*]=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	HOOO [*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:1OC	-0.484	1 out of 7
SCFP_6	1434056623	HO O [*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:10C	-0.278	0 out of 1
SCFP_6	1726855784	[*]O[c]1:[cH]:[cH]:[c H]:[cH]:[c]:1C(=O)O	-0.278	0 out of 1



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.

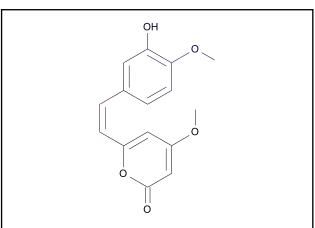
Structural Similar Compounds					
Name	Ephedrine	Phenylephrine	Pronetalol		
Structure	OH N H	HO	HN		
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen		
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Distance	0.399	0.431	0.543		
Reference	US FDA (Centre for Drug Eval.& Res./Off. 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.

Feature Co	ntribution			
	Top fe	eatures for positive c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	611156666	AND Enantiomer OH [*][c]1:[*]:[cH]:[cH] :[c](O):[cH]:1	0.186	6 out of 15
	Top Fe	eatures for negative of	contribution	า
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
	-	•	-	•

SCFP_6	1702724181	HO OH :	-1.22	1 out of 18
SCFP_6	-1632615624	(*)]C([*))[c]1:[cH]:[c H]:[cH]:[c](O):[cH]:	-1.07	0 out of 6
SCFP_6	-561151481	AND Enantiomer OH [*]C([*)](C@H)(O)(c]1 :[cH];(cH]:1 [cH]:1	-1.07	1 out of 15



C₁₅H₁₄O₅

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Cytembena	Atropine	Scopolamine	
Structure	OH Br	N OH	O NE NO OH	
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen	

Model Applicability

Predicted Endpoint

Distance

Reference

Structural Similar Compounds

Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.478

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

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.588

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

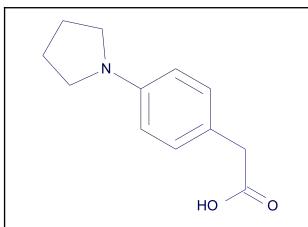
Eval.& Res./Off. Testing &

0.588

1. OPS PC10 out of range. Value: -4.2526. Training min, max, SD, explained variance: -4.2502, 5.657, 1.784, 0.0297.

Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set SCFP_6 -1971137145 0.434 5 out of 9

SCFP_6	392579710	[*][c]1:[*]:[cH]:[c](C=[*]):[cH]:[c]:10	0.425	2 out of 3
SCFP_6	-1977229858	[*]C(=[*])\C=C\[c](:[cH]:[*]):[e]([*]):[*]	0.425	2 out of 3
	Top Feat	tures for negative of	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	2116304939	[*]O[c]1:[cH]:[*]:[c] ([*]):[cH]:[c]:10	-0.825	0 out of 4
SCFP_6	1559374977	OH O O O O O O O O O O O O O O O O O O	-0.278	0 out of 1
SCFP_6	1616083408	OH O OH O OH O OH OH OH OH OH OH OH OH O	-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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Naproxen	Methylphenidate	Eugenol	
Structure	OH NO	N H O O	OH OH	
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.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1380909229	[*]N([*])[c]1:[cH]:[cH]:[cH]:11	0.287	17 out of 39

SCFP_6	10	HO O	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
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-2056996303	HO O	-0.825	0 out of 4
SCFP_6	-1576616475	HO 0	-0.674	0 out of 3
SCFP_6	-1476721585	[,]:[eH]:[c](CC(=0)0)	-0.674	0 out of 3

US FDA (Centre for Drug

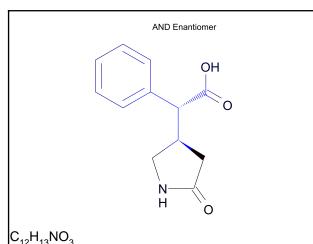
Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



1211131103

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Phenobarbital	Aminoglutethimide	Tocainide	
Structure	HN O	NH ₂	NH NH 2	
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen	
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen	
Distance	0.470	0.502	0.559	

Model Applicability

Reference

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.

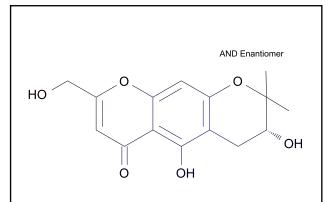
US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Feature Co	ntribution			
	Top fea	ntures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1272768868	AND Enantiomer OH N N O [*]C([*])CC(=[*])[*]	0.242	26 out of 63

SCFP_6	13	AND Enantiomer	0.0717	90 out of 261
		OH OH		
		N H		
		[*]=O		
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	699559848	AND Enantiomer OH	-1.44	0 out of 10
		N-O		
		[*]C([*])[C@@H](C(=[*])[*])[e](:[*]):[*]		
SCFP_6	-1214451192	AND Enantiomer OH N N H O	-1.27	0 out of 8
		[*]C([*])[C@@H](C(=[*))[*])[c]1:[cH];cH] :[*]:[cH]:[cH]:1		
SCFP_6	413428563	AND Enantiomer OH OH	-0.957	0 out of 5
		NH O		
		[*]C([*])[C@@H](C(=O) O)[c](:[*]):[*]		



C₁₅H₁₆O₆

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Hydrocortisone	Piroxicam	Monocrotaline	
Structure	OH OH OH OH	OH HN PP	NO OH OH	
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.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971196727	AND Enantiomer HO OH [*]C(=CC(=[*])[*])[*]	0.361	17 out of 36

SCFP_6	-711656408	AND Enantiomer HO OH OH	0.266	7 out of 16
		[*]C(=[*])CO		
SCFP_6	1416196903	AND Enantiomer HO O OH OH	0.213	5 out of 12
		[*]OC(C)(C)C([*])[*]		
	Top Feat	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1681833008	AND Enantiomer HO OH OH [*]C[c]1:[c]([*]):[*] :[c]([*]):[c]:1O	-0.496	0 out of 2
SCFP_6	-1272709286	AND Enantiomer HO O OH OH [*]CC[c](:[*]):[*]	-0.459	12 out of 61
SCFP_6	-617610981	AND Enantomer HO OH OH [*][c](:[*]):[c]1C(=O)C=C([*])[*][c]:1:[*	-0.278	0 out of 1

AND Enantiomer HO HO CI

C₁₇H₁₇CIO₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Hydrocortisone	Beclomethasone	Naltrexone
Structure	OH OH OH OH	HO state of the st	HO the state of th
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

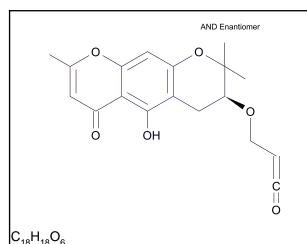
Model Applicability

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 Co	ntribution			
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971196727	AND Enantiomer HO CI [*]C(=CC(=[*])[*])[*]	0.361	17 out of 36

SCFP_6	55464376	AND Enantiomer OH O	0.345	14 out of 30
		Ho Ci		
		[*]CC(=C([*])[*])C(=[*])[*]		
SCFP_6	-52163885	AND Enantiomer HO CI	0.32	7 out of 15
		[*]C([*])Cl		
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1211866396	AND Enantomer HO CI	-1.1	2 out of 25
		[*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1		
SCFP_6	-1640858361	AND Enantiomer HO CI [*]CC[c]1:[cH]:[cH]:[-0.817	1 out of 11
SCFP_6	-1852892018	AND Enuntomer HO CI (*)C(=[*])CC[c](:[cH] :[*]):[cH]:[*]	-0.674	0 out of 3



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Naltrexone	Tolazamide	Indapamide
Structure	HO the Control of the	T N N N N N N N N N N N N N N N N N N N	HN Zz S=0

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

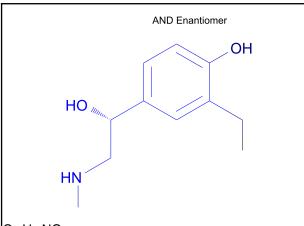
Model Applicability

Structural Similar Compounds

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

Feature Co	ntribution			
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1971196727	AND Enantiomer OH OH [*]C(=CC(=[*])[*])[*]	0.361	17 out of 36

SCFP_6	-711656408	AND Enantiomer O OH C C O	0.266	7 out of 16
		[*]C(=[*])CO		
SCFP_6	2	AND Enantiomer OH OH	0.238	8 out of 19
		C		
		[*]=C=[*]		
	Top Featur	es for negative c	ontribution	
Fingerprint	<u> </u>			Carcinogen in training set
SCFP_6	-1681833008	AND Enantiomer O OH C O (*]C[c]1:[c](([*]):[*] :[c](([*]):[c](C(=[*]) [**]\\[[*](1.40)	-0.496	0 out of 2
SCFP_6	-1272709286)[*]):[c]:10 AND Enantiomer OH [*]CC[c](:[*]):[*]	-0.459	12 out of 61
SCFP_6	-1849894309	AND Enantiomer OH OH C C OH (**)[:c]:1C[C@ @H)(O)C([**)]([**)][**] [c]:1:[*]	-0.278	0 out of 1



C₁₁H₁₇NO₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Phenylephrine	Terbutaline	Metaproterenol
Structure	HO	HO HO	HO NO HIN
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 of 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.

Feature Co	ntribution			
	Top Fea	atures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1702724181	AND Enantiomer HO	-1.22	1 out of 18

SCFP_6	1578545183	AND Enantiomer OH HO HO ["]C[C@H](O)[6]1:[cH] :[cH]:[c](O):[c](['])[[cH]:1	-1.07	0 out of 6
SCFP_6	-561151481	AND Enantiomer HO (*)*C(!*)*(C@H)*(O)*c!1 :[cH]:(cH]:(cH]:(cH]:	-1.07	1 out of 15

AND Enantiomer
OH
OH J
Milian
Alluman
OH
OH

Molecular Weight: 276.371

ALogP: 2.165 Rotatable Bonds: 4

Acceptors: 3
Donors: 3

 $C_{17}H_{24}O_{3}$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Pindolol	Carteolol	Procarbazine	
Structure	HN OH H	HN OH H	N H NH	
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	AND Enantiomer OH OH (*)C(*))(C@@H)(Ctc)(:(cH):(*),(cH):(*)) C=(*)	0.415	1 out of 1	

SCFP_6	55434585	AND Enantiomer OH OH OH (*)\C=C(\C)\C([*])[*]	0.331	12 out of 26
SCFP_6	-1043310069	AND Enantiomer OH OH (*)C(C@@H)(C=(*))C((*))(*)	0.296	15 out of 34
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1211866396	AND Enantiomer OH OH (*)CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.1	2 out of 25
SCFP_6	-1640858361	AND Enantiomer OH OH (*)CC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.817	1 out of 11
SCFP_6	1653911926	AND Enantiomer OH OH OH (*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.504	12 out of 64

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

O NH N S N N N N N N N N N N N N N N N N
N

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Reserpine	Deserpidine	Bromocriptine	
Structure		OKAN MANO MANO MANO MANO MANO MANO MANO MA	Br M N N N N N N N N N N N N N N N N N N	
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen	
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen	
Distance	0.771	0.812	0.842	

Model Applicability

Reference

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.

US FDA (Centre for Drug

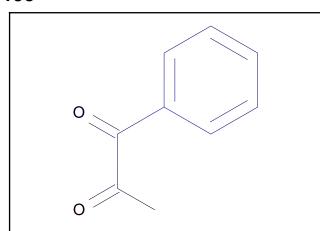
Eval.& Res./Off. Testing &

Res.) Sept. 1997

2. OPS PC22 out of range. Value: 3.5288. Training min, max, SD, explained variance: -3.2994, 3.2573, 1.197, 0.0134.

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set SCFP_6 -1903488337 0.603 2 out of 2

SCFP_6	-1379673609	AND Enantiomer	0.526	11 out of 19
		[*][c](:[*]):[c]1:[cH]:[cH]:[c]([*]):[*]: [c]:1:[*]		
SCFP_6	1310748454	AND Enantiomer	0.437	7 out of 13
		A SUN OF OUR OWN OF OUR OWN OF OUR OWN		
		[*][c]1:[*]:[cH]: s:1		
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	182902497	AND Enantiomer	-0.825	0 out of 4
		[*]NS(=O)(=O)C1[*][*] 1		
SCFP_6	306578635	AND Enantiomer	-0.825	0 out of 4
		[*]CCCN(C)C(=[*])[*]		
SCFP_6	1260369147	AND Enantiomer	-0.674	0 out of 3
		A SUBJECTION OF SUBJECT OF SUBJEC		
		[*]CCC\C=C/[*]		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Coumarin	Antipyrine	Phenacetin	
Structure		N N N N N N N N N N N N N N N N N N N	O H	
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	

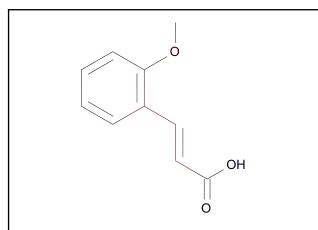
Model Applicability

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

	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	571795252		0.201	2 out of 4
		[*]C(=[*])C(=O)C		

SCFP_8	13		0.172	39 out of 90
SCFP_8	1	[*]=O [*]C(=[*])[*]	0.146	38 out of 90
	Top Feat	tures for negative of	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(=0)[c](:[*]);[*]	-0.737	0 out of 3
SCFP_8	136627117	[*]C(=[*])C	-0.41	4 out of 18

Res.) Sept. 1997



C₁₀H₁₀O₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Phenacetin	Coumarin	Cytembena
Structure	н		
			OH Br
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 &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &

Model Applicability

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

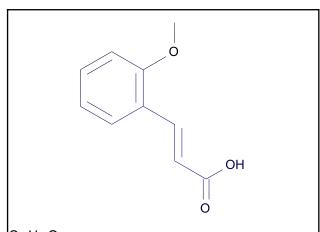
Res.) Sept. 1997

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

Res.) Sept. 1997

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=C\[c](:[0.553	2 out of 2

SCFP_8	966282057	C !	0.553	2 out of 2
		(*]O[c]1:[cH]:[*]:[cH]:[c]:1\C=C\[*]		
SCFP_8	2109620068	[*]O[c]1:[cH]:[*]:[cH	0.553	2 out of 2
]:[cH]:[c]:1\C=C\C(= O)O		
	Top Featu	res for negative c	ontribution	
Fingerprint	Bit/Smiles		Score	Multiple- Carcinogen in training set
SCFP_8	-424485343	OH	-0.584	3 out of 17
		[*]C(=[*])O		
SCFP_8	1132907712	-° H	-0.58	2 out of 12
		[*]=CC(=O)O		
SCFP_8	2034654200	OH OH	-0.546	0 out of 2
		[*]\C=C\C(=O)O		



 $C_{10}H_{10}O_3$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sim	Cinnamic acid	· · · · · · · · · · · · · · · · · · ·	
Structure	ОН	OH O	CI C
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

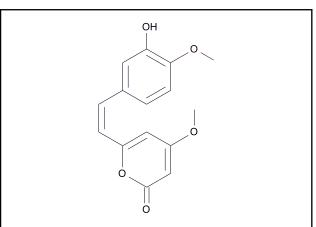
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

Feature Co	ntribution			
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
	!	<u>'</u>		<u>'</u>

FCFP_12	1458856986	[*]=C[c]1:[cH]:[cH]:[cH]:[cH]:[c]:10C	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
	Ton Featur	es for negative c	ontribution	

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	-395337223	[*]O[c]1:[cH]:[*]:[cH]:[c]:1\C=C\[*]	-0.733	0 out of 3
FCFP_12	-1996489595	[*]C(=[*])\C=C\[o]1:[cH]:[cH]:[cH]:[cH]:[c]:10C	-0.543	0 out of 2

FCFP_12	-2107131107		-0.458	1 out of 6
		ОН		
		0		
		[*]\C=C\C(=O)O		



C₁₅H₁₄O₅

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	O OH N	O OH	HO		
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

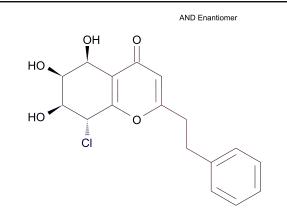
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

Feature Contribution					
	Top fe	eatures for positive of	ontribution		
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	OH O	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	OH O O	-0.543	0 out of 2	
FCFP_12	451371068	[*]\C=C\[c](:[*]):[*]	-0.226	9 out of 32	

FCFP_12	7	ŏ-°	-0.223	20 out of 70
		[*]0		



 $C_{17}H_{17}CIO_5$

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 nonnormal 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	H ₂ N ₁ , NH ₂	OH OH	OHCI CI CI OH	
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.

 OPS PC16 out of range. Value: 3.6979. Training min, max, SD, explained variance: -4.234, 3.1326, 1.134, 0.0174.

Top featu	res for positive c	ontribution	
it/Smiles	Feature Structure	Score	Moderate_Severe in training set
	•		Top features for positive contribution t/Smiles Feature Structure Score

FCFP_12	113	68 out of	0.503	AND Enantiomer HO CI [*]C([*])C(=C([*))[*])C(=[*])[*]	436886043	FCFP_12
HO OH OH	3	4 out of 6	0.458	HO CI [*]C(=[*])CC[c]1:[cH] :[cH]:[cH]:[cH]	1388176727	FCFP_12
[*]CC[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	41	21 out of	0.332	HO Ci	-497728148	FCFP_12

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	32	AND Enantiomer HO CI	-0.789	16 out of 101
		[*]CI		
FCFP_12	71953198	AND Enentiomer HO CI State (5th) Old	-0.548	10 out of 50
		[*]C([*])Cl		

FCFP_12	-1305493555	oн ü	-0.308	0 out of 1
		HO		
		[#10004-00/-P1\[#1-0		
		[*]CCC1=CC(=[*])[*]=C ([*])O1		

C₁₈H₁₈O₆

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 nonnormal 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	OHCI CI CI OH	H ₂ N ₁ , NH ₂	HO O O O O O O O O O O O O O O O O O O		
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

	Top fe	atures for positive of	ontribution	
ingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
				L

FCFP_12	-2000646149	AND Enantiomer AND Enantiomer C C (')O[C@H]1C[o](:[c](['1)):[']):[o](:['1])[']C1(['1])[']	0.365	2 out of 3
FCFP_12	-1099193755	AND Enantiomer OH C C O [*]=C[c]1:[cH]:[*]:[c H]:[cH]:[c]:1OC	0.343	6 out of 11
FCFP_12	451847724	AND Enantiomer OH C C OH C OH	0.257	101 out of 216

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	-1549192822	AND Enantiomer OH C C O	-0.909	1 out of 11
FCFP_12	-1979984779	[*]=CC(=O)[c](:[*]):[-0.308	0 out of 1
		[*][C@H]1[*][c](:[*]) :[c](OC1(C)C):[cH]:[*]		

FCFP_12	1244036906	AND Enantiomer	-0.308	0 out of 1
		\ \\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\		
		O OH		
		č Ji		
		° [*][c](:[*]):[c]1C(=O		
)C=C([*])[*][c]:1:[*		
]		

AND Enantiomer
ОН ОН
n _{nn}
Mining
· · · · · · · · · · · · · · · · · · ·
OH

 $|C_{17}H_{24}O_3|$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

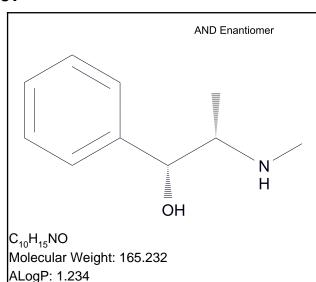
Name	o-Toluidine, N,N-bis(2- hydroxyethyl)-	Phenol, 4-(3- carbazolylamino)-	2-Imidazoline, 2-(2,6-dichloroanilino)-	
Structure	OH OH	H N N N N N N N N N N N N N N N N N N N	CI H N HN	
Actual Endpoint	Mild	Mild	Mild	
Predicted Endpoint	Mild	Mild	Mild	
Distance	0.714	0.747	0.754	
Reference	85JCAE "Prehled 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/yea 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.

Feature Co	Feature Contribution						
Top features for positive contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set			
FCFP_12	-2005486458	AND Enantiomer OH OH (*)C((*))(C@@H)(C(c)(:(cH)(*))(cH)(*)) =(*)	0.717	4 out of 4			

FCFP_12	1384456758	AND Enantiomer OH OH OH (C) (C) (C) (C) (C) (C) (C) (C	0.653	3 out of 3
		:[cH]:[cH]:[cH] :[cH]:1)C=[*]		
FCFP_12	436886043	AND Enantiomer OH	0.503	68 out of 113
		[*]C([*])C(=C([*])[*])C(=[*])[*]		
	Ton Foot			
F*		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	-1272709286	AND Enantiomer OH OH OH OH	-0.475	18 out of 82
		[*]C([*])([*])CO		
FCFP_12	3	AND Enantiomer OH OH OH OH OH	-0.234	58 out of 204
FCFP_12	-2093839777	AND Enantiomer OH OH OH (*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.207	30 out of 103



Model Prediction

Rotatable Bonds: 3

Acceptors: 2 Donors: 2

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

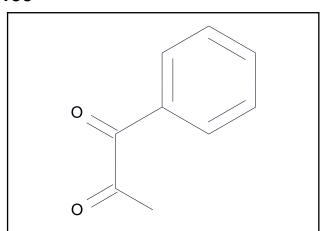
Structural Similar Compounds					
Name	Ethanol, 2-toluidino-	Benzyl alcohol, alpha- methyl-	1,3-Hexanediol, 2-ethyl-		
Structure	HO NH	OH	ОН		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Non-Irritant	Irritant	Irritant		
Distance	0.439	0.519	0.557		
Reference	85JCAE "Prehled Prumyslove 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.

Feature Co	Feature Contribution						
Top Features for negative contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set			
FCFP_12	136686699	AND Enantiomer OH H	-0.484	3 out of 6			
		[*]NC					

FCFP_12	-1083860676	AND Enantiomer	-0.153	3 out of 4
		["]C(("))[C@H](O)[c]1 :[cH]:[cH]:[cH] :[cH]:1		
FCFP_12	1618154665	AND Enantiomer	-0.0845	412 out of 490
		 И Н		
		[*][c](:[*]):[cH]:[cH]:[*]		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Propiophenone	Acetophenone, 4'- methoxy-	o-Anisaldehyde		
Structure	0				
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.

Feature Co	ntribution					
	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		-0.261	4 out of 6
		[*]C(=[*])C(=0)[c]1:[cH]:[cH]:[cH]:[cH]: cH]:1		
FCFP_12	-581162801		-0.101	9 out of 11
		[*]C(=[*])C(=0)[c]1:[cH]:[cH]:[t]:[cH]:[c H]:1		
FCFP_12	1618154665		-0.0845	412 out of 490
		[*][c](:[*]):[cH]:[cH]:[*]		

AND Enantiomer HO HO O

C₁₇H₁₈O₄ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

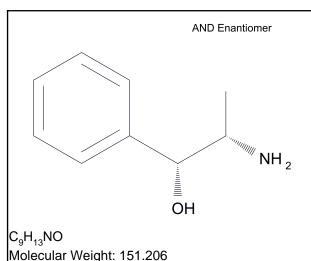
Name	Benzoin, oxime	Benzophenone, 2,4- dihydroxy-	1-Amino-2-bromo-4- hydroxyanthraquinone
Structure	N OH	HONNOH	HO the Br
Actual Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.639	0.668	0.669
Reference	28ZPAK -,111,72	28ZPAK -,101,72	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.

Feature Co	Feature Contribution					
	Top fe	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	436886043	AND Enantiomer HO HO (*)CC(=C((*))(*))C(=[0.0804	129 out of 130		

FCFP_12	436915834	AND Enantiomer HO HO	0.0756	6 out of 6
FCFP_12	565968762	[*]C\C(=C\[*])\O[*] AND Enantiomer HO HO [*]C(=[*])C(=O)C	0.075	78 out of 79
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1388176727	AND Enantiomer HO	-0.12	8 out of 10
FCFP_12	1618154665	AND Enantiomer HO HO (*)[c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490
FCFP_12	16	AND Enantiomer HO	-0.0843	423 out of 503



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

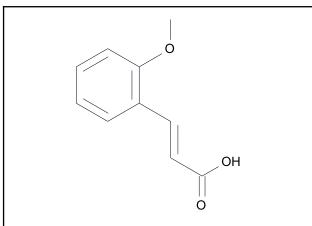
Structural Similar Compounds				
Name	Methylamine, m- phenylenebis-	Methylamine, p- phenylenebis-	Acetanilide, 2'-hydroxy-	
Structure	H ₂ N NH ₂	H ₂ N NH ₂	OH H	
Actual Endpoint	Irritant	Irritant	Non-Irritant	
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant	
Distance	0.460	0.468	0.492	
Reference	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.

Feature Co	ntribution			
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1083860676	AND Enantiomer NH2	-0.153	3 out of 4
		["]C(("))[C@H](O)[c]1 :[cH]:[cH]:[cH] :[cH]:1		

FCFP_12	1618154665	AND Enantiomer	-0.0845	412 out of 490
		·, NH2		
		ОН		
		[*][c](:[*]):[cH]:[cH]:[*]		
FCFP_12	16	AND Enantiomer	-0.0843	423 out of 503
		OH NH ₂		
		[*][c](:[*]):[*]		



 $|C_{10}H_{10}O_3|$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simila	ar Compounds		
Name	Cinnamic acid	Salicylic acid, ethyl ester	Acetic acid, ((4-chloro-o-tolyl)oxy)-
Structure	ОН	OH O	CI WHAT OO OH
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

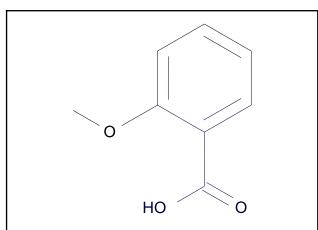
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

ntribution			
Top fe	atures for positive c	ontribution	
Bit/Smiles	Feature Structure	Score	Irritant in training set
			•
	Top fe	Top features for positive of	Top features for positive contribution

FCFP_12	-146015125		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(=[,])[,])[,] OH	0.0737	270 out of 274
	Ton Foatur	es for negative c	ontribution	

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-548632217	OH	-0.128	49 out of 61
		[*]C(=[*])O		
FCFP_12	7	OH	-0.118	104 out of 128
		[*]O		

FCFP_12	1458856986	- 0	-0.109	4 out of 5
		OH		
		(*]=C[c]1:[cH]:[cH]:[
		cH]:[cH]:[c]:1OC		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

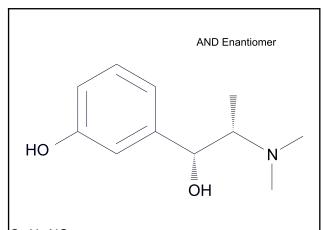
Structural Similar Compounds				
Name	Salicylic acid, methyl ester	Acetic acid, phenoxy-	Benzoic acid	
Structure	OH O	О	OH	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant	
Distance	0.355	0.381	0.393	
Reference	FCTXAV Food and Cosmetics Toxicology. (London, 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.

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	HO (=O)O	-0.612	5 out of 11
FCFP_12	-548632217	HO 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 nonnormal 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	OH OH	OH N OH	НО	
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.

Feature Contribution						
	Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	949015626	AND Enantiomer : OH [*]C([*])[c]1:[cH]:[*]:[cH]:[c](O):[cH]:1	-0.222	2 out of 3		

FCFP_12	-1083860676	AND Enantiomer OH [']C(['])[C@H](O)[c]1 :[cH][cH][cH][cH]	-0.153	3 out of 4
FCFP_12	7	:[cH]:1	-0.118	104 out of 128

C₁₅H₁₄O₅

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 nonnormal 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)undecan e, 1-(3-aminopropyl)-	1-Valeryl-2-pipecoline	
Structure	O S N N N N N N N N N N N N N N N N N N	H ₂ N O Si O	O OH	
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

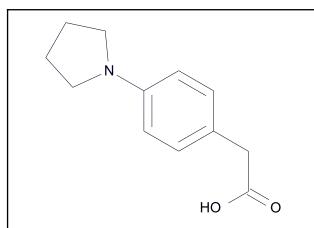
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

Feature Contribution					
Top features for positive contribution					
Bit/Smiles	Feature Structure	Score	Irritant in training set		
•	·	-	•		
	Top fe	Top features for positive of	Top features for positive contribution		

FCFP_12	-1251367201	OH O	0.0858	31 out of 31
FCFP_12	-146015125	[*]C(=[*]))C=C\[o](:[0.085	24 out of 24
FCFP_12	523826990	[*]O[c]1:[cH]:[*]:[c] ([*]):[cH]:[c]:10	0.0756	6 out of 6

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	OH O\ \(\sigma\)	-0.118	104 out of 128

FCFP_12	-549108873	○H ○ ○ ○ ○	-0.11	54 out of 66
		[*]:[c](:[*])O		



C₁₂H₁₅NO₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name Acetic acid, ((4-chloro-o-tolyl)oxy)- 2		2-Butanone, 4-(4-hydroxy- 3-methoxyphenyl)-	Acetic acid, (2,4- dichlorophenoxy)-		
Structure	O OH	OH OH	CI OH		
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.

Feature Contribution				
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
			!	'

FCFP_12	1985089045	[*]N([*])[c]1:[cH]:[c H]:[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	HOOO [*][e]1:[eH]:[eH]:[e] (:[eH]:[eH]:1)N2CCCC	0.0583	2 out of 2
	Top Featur	[*][c]1:[cH]:[cH]:[c] (:[cH]:[cH]:1)N2CCCC	contribution	

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-822674211	HOOO	-0.65	0 out of 1
FCFP_12	-548632217	[*]:[c](:[*])N1CCCC1	-0.128	49 out of 61
		HO O		

FCFP_12	7		-0.118	104 out of 128
		\sim		
		но∕о		
		[*]O		

28ZPAK -,441,72

AND Enantiomer OH NH O

 $|C_{12}H_{13}NO_3|$

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

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Acetanilide, 3'-amino-4'- ethoxy-	Ethyl 4-hydroxy-2,6- dimethylcyclohexane carboxylate	Anthranilic acid, N-methyl-
Structure	O H Number of the NH ₂	OH WHO O	O OH
Actual Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.504	0.560	0.564
			1

Model Applicability

Reference

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

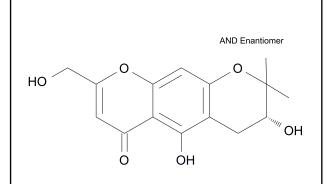
US ARMY

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

28ZPAK -,115,72

Feature Contribution				
	Top fea	atures for positive c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1186303932	AND Enantiomer OH N N H O	0.0838	18 out of 18
		[*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]		

FCFP_12	-1133295320	AND Enantiomer OH NH O [*]C([*])[C@@H](C(=[*])[*)](-][:[cH]]	0.0816	12 out of 12
FCFP_12	-1716639150	:[']:[cH]:[cH]:1 AND Enantiomer OH OH OH (')(('))(C@@H](C(=!*)),'')(c1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]	0.0658	3 out of 3
	Top Featu	res for negative of	contribution)
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	566058135	AND Enantiomer OH NH O=C1C[*][*]N1	-0.367	13 out of 21
FCFP_12	-547731249	AND Enantiomer OH NH O [*][C@@H]1CNC(=0)C1	-0.222	2 out of 3
FCFP_12	-548632217	AND Enantiomer OH N N H O [*]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 nonnormal 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	HO WN H	O OH O OH O OH	H ₂ N ₁₁ NH ₂	
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

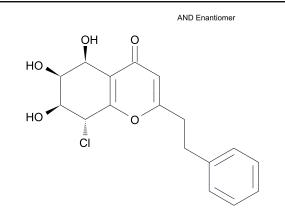
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

Feature Co	ntribution			
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
	•	•	·	•

FCFP_12	436915834	AND Enantiomer HO OH OH [*]C\C(=C\[*])\O[*]	0.0756	6 out of 6
FCFP_12	451847724	AND Enantiomer HO O OH OH [*]C(=CC(=[*])[*])[*]	0.0737	270 out of 274
FCFP_12	-1601875224	AND Enantomer HO OH OH [*]C[c]1:[c]([*]):[*] :[c]((*]):[c](C(=[*])[*]):[c]:10	0.0734	5 out of 5

Fire are any mine t		atures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	7	AND Enantiomer	-0.118	104 out of 128
		HO OH	1	
		[*]0		
FCFP_12	-549108873	AND Enantiomer	-0.11	54 out of 66
		HO OH OH	1	
		[*]:[c](:[*])O		

FCFP_12	74595001	AND Enantiomer	-0.11	54 out of 66
		HO ~ O ~ O ~ O ~ O ~ O ~ O ~ O ~ O ~ O ~		
		о он • он		
		[*]:[cH]:[c](O):[cH]: [*]		
		[]		



C₁₇H₁₇CIO₅

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	HO NH H	O N N N OH	H ₂ N ₁ , NH ₂	
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.

Feature Co	ntribution			
	Top fe	eatures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
	<u> </u>	•	·!	

FCFP_12	436886043	AND Enantiomer OH O Ci (*)CC(=C([*))[*])C(=[0.0804	129 out of 130
FCFP_12	-879019572	AND Enantiomer HO CI ("1C(("))(C@@H)(CI)C(=["1)(")	0.0772	7 out of 7
FCFP_12	436915834	AND Enantiomer HO CI [*]C\C(=C\[*])\O[*]	0.0756	6 out of 6

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1388176727	AND Enantiomer OH OH OCI (*)C(=[*))CC[c]1:[cH] ::[cH]:[cH]:[cH]:[:1	-0.12	8 out of 10
FCFP_12	1618154665	AND Enantiomer HO CI [*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490

FCFP_12	16	AND Enantiomer HO CI	-0.0843	423 out of 503
		[*][0](:[*]):[*]		

C₁₈H₁₈O₆

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sim	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	OHCI CI CI OH	HO MANH 2	HO MNH ₂
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

- 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	AND Enantiomer OH C OH [*]C([*])OCC=C=[*]	0.0863	37 out of 37
FCFP_12	-1687409379	AND Enuntiomer OH C ("C[C@H](OCC=[")]C(["])(")]"]	0.0795	9 out of 9
FCFP_12	436915834	AND Enantiomer OH C C OH (*]C\C(=C\[*])\O[*]	0.0756	6 out of 6

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	7	AND Enantiomer O OH C C	-0.118	104 out of 128
FCFP_12	-549108873	AND Enantiomer OH C C C C C C C C C C C C C	-0.11	54 out of 66
		[*]:[c](:[*])O		

FCFP_12	74595001	AND Enantiomer	-0.11	54 out of 66
		\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\		
		O OH		
		C: 0		
		[*]:[cH]:[c](O):[cH]:		
		[*]		

AND Enantiomer OH HO MINISTRACTOR OF THE PROPERTY OF THE PR

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 nonnormal 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	OH OH	OH Z OH	HONNOH	
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.

Feature Co	Feature Contribution					
	Top fea	tures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	-1710155973	AND Enantiomer HO	0.0658	3 out of 3		

FCFP_12	-542873837	AND Enantioner HO . HN [*][c]1:[*]:[cH]:[c](O):[c](CC):[cH]:1	0.0658	3 out of 3
FCFP_12	-306804326	AND Enantiomer HO. HO. HO. HO. CHO. CHO. CHO. CHO. CHO.	0.0658	3 out of 3
	Top Featu	res for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	136686699	AND Enantiomer OH HO, [*]NC	-0.484	3 out of 6
FCFP_12	-1604301295	AND Enantioner HO . HN OH *]C[c]1:[cH]:[*]:[cH]:[cH]:[c]:10	-0.18	22 out of 29
FCFP_12	-451251206	AND Enantiomer OH HO. HO. (*)C[c]1:[cH]:[c](:[c H]:[*]:[c]:1[*])C([*])[*]	-0.132	44 out of 55

AND Enantiomer
OH
OH Manna
Allum,
OH

C₁₇H₂₄O₃ 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 nonnormal 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	OH NOH	OH OH	HONNOH	
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.

Feature Co	Feature Contribution					
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	-415156552	AND Enantlomer OH OH (*)C(C@)(C)(O)C((*))(1)	0.0854	27 out of 27		

FCFP_12	436886043	AND Enantiomer OH OH OH	0.0804	129 out of 130
FCFP_12	451847724	[*]CC(=C([*])[*])C(=[*])[*] AND Enantiomer OH OH OH	0.0737	270 out of 274
	Top Fee	rjc(=cc(=(rj)(rj)(rj	Contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1618154665	AND Enantiomer OH OH (*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490
FCFP_12	16	AND Enantiomer OH OH OH (*)[c](:[*]):[*]	-0.0843	423 out of 503
FCFP_12	3	AND Enantiomer OH OH OH OH OH	-0.0812	291 out of 345

 $C_{38}H_{47}N_5O_7S_2$

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		The state of the s		
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

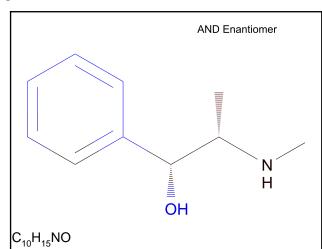
- 1. OPS PC6 out of range. Value: 7.6964. Training min, max, SD, explained variance: -5.7225, 6.867, 1.786, 0.0403.
- 2. OPS PC22 out of range. Value: 5.13. Training min, max, SD, explained variance: -2.9568, 3.7845, 1.016, 0.0131.

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		

FCFP_12	-415156552	AND Enantiomer 0.0854	27 out of 27
FCFP_12	-53728878	AND Enantiomer 0.0845	21 out of 21
FCFP_12	1186303932	AND Enantiomer 0.0838	18 out of 18

Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	690481386	AND Enantiomer AND Enantiomer N (*):[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	AND Enantiomer	-0.0893	20 out of 24
		A. O. Nes		
		A SUP N-O.		
		[*][c](:[*]):[c]1:[cH]:[cH]:[c]([*]):[*]:		
		[c]:1:[*]		



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	NH HN O	NH NH NH 2	CI CI		
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		

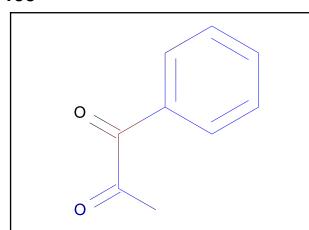
Model Applicability

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.
- $2. \qquad \text{Unknown ECFP_2 feature: -222940837: } [*]C([*])[C@H](O)[c](:[*]):[*]$
- 3. Unknown ECFP_2 feature: -871548073: [*]N[C@@H](C)C([*])[*]
- 4. Unknown ECFP_2 feature: 1336634178: [*]C([*])NC

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score ECFP_6 734603939 AND Enantiomer 0.0424

ECFP_6	864518973	AND Enantiomer	0.0366
		N	
		- Он ^Н	
		[*]C([*])C	
ECFP_6	-1897341097	AND Enantiomer	0.0284
		OH H	
		[*]N[*]	
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	AND Enantiomer	-0.251
		OH H	
		[*][c](:[*]):[cH]:[cH]:[*]	
ECFP_6	642810091	AND Enantiomer	-0.247
		OH H	
		[*][c](:[*]):[*]	
ECFP_6	-182236392	AND Enantiomer	-0.232
		OH H	
		[*]:[cH]:[*]	



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

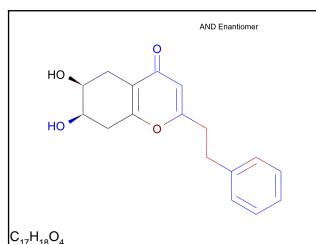
Structural Similar Compounds					
Name	Benzyl acetate	Nitrobenzene	Benzaldehyde		
Structure		0 N	0		
Actual Endpoint (-log C)	2.01823	2.619	1.85262		
Predicted Endpoint (-log C)	2.37939	2.53727	2.81264		
Distance	0.417	0.464	0.500		
Reference	CPDB	CPDB	CPDB		

Model Applicability

- 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 fo	r negative contribution	on
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	[*][c](:[*]):[*]	-0.247
ECFP_6	-182236392		-0.232
		[*]:[cH]:[*]	



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

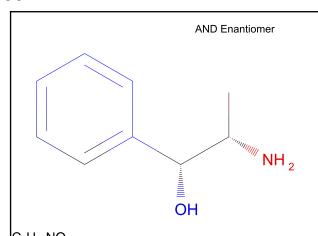
Structural Similar Compounds					
Name	C.I. disperse yellow 3	Phenobarbital s	581		
Structure	NH NH	HN O HN O	H N O Na		
Actual Endpoint (-log C)	2.42163	4.49846	3.89128		
Predicted Endpoint (-log C)	3.11116	3.14828	3.31105		
Distance	0.586	0.592	0.633		
Reference	CPDB	CPDB	CPDB		

Model Applicability

- 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	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	1559650422	AND Enantiomer HO HO [*]C[*]	0.203		

ECFP_6	-2024255407	AND Enantiomer HO HO	0.172
ECFP_6	683445015	[*]C[c](:[cH]:[*]):[c H]:[*]	0.136
LOTT_0	000440010	HO O	0.130
		[*]0[*]	
	Top Features f	or negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer HO HO O [*]C(=O)[*]	-0.275
ECFP_6	1996767644	AND Enantiomer HO HO [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	AND Enantiomer HO HO	-0.247
		[*][c](:[*]):[*]	



C₉H₁₃NO

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

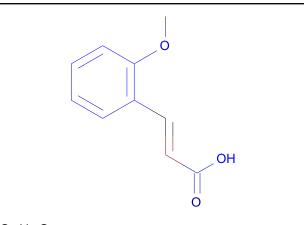
Name	485	Benzoyl hydrazine	529
Structure	CI CI NH NH 2	H ₂ N N H	NH INH 2
Actual Endpoint (-log C)	3.15601	4.1522	3.96981
Predicted Endpoint (-log C)	3.43754	3.02635	3.43178
Distance	0.461	0.470	0.477
Reference	CPDB	CPDB	CPDB

Model Applicability

- 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					
ECFP_6	1572579716	AND Enantiomer ., NH ₂ OH [*]N	0.225		

ECFP_6	734603939	AND Enantiomer	0.0424
		OH 'NH ₂	
		[*]C	
ECFP_6	864518973	AND Enantiomer	0.0366
		NH ₂	
		: OH	
		[*]C([*])C	
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	AND Enantiomer	-0.251
		.'NH ₂	
		[*][c](:[*]):[cH]:[cH]:[*]	
ECFP_6	642810091	AND Enantiomer	-0.247
		NH ₂	
		ОН	
		[*][c](:[*]):[*]	
ECFP_6	-182236392	AND Enantiomer	-0.232
		NH ₂	
		он	
		[*]:[cH]:[*]	



C₁₀H₁₀O₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

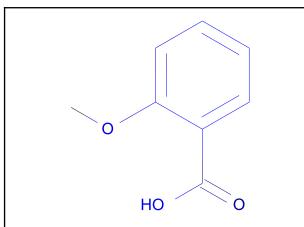
Structural Similar Compounds				
Name	o-Ethoxybenzamide	o-Nitroanisole	153	
Structure	O NH ₂		NH 2	
Actual Endpoint (-log C)	2.50786	2.93466	2.10547	
Predicted Endpoint (-log C)	2.72346	2.36925	2.33861	
Distance	0.479	0.499	0.507	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

Top features f							
	for positive contribution	Top features for positive contribution					
Bit/Smiles	Feature Structure	Score					
-1925046727	[*]C=[*]	0.145					
		-1925046727	-1925046727 0.145				

734603939	[*]C	0.0424
1307307440	[*]:[c](:[*])OC	0.0156
	Feature Structure	Score
1996767644	[*][c](:[*]):[cH]:[cH]:[*]	-0.251
642810091	[*][c](:[*]):[*]	-0.247
-182236392	[*]:[cH]:[*]	-0.232
	1307307440 Top Features f Bit/Smiles 1996767644 642810091	Top Features for negative contribution



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	o-Ethoxybenzamide	153	o-Nitroanisole
Structure	O NH ₂	NH 2	
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

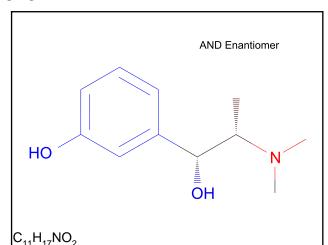
Model Applicability

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.

Top features Bit/Smiles 734603939	for positive contribution Feature Structure	Score	
	Feature Structure	Score	
734603939			
	HO O	0.0424	
_			

ECFP_6	1307307440	HOO [*]:[c](:[*])OC	0.0156
	Top Features f	or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	*][c](:[*]):[cH]:[cH :[*]	-0.251
ECFP_6	642810091	[*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	HO O	-0.232



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	HO HO O	NH HN O	NH NH NH 2		
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

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score ECFP_6 -1072294614 -1072294 -1072294614 -1072294614 -1072294614 -1072294614 -107229461

ECFP_6	865379614	AND Enantiomer HO OH [*]N([*])C	0.219
ECFP_6	734603939	AND Enantiomer HO OH [*]C	0.0424
Fingerprint	Top Features f	for negative contribution Feature Structure	n Score
ECFP_6	2019062761	AND Enantiomer OH [*]:[c](:[*])O	-0.258
ECFP_6	1996767644	AND Enantiomer : HO : OH [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	AND Enantiomer HO : OH N OH OH OH OH OH OH	-0.247

C₁₅H₁₄O₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	C.I. disperse yellow 3	Ciprofibrate	581		
Structure	NH NH	CI	H N Na		
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

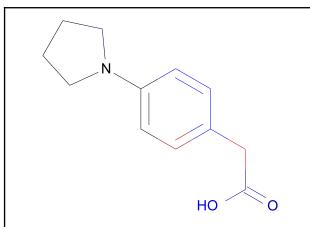
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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	OH O\ \(\circ\) \(\circ\) \(\circ\) \(\circ\) \(\circ\) \(\circ\) \(\circ\) \(\circ\)	0.145		

ECFP_6	683445015	OH O	0.136
ECFP_6	734603939	[*]O[*]	0.0424
		or negative contributio	
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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	666	1'-Hydroxysafrole	Ciprofibrate		
Structure	HN	OH	CI		
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

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score ECFP_6 1559650422 0.203

ECFP_6	-2024255407	[*]C[c](:[cH]:[*]):[c H]:[*]	0.172
ECFP_6	670515721	(*)N([*])[*]	0.00735
		for negative contributio	
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 OH N H O

Molecular Weight: 219.237

ALogP: 0.621 Rotatable Bonds: 3

Acceptors: 3
Donors: 2

C₁₂H₁₃NO₃

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.

Name	Phenobarbital s	Primidone	3-Amino-4- ethoxyacetanilide
Structure	HN HN O	HN	O NH ₂
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

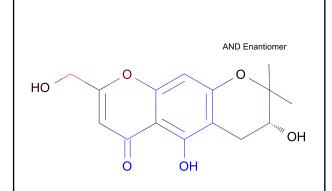
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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	Fingerprint Bit/Smiles Feature Structure Sc				
ECFP_6	-167460056	AND Enantiomer OH N N H O [*]C([*])[*]	0.0596		
		[]0([])[]			

ECFP_6	-1910270391	AND Enantiomer OH N N (*) (*) (*) (*) (*)	0.0279
ECFP_6	1571214559	AND Enantiomer OH OH N OH CH CH CH CH CH CH CH CH CH	0.0145
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	[*]C(=O)[*]	-0.275
ECFP_6	1996767644	AND Enantiomer OH OH O(H) (S) (S) (S) (S) (S) (S) (S)	-0.251
ECFP_6	642810091	AND Enantiomer OH	-0.247



C₁₅H₁₆O₆

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Riddelliine	Zearalenone	71		
Structure	HO m OH OH	HO OH O	HO OH O		
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		

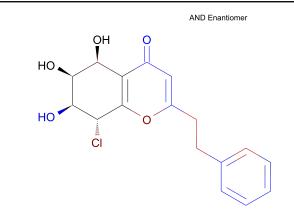
Model Applicability

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([*])[*]

oution					
Top features	for positive contribution	1			
Fingerprint Bit/Smiles Feature Structure Score					
1559650422	AND Enantiomer HO OH [*]C[*]	0.203			
	Bit/Smiles	1559650422 AND Enantiomer HO OH OH			

ECFP_6	683445015	AND Enantiomer HO O OH [*]O[*]	0.136
ECFP_6	657586427	AND Enantiomer HO O OH O OH [*]C([*])([*])[*]	0.0789
Fingerprint	Top Features f	or negative contribution Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer HO O OH OH [*]C(=O)[*]	-0.275
ECFP_6	2019062761	AND Enantiomer HO OHO OH [*]:[c](:[*])O	-0.258
ECFP_6	642810091	AND Enantiomer HO O OH OH [*][C](:[*]):[*]	-0.247



C₁₇H₁₇CIO₅

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 nonnormal 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	NH NH	AND Enantomer AND Enantomer HN HO O	OH HO CI		
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](CI)C(=[*])[*]
- 5. Unknown ECFP_2 feature: 1652274794: [*]OC(=C([*])[*])C([*])[*]

ECFP_6	-2024255407	AND Enantlomer HO CI [*]C[c](:[cH]:[*]):[c H]:[*]	0.172
ECFP_6	683445015	AND Enantiomer HO OH OCI [*]O[*]	0.136
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer HO CI [*]C(=O)[*]	-0.275
ECFP_6	1996767644	AND Enantlomer HO OH OH OH OH OH OH OH OH O	-0.251
ECFP_6	642810091	AND Enantiomer HO CI [*][c](:[*]):[*]	-0.247

C₁₈H₁₈O₆ 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	O NH NH NH	HO OH O	HO OH O	
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

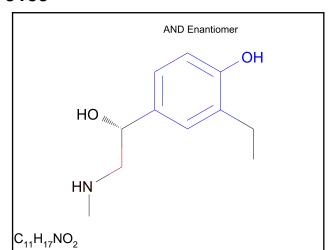
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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

Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	1559650422	AND Enantiomer OH CH CH CH CH CH CH CH CH CH	0.203	
		[*]C[*]		

ECFP_6	-1925046727	AND Enantiomer OH OH [*]C=[*]	0.145
ECFP_6	683445015	AND Enantiomer OH C C O [*]O[*]	0.136
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer OH C C O [*]C(=O)[*]	-0.275
ECFP_6	2019062761	AND Enantiomer OH C:	-0.258
ECFP_6	642810091	AND Enantiomer OH [*][c](:[*]):[*]	-0.247



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 nonnormal 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	HO NH HN O	HO THE NAME OF THE	O NH ₂		
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					
ECFP_6	1559650422	HO, . [*]C[*]	0.203		
	I				

ECFP_6	734603939	AND Enantiomer OH HO,	0.0424
ECFP_6	-1897341097	AND Enantiomer OH HO, [*]N[*]	0.0284
Fig. or a manufact	Top Features f	or negative contribution	
Fingerprint ECFP_6	2019062761	Feature Structure AND Enantiomer	Score -0.258
		HO,	
ECFP_6	1996767644	AND Enantiomer HO	-0.251
ECFP_6	642810091	AND Enantiomer OH HO,	-0.247

C₁₇H₂₄O₃ 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 nonnormal 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	HN O	NH NH	H N O O		
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](:[*]):[*]

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score ECFP_6 1559650422 AND Enantiomer 0.203

ECFP_6	-2024255407	AND Enantiomer OH OH OH	0.172
		[*]C[c](:[cH]:[*]):[c H]:[*]	
ECFP_6	-167460056	AND Enantiomer OH OH	0.0596
		ОН.	
		[*]C([*])[*]	
	Top Features f	or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	AND Enantiomer OH	-0.251
		[*][c](:[*]):[cH]:[cH]:[*]	
ECFP_6	642810091	AND Enantiomer OH OH OH OH	-0.247
ECFP_6	-182236392	[*][c](:[*]):[*] AND Enantiomer OH	-0.232
		ОН	
		[*]:[cH]:[*]	

Simeprevir

TOPKAT_Carcinogenic_Potency_TD50_Mouse

AND Enantiomer
O NH N S N N O O O O O O O O O O O O O O O O
, N

C₃₈H₄₇N₅O₇S₂ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	223	D & C red no. 5	Salicylazosulfapyridine	
Structure	AND Enastromer	OHO Na Na	HN N H	
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

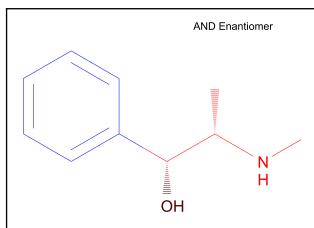
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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 contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
	1	,		

ECFP_6	655739385	AND Enantiomer No. S No. S	0.229
ECFP_6	834876373	[*]:n:[*] AND Enantiomer [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	657586427	AND Enantiomer AND Enantiomer N S N S N S N S N S N S N S N S N S N	0.0789
	Top Features for	negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer AND Enantiomer NS N (*]C(=O)[*]	-0.275
ECFP_6	1996767644	AND Enantiomer AND Enantiomer (*][c](:[*]):[cH]:[cH]:[*]	-0.251

ECFP_6	642810091	AND Enantiomer AND Enantiomer	-0.247
		[*][c](:[*]):[*]	



C₁₀H₁₅NO

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	a-Methylbenzyl alcohol	665	N-Nitrosoephedrine
Structure	OH	НО	N OH
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

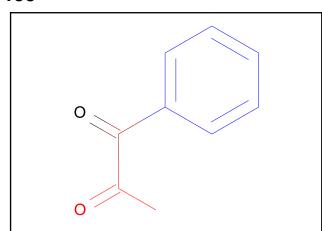
Model Applicability

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	AND Enantiomer OH I*INIC@@HI(C)C([*])[*]	1.15		

FCFP_6	-885550502	AND Enantiomer	0.229
FOED 0	000077700	[*]C([*])NC	0.407
FCFP_6	203677720	AND Enantiomer	0.137
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	
	Top Features f	or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer	-0.422
FCFP_6	-2093839777	AND Enantiomer OH H [*][c]1:[cH]:[cH]:[cH]:[cH]:1	-0.378
FCFP_6	16	AND Enantiomer	-0.354



C₉H₈O₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

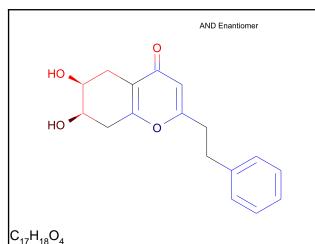
Name	N-Methyl-N- nitrosobenzamide	o-Nitrosotoluene	Nitrosomethylaniline
Structure	0 N = 0	O N	N N N N N N N N N N N N N N N N N N N
Actual Endpoint (-log C)	4.70607	3.37827	5.98173
Predicted Endpoint (-log C)	4.64533	3.14846	4.43574
Distance	0.457	0.466	0.471
Reference	CPDB	CPDB	CPDB

Model Applicability

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

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	565968762	[*]C(=[*])C(=O)C	0.266	

		_	
FCFP_6	1	[*]=0	0.234
FCFP_6	203677720	[*]C([*])[o](:[cH]:[*]):[cH]:[*]	0.137
		egative 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



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

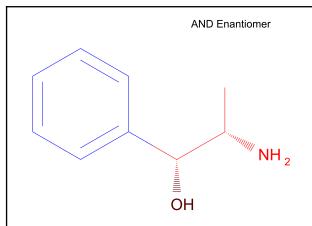
Structural Similar Compounds				
Name	C.I. disperse yellow 3	Indolidan	581	
Structure	NH NH	HN N N N N N N N N N N N N N N N N N N	H N Na	
Actual Endpoint (-log C)	2.85045	5.10722	3.42954	
Predicted Endpoint (-log C)	3.2505	4.15845	3.78228	
Distance	0.562	0.571	0.606	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	-1043250487	AND Enantiomer HO HO [*]N[C@@H](C)C([*])[*]	1.15	

FCFP_6	565968762	AND Enantiomer HO HO C C C C C C C C C C C C C C C C C C C	0.266
FCFP_6	1		0.234
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	AND Enantiomer HO HO O [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	991735244	AND Enantiomer HO HO (*)[c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422
FCFP_6	436886043	AND Enantiomer HO HO (*)CC(=C((*))(*))C(=[-0.383



C₉H₁₃NO

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	a-Methylbenzyl alcohol	665	N-Nitrosoephedrine	
Structure	OH	но	OH NON NO	
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	

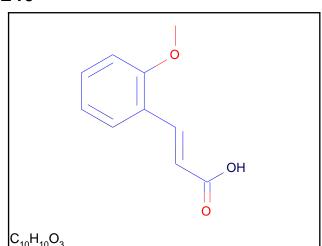
Model Applicability

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

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	-1043250487	AND Enantiomer NH 2 OH [']N[C@@H](C)C(['))[*	1.15	

FCFP_6	203677720	AND Enantiomer	0.137
		·, NH ₂	
		ŌН	
		[*]C([*])[c](:[cH]:[*	
]):[cH]:[*]	
FCFP_6	136597326	AND Enantiomer	0.0695
		· NIL	
		OH NH ₂	
		[*]C([*])C	
	Top Features f	or negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer	-0.422
		.,NH₂	
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	
-CFP_6	-2093839777	AND Enantiomer	-0.378
		ÖH .,NH²	
		On	
		[*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	
FCFP_6	16	AND Enantiomer	-0.354
		· NH ₂	
		ОН	
		[*][c](:[*]):[*]	

┖



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

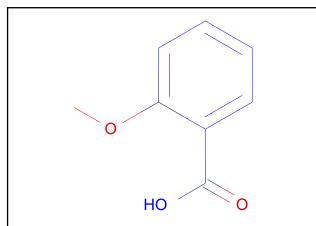
Structural Similar Compounds				
Name	Coumarin s	N-Nitrosoephedrine	Phenacetin	
Structure	0 0	OH Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z	O H	
Actual Endpoint (-log C)	3.57149	3.30968	2.15647	
Predicted Endpoint (-log C)	3.23061	4.9805	2.91803	
Distance	0.468	0.482	0.497	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

Feature Cont	ribution					
	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]:[c H]:[cH]:[c]:1[*]	0.245
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	[*]C(=CC(=[*])[*])[*] OH	-0.436
FCFP_6	991735244	[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422
FCFP_6	7	OH (**)O	-0.372



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

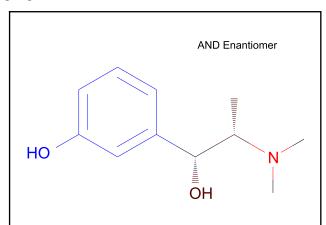
Structural Similar Compounds				
Name	153	p-Cresidine	m-Cresidine	
Structure	NH 2	NH ₂	NH ₂	
Actual Endpoint (-log C)	3.61769	3.14606	2.46519	
Predicted Endpoint (-log C)	3.58569	3.83612	3.87814	
Distance	0.433	0.441	0.458	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_6	136627117	HO O	0.69		

FCFP_6	-2090462286	(*]O[c]1:[cH]:[cH]:[c	0.245
FCFP_6	1	H]:[cH]:[c]:1[*]	0.234
	Ton Footures ([*]=0	un .
Fingerprint	Bit/Smiles	for negative contributio Feature Structure	Score
FCFP_6	991735244	[*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	7	HO 0	-0.372
FCFP_6	16	[*][c](:[*]):[*]	-0.354



C₁₁H₁₇NO₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	N-Nitrosoephedrine	Acetaminophen	3-Methoxycatechol	
Structure	H N N N N N N N N N N N N N N N N N N N	HONH	НООН	
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	

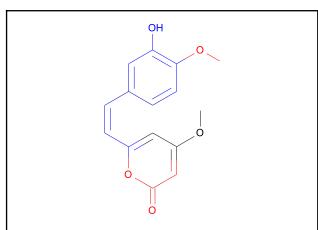
Model Applicability

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

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_6	9	AND Enantiomer OH [*]N([*])[*]	0.385		

FCFP_6	203677720	AND Enantiomer	0.137
		HO ;	
		он Т	
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	
FCFP_6	136597326	AND Enantiomer	0.0695
		но Он N	
		[*]C([*])C	
	Top Features for ne	egative contribution	
ingerprint		Feature Structure	Score
FCFP_6	991735244	AND Enantiomer	-0.422
		HO OH N	
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	
FCFP_6	7	AND Enantiomer	-0.372
		HO OH I	
		[*]O	
-CFP_6	16	AND Enantiomer	-0.354
		HO OH N	
		[*][c](:[*]):[*]	

┖



C₁₅H₁₄O₅ 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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

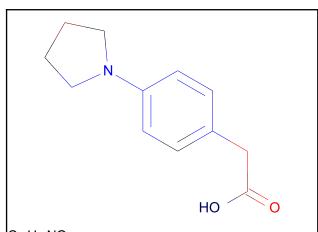
Structural Similar Compounds				
Name	Nalidixic acid	Ciprofibrate	1´-Acetoxysafrole	
Structure	HOOO	CI CI OH		
Actual Endpoint (-log C)	3.06273	5.16666	3.94492	
Predicted Endpoint (-log C)	3.02209	3.87886	3.37654	
Distance	0.572	0.582	0.583	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_6	136627117	[*]OC	0.69			

FCFP_6	565998553	OH O	0.357
FCFP_6	1	[*]OC(=O)C=[*]	0.234
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	[,]C(=CC(=[,])(,])(,]	-0.436
FCFP_6	7	OH O\ (*)0	-0.372
FCFP_6	16	[*][c](:[*]):[*]	-0.354



C₁₂H₁₅NO₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

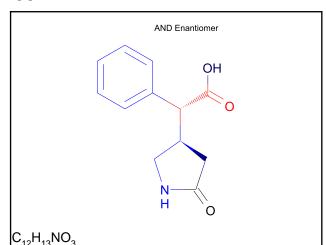
Structural Similar Compounds				
Name	1´-Hydroxysafrole	Ciprofibrate	Phenacetin	
Structure	O O O O O	CI	O H N	
Actual Endpoint (-log C)	3.98605	5.16666	2.15647	
Predicted Endpoint (-log C)	3.25547	3.87886	2.91803	
Distance	0.495	0.496	0.552	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

Feature Contribution Top features for positive contribution				
FCFP_6	-1176841573	[*]=CC(=O)O	0.277	

FCFP_6	203677720	[*]=O	0.234
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	7	HO O	-0.372
FCFP_6	16	[*][c](:[*]):[*]	-0.354
FCFP_6	1674451008	(*)O[c]1:[cH]:[cH]:[c H]:[*]:[c]:1[*]	-0.233



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	581	4-Ethoxy-phenylurea	Indolidan	
Structure	H N O Na	O H N O O	HN N N N N N N N N N N N N N N N N N N	
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

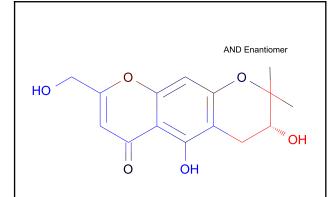
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

 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	AND Enantiomer OH NH O [*]=CC(=O)O	0.277			

FCFP_6	1	AND Enantiomer OH N H O [*]=O	0.234
FCFP_6	-885550502	AND Enantiomer OH OH O [*]C([*])NC	0.229
-		egative contribution	
Fingerprint		Feature Structure	Score
FCFP_6	-1272709286	AND Enantiomer OH N N N (*) [*] [C@H] 1[*] [*] NC1	-0.526
FCFP_6	991735244	AND Enantiomer OH OH ON (*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422
FCFP_6	-2093839777	AND Enantiomer OH OH NHO [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.378

┖



C₁₅H₁₆O₆

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Purpurin	Prednisolone	1085	
Structure	OH O OH O	OH OH OH	AND Enuntomer OH HO OH	
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	

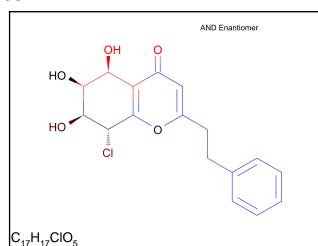
Model Applicability

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

Feature Cont	Feature Contribution				
	Top features	for positive contributio	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_6	-1043250487	AND Enantiomer HO OH [*]N[C@@H](C)C([*])[*]	1.15		

FCFP_6	1	AND Enantiomer HO OH OH [*]=O	0.234
FCFP_6	203677720	AND Enantiomer HO OH OH [*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137
	Top Features for ne	gative contribution	
			Score
FCFP_6	-1272709286	AND Enantiomer HO OH OH [*][C@H]1[*][*]NC1	-0.526
FCFP_6	451847724	AND Enantiomer HO OH OH [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	7	AND Enantiomer HO OH OH [*]O	-0.372

┖



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Prednisolone	1085	Purpurin	
Structure	OH OH OH OH	AND Enantomer HO HO OH	OH O HO OH O	
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	СРДВ	CPDB	CPDB	

Model Applicability

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

Feature Contribution				
	Top features	for positive contribution	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	-1043250487	AND Enantiomer HO CI [*]N[C@@H](C)C([*])[*]	1.15	

FCFP_6	565968762	AND Enantiomer HO HO CI [*]C(=[*])C(=O)C	0.266
FCFP_6	1	AND Enantiomer HO CI [*]=O	0.234
	Top Features for ne		
		Feature Structure	Score
FCFP_6	451847724	AND Enantiomer HO OH O	-0.436
FCFP_6	991735244	AND Enantiomer HO CI [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422
FCFP_6	436886043	AND Enantiomer HO CI [*]CC(=C([*])[*])C(=[*])[*]	-0.383

┖

C₁₈H₁₈O₆ 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 nonnormal 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-nosulpho nyl)benzoic acid		
Structure	HOOO	NH N	HO SINO O		
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

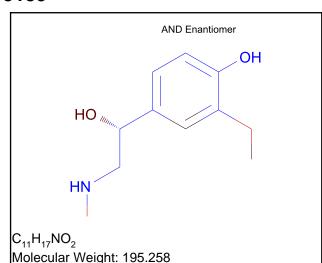
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Cont	Feature Contribution				
	Top features	for positive contribution	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_6	1	AND Enantiomer OH C C O [*]=O	0.234		
		[*]=O		_	

FCFP_6	203677720	AND Enantiomer	0.137
		OH CHARLES	
		= 0:0	
		[*]C([*])[c](:[cH]:[*	
FCFP_6	-1272768868]):[cH]:[*] AND Enantiomer	0.127
G. 1 _0	12,2,0000	\^\\\	0
		ООН	
		<u> </u> 0:0	
		[*]N1[*][*]CC1	
	Top Features for ne		
= -		Feature Structure	Score
-CFP_6	451847724	AND Enantiomer	-0.436
		о он	
		C: O	
		[*]C(=CC(=[*])[*])[*]	
FCFP_6	7	AND Enantiomer	-0.372
		OH OH	
		= 0:0	
-CFP_6	16	[*]O AND Enantiomer	-0.354
0.1 _0		\^\\\	0.001
		O OH	
] 0:0	
		[*][c](:[*]):[*]	

┖



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Salbutamol	Terbutaline	Procarbazine
Structure	HO OH NH	НО	HN

Actual Endpoint (-log C) 3.7769 2.73995 4.74183 Predicted Endpoint (-log 2.66345 5.26514 2.96169 Distance 0.408 0.476 0.513 **CPDB** CPDB Reference **CPDB**

Model Applicability

Structural Similar Compounds

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

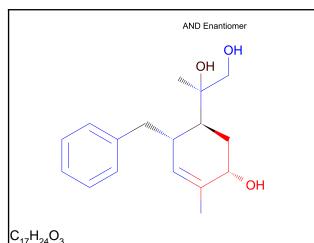
1. All properties and OPS components are within expected ranges.

Fingerprint Bit/Smiles Feature Structure Score FCFP_6 -885550502 -885550502 -0.229

FCFP_6	203677720	AND Enantiomer OH	0.137
		но.	
		HN	
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	
FCFP_6	136597326	AND Enantiomer	0.0695
		но.	
		HN	
		[*]C([*])C	
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1272709286	AND Enantiomer	-0.526
		HO,	
		HN	
		[*][C@H]1[*][*]NC1	
FCFP_6	7	AND Enantiomer OH	-0.372
		HO, .	
		HN	
FCFP_6	16	AND Enantiomer	-0.354
		HO,.	
		HN	
		[*][c](:[*]):[*]	

0.623

CPDB



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Procarbazine	C.I. disperse yellow 3	Terbutaline		
Structure	H N N N N N N N N N N N N N N N N N N N	O NH	HO NH		
Actual Endpoint (-log C)	4.74183	2.85045	2.73995		
Predicted Endpoint (-log	5.26514	3.2505	2.96169		

Model Applicability

Distance

Reference

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

0.619

CPDB

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

0.618

CPDB

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_6	-1043250487	AND Enantiomer OH OH (*)N[C@@H](C)C((*))(*)	1.15			

FCFP_6	203677720	AND Enantiomer OH OH (*]C([*])[c](:[cH]:[*]):[cH]:[*]	0.137
FCFP_6	-1272798659	AND Enantiomer OH OH OH (*]C([*])CC(=[*])[*]	0.11
		or negative contribution	
Fingerprint FCFP_6	Bit/Smiles -1272709286	Feature Structure	Score -0.526
		AND Enantiomer OH OH OH (*][C@H]1[*][*]NC1	
FCFP_6	451847724	AND Enantiomer OH OH (*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	991735244	AND Enantiomer OH OH OH (*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422

Simeprevir

TOPKAT_Carcinogenic_Potency_TD50_Rat

AND Enantiomer
O S NH S N S N S N S N S N S N S N S N S
, N

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	411	223	188	
Structure	Han .	AND Enastrones	No No No	
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	

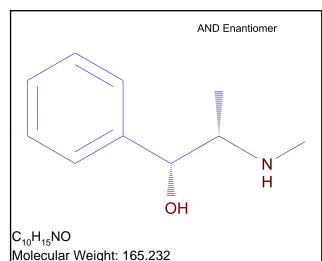
Model Applicability

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	AND Enantiomer NS NS N N N N N N N N N N N N N N N N	0.69			

FCFP_6	565998553	AND Enantiomer AND Enantiomer No. S No. S	0.357
FCFP_6	690511177	AND Enantiomer NS N S N (*):[cH]:[c](:n:[*])[C](:[*]):[*]	0.293
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1175638033	AND Enantiomer AND Enantiomer N S N S N S N S N S N S N S N S N S N	-0.512
FCFP_6	451847724	AND Enantiomer NS N S N (*)C(=CC(=[*])[*])[*]	-0.436
FCFP_6	16	AND Enantiomer AND Enantiomer NS NS N S N S N S N S N S N S N S N S	-0.354



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simila	EPHEDRINE SULPHATE	ALPHA-METHYLBENZYL ALCOHOL	DL-AMPHETAMINE SULFATE
Structure	OH N H	ОН	NH ₂
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

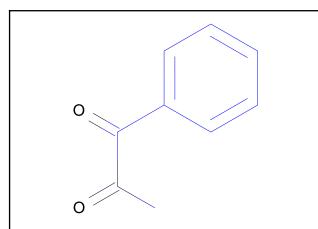
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- All properties and OPS components are within expected ranges. 1.
- Unknown ECFP 6 feature: 2023785560: [*]C([*])O 2.
- Unknown ECFP_6 feature: -222940837: [*]C([*])[C@H](O)[c](:[*]):[*] 3.
- 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]:[*] Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*] 7.
- Unknown ECFP 6 feature: 864287155: [*]NC 8.

Feature Contribution							
Top features for positive contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score				
	•	•	•				

FCFP_6	3	AND Enantiomer	0.0924
		OH H	
		[*]O	
FCFP_6	-2093839777	AND Enantiomer	0.078
		[*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	
FCFP_6	-885550502	AND Enantiomer OH H	0.0684
		[*]C([*])NC	
	Top Features f	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer OH H	-0.134
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	
ECFP_6	1564392544	AND Enantiomer I N OH H	-0.133
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	

FCFP_6	-1698724694	AND Enantiomer	-0.0944
		Р Н 2 ОН	
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	BENZYL ACETATE	BENZALDEHYDE	ALPHA-METHYLBENZYL ALCOHOL		
Structure		0	OH		
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		

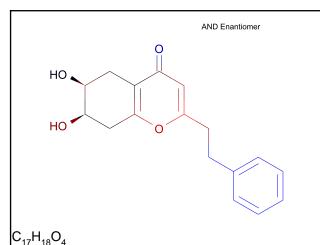
Model Applicability

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]:1	0.078				

ECFP_6	2099970318	[*]C(=O)[*]	0.0766
ECFP_6	642810091	[*][c](:[*]):[*]	0.0424
	Top Features f	or negative contribution	on
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		-0.102



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	NALTREXONE.HCL	NALIDIXIC ACID	CARBOXIN	
Structure	HO MAN NON NON NON NON NON NON NON NON NON N	HOO	HN	
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

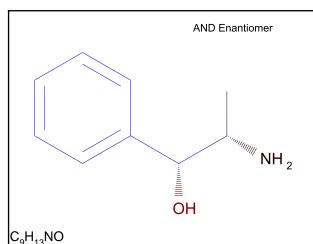
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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	AND Enantiomer	0.16
		но	
		но	
		[*]C(=CC(=[*])[*])[*]	
ECFP_6	-167460056	AND Enantiomer	0.136
		HO	
		[*]C([*])[*]	
FCFP_6	-1143715940	AND Enantiomer	0.13
		HO	
		но	
		[*]C(=[*])OC(=[*])[*]	
		gative contribution	
Fingerprint		Feature Structure	Score
FCFP_6	991735244	AND Enantiomer	-0.134
		HO	
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	
ECFP_6	1564392544	AND Enantiomer	-0.133
		HO HO	
		[*][c]1:[*]:[cH]:[cH]	
		:[cH]:[cH]:1	

ECFP_6	2106656448	AND Enantiomer	-0.11
		HO HO HO	
		[*]C(=O)[*]	



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	EPHEDRINE SULPHATE	ACETAMINOPHEN	TOCAINIDE.HCL		
Structure	OH N H	HO	NH NH 2		
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

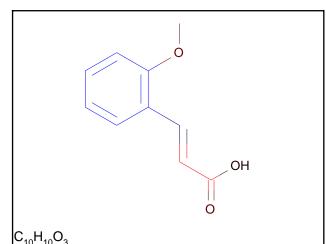
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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]:[*]

Top features for positive contribution				
Fingerprint Bit/Smiles Feature Structure Score				

FCFP_6	3	AND Enantiomer NH ₂ OH	0.0924
FCFP_6	-2093839777	[*]O AND Enantiomer NH 2 OH [*][c]1:[cH]:[cH]:[cH	0.078
ECFP_6	1572579716]:[cH]:[cH]:1 AND Enantiomer ., NH ₂ OH [*]N	0.0576
	Top Features f	for negative contributior	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer OH [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134
ECFP_6	1564392544	AND Enantiomer ., NH 2 OH [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133

FCFP_6	-1698724694	AND Enantiomer	-0.0944
		ÖH NH₂	
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	11	2; 4-DICHLOROPHENOXY ACETIC ACID	EUGENOL	
Structure	CI MUNICIPAL OF OH	CI OH	OH NO	
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

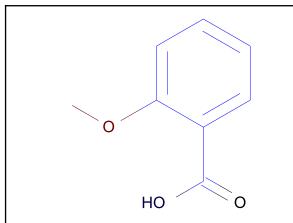
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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
		or negative contributio	n
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		-0.102
		ОН	
		[*]=O	



C₈H₈O₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	O-NITROANISOLE	O-ANISIDINE .HCL	P-CRESIDINE	
Structure	NO -	NH ₂	NH ₂	
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	

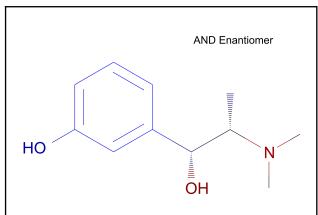
Model Applicability

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

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	2099970318	HOO	0.0766		
		[*]C(=O)[*]			

FCFP_6	1036089772	НОО	0.073
FCFP_6	136627117	[*]:[c](:[*])OC	0.0538
		HO O	
	Top Features f	for negative contributio	n
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



 $C_{11}H_{17}NO_2$

818

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	EPHEDRINE SULPHATE	PHENYLEPHRINE .HCL	TOCAINIDE.HCL	
Structure	OH N H	HO	NH NH 2	
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

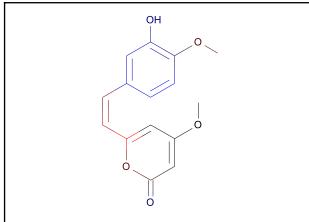
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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

i catale contribution				
Top features for positive contribution				
Bit/Smiles	Feature Structure	Score		
•	•			
	Top features	Top features for positive contribution	Top features for positive contribution	

FCFP_6 3 0.0924	
он î	
он î	
[*]O	
FCFP_6 9 0.0797	
HO HO	
OH OH	
[*]N([*])[*]	
ECFP_6 2007300961 AND Enantiomer 0.0564	
HO' OH N	
[*][c]1:[*]:[c]([*]):	
[cH]:[cH]:1	
Top Features for negative contribution	
Fingerprint Bit/Smiles Feature Structure Score	
FCFP_6 991735244 -0.134	
HO HO	
o _H ï	
[#3[-3]4_[#3],[-1]3,[-1]3	
[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	
FCEP 6 L453677277 L0 0906	
AND Enantiomer CO.0300	
HO OH	
[*]C([*])[c]1:[cH]:[*	
]:[cH]:[cH]:1	

136597326	AND Enantiomer	-0.0815
	HO N	
	136597326	AND ENGINORIES



C₁₅H₁₄O₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	NALIDIXIC ACID	11	DIMETHYL PHTHALATE	
Structure	HO Z	CIMPO	0	
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

Top features for positive contribution				
Fingerprint Bit/Smiles Feature Structure Score				

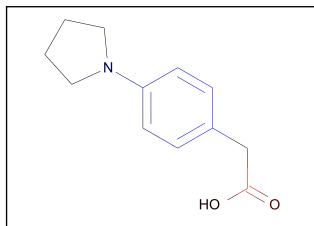
		[*]C(=CC(=[*])[*])[*]	
FCFP_6	-1143715940	[*]C(=[*])OC(=[*])[*]	0.13
ECFP_6	-1925046727	[*]C=[*]	0.0915
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	[*]C(=O)[*]	-0.11
FCFP_6	1	○H ○ ○ [*]=O	-0.102

0.16

FCFP_6

451847724

FCFP_6	-453677277	OH O	-0.0906
		[*]C([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	



 $|C_{12}H_{15}NO_{2}|$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	2; 4-DICHLOROPHENOXY ACETIC ACID	11	DICHLOROPROP
Structure	CI OH	CI no OH	CI NO OH
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.
- 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	HOOO	0.129
FCFP_6	-1176841573	[*]C[*]	0.0963
ECFP_6	2099970318	[*]C(=O)[*]	0.0766
	Top Features	for negative contribution	n
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	HOOO	-0.0713
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	

AND Enantiomer OH

 $C_{12}H_{13}NO_3$

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 nonnormal 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	O Nhy	OH OH OH	NH 2	
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

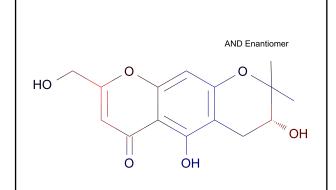
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- All properties and OPS components are within expected ranges. 1.
- Unknown ECFP 6 feature: -154530762: [*]N[*] 2.
- 3. Unknown ECFP 6 feature: -1694930393: [*]=C1[*][*]CN1
- Unknown ECFP_6 feature: -742538367: O=C1C[*][*]N1 4.
- 5. Unknown ECFP 6 feature: 53207596: [*]C([*])CC(=[*])[*] 6.
- Unknown ECFP_6 feature: -858846751: [*]C([*])C1C[*][*]C1
- Unknown ECFP_6 feature: -1457159889: [*][C@H]1[*][*]NC1 7.
- Unknown ECFP 6 feature: -1905455774: [*]C([*])C(=O)O 8.
- 9. Unknown ECFP_6 feature: 1603312431: [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]
- Unknown ECFP 6 feature: -176846085: [*]C([*])[c](:[cH]:[*]):[cH]:[*] 10.
- Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

Feature Contribution Top features for positive contribution Bit/Smiles Feature Structure Score Fingerprint

ECFP_6 -167460056		0.136
	○ OH OH	
	$\langle \cdot \rangle$	
	N—(H O	
	[*]C([*])[*]	
FCFP_6 -1176841573		0.0963
	,	
	и—(N—(
	[*]=CC(=O)O	
FCFP_6 3	AND Enantiomer	0.0924
	¢°	
	N——	
	[*]0	
Top Features for nega		
		Score
FCFP_6 991735244	_	-0.134
	OH OH	
	$\langle \cdot \rangle$	
	H O	
][c]1:[]:[cH]:[cH] :[cH]:[cH]:1	
ECFP_6 1564392544	AND Enantiomer	-0.133
	OH LO	
	$\langle \cdot \rangle$	
	N-K H O	
][c]1:[]:[cH]:[cH] :[cH]:[cH]:1	

ECFP_6	2106656448	AND Enantiomer	-0.11
		OH OH	
		N—N	
		[*]C(=O)[*]	



 $C_{15}H_{16}O_{6}$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

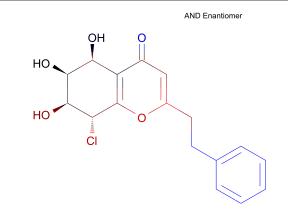
Structural Similar Compounds				
Name	NALTREXONE.HCL	PROPYL GALLATE	ZERANOL	
Structure	HO 11 NOH	HO WHO O	HO range of the control of the contr	
Actual Endpoint (-log C)	4.05615	2.84963	4.20737	
Predicted Endpoint (-log C)	4.77886	2.91297	3.4948	
Distance	0.612	0.619	0.646	
Reference	NDA-18932	NTP REPORT # 240	REGULAT. TOXICOL. PHARMACOL. 1983; 3: 9-	

Model Applicability

- 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 Cont	ribution		
	Top features	for positive contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	AND Enantomer HO O O OH O OH [*]C(=CC(=[*])[*])[*]	0.16
ECFP_6	-167460056	AND Enantiomer HO O OH O OH [*]C([*])[*]	0.136
ECFP_6	1559650422	AND Enantiomer HO OH OH [*]C[*]	0.129
	Top Features	for negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer HO OH OH [*]C(=O)[*]	-0.11

FCFP_6	1	AND Enantiomer HO OH OH [*]=O	-0.102
FCFP_6	136597326	AND Enantiomer HO OH OH [*]C([*])C	-0.0815



 $C_{17}H_{17}CIO_5$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	NALTREXONE.HCL	PROPYL GALLATE	ZEARALENONE	
Structure	HO IN THOM	HO NOT NOT NOT NOT NOT NOT NOT NOT NOT NO	HO the state of th	
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	

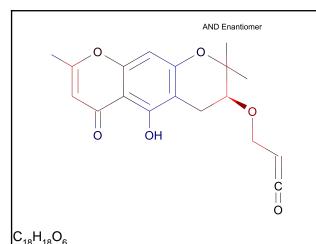
Model Applicability

- 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](CI)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	AND Enantiomer HO CI	0.16
ECFP_6	-167460056	[*]C(=CC(=[*])[*])[*] AND Enantiomer HO CI	0.136
FCFP_6	-1143715940	[*]C([*])[*] AND Enantiomer HO CI [*]C(=[*])OC(=[*])[*]	0.13
	Top Features	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantlomer HO CI [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134
ECFP_6	1564392544	AND Enantiomer HO CI [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133

ECFP_6	2106656448	AND Enantiomer	-0.11
		HO OH OH	
		[*]C(=O)[*]	



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	NALTREXONE.HCL	ZEARALENONE	NALIDIXIC ACID	
Structure	HO 14 NOH	HO O	HOO	
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	

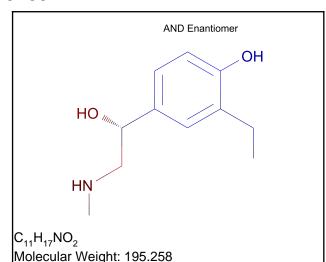
Model Applicability

- 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(=0)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=[*] Unknown ECFP_6 feature: -1250019913: [*]COC([*])[*] 22.

Feature Contribution						
reature Contr						
-		for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_6	451847724	AND Enantiomer OH C C C O (')C(=CC(=('))('))(')	0.16			
ECFP_6	-167460056	AND Enantiomer OH [*]C([*])[*]	0.136			
FCFP_6	-1143715940	AND Enantiomer OH C OH [*]C(=[*])OC(=[*])[*]	0.13			
		for negative contributio				
Fingerprint	Bit/Smiles	Feature Structure	Score			
ECFP_6	2106656448	AND Enantiomer OH CH CH CH CH CH CH CH CH CH	-0.11			
		[*]C(=O)[*]				

FCFP_6	1	AND Enantiomer OH C C OH C C O T O C O C O C O T O C O T	-0.102
FCFP_6	136597326	AND Enantiomer OH [*]C([*])C	-0.0815



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	PHENYLEPHRINE .HCL	ALBUTEROL.SULFATE	ALPHA- METHYLDOPASESQUITT YDRATE	
Structure	HO	HO THE OH	HO MH 2	
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

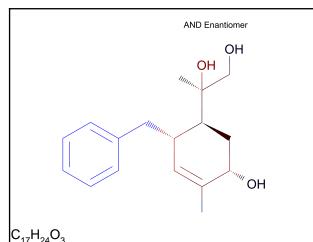
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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 OH HO, [*]C[*]	0.129
FCFP_6	3	AND Enantiomer OH HO, [*]O	0.0924
FCFP_6	-885550502	AND Enantiomer OH HO,	0.0684
		for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-453677277	AND Enantiomer HO	-0.0906
FCFP_6	136597326	AND Enantiomer OH HO, HN [*]C([*])C	-0.0815

FCFP_6	203677720	AND Enantiomer	-0.0713
		HOHOHOHOHOHOHOHO.	



17 1₂₄ 0₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	ALBUTEROL.SULFATE	PINDOLOL	PROPYL GALLATE	
Structure	HO MOH	HN OH H	HO NO O	
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	

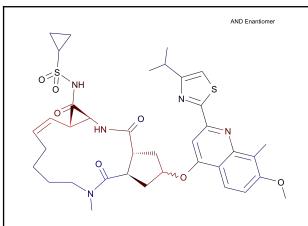
Model Applicability

- 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	AND Enantiomer	0.16
011_0	101011121	ОН ОН]	0.10
		,	
		. ОН	
		[*]C(=CC(=[*])[*])[*]	
ECFP_6	-167460056	AND Enantiomer	0.136
		OH OH	
		ОН	
		[*]C([*])[*]	
ECFP_6	1559650422	AND Enantiomer	0.129
		OH OH	
		, он	
		[*]C[*]	
	Top Features	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer	-0.134
		· OH OH	
		. ОН	
		F+1F 14 F+1 F 1 17 F 1 17	
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	
ECFP_6	1564392544	AND Enantiomer	-0.133
		OH OH	
		' '	
		[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	
		if an illiformity	

FCFP_6	-1698724694		-0.0944
		, OH OH	
		, OH	
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	



 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal 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	O Ray of the second of the sec		HN OH OWN	
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

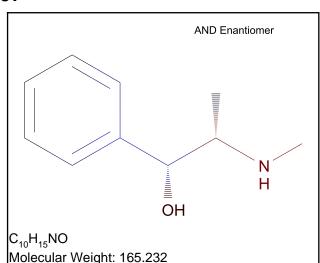
- 1. OPS PC1 out of range. Value: 18.64. Training min, max, SD, explained variance: -9.2986, 15.594, 5, 0.1094.
- 2. Unknown FCFP_2 feature: 1018942292: [*]NS(=O)(=O)C1[*][*]1
- Unknown ECFP_6 feature: -154530762: [*]N[*]
 Unknown ECFP 6 feature: 914325265: [*]:s:[*]
- 5. Unknown ECFP_6 feature: -797085356: [*]S(=[*])(=[*])[*]
- 6. Unknown ECFP_6 feature: -2095963820: [*][C@@H]1[*][*]C[C@H]1C(=[*])[*]
- 7. Unknown ECFP_6 feature: -867777309: [*]NC(=0)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	AND Enantiomer AND Enantiomer N S N S N S N S N S N S N S N	0.16		
ECFP_6	-167460056	AND Enantiomer AND Enantiomer N S N C ([*])[*]	0.136		

FCFP_6	3	AND Enantlomer O.0924 I*]O		
Top Features for negative contribution				

	Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	2106656448	AND Enantiomer AND Enantiomer Nus Nus Nus Nus Nus Nus Nus Nu	-0.11		
FCFP_6	1	AND Enantiomer AND Enantiomer NS NS NS NS NS NS NS NS NS N	-0.102		
FCFP_6	136597326	AND Enantiomer AND Enantiomer NS NS N I*]C([*])C	-0.0815		



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

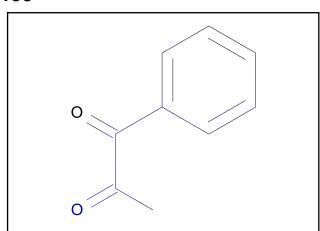
Structural Similar Compounds					
Name	EPHEDRINE SULFATE	DL-AMPHETAMINE SULFATE	PRIMIDONE (PRIMACLONE)		
Structure	N H	NH 2	HN N H		
Actual Endpoint (-log C)	4.16695	4.43204	3.94103		
Predicted Endpoint (-log C)	3.11455	3.18009	3.24349		
Distance	0.000	0.355	0.474		
Reference	NCI/NTP TR-307	NCI/NTP TR-387	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.

Feature Contribution					
	Top features	for positive contribution	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	-885550502	AND Enantiomer	0.115		
		[*]C([*])NC			

FCFP_2	3	AND Enantiomer	0.0737
		E	
		он н [*]О	
	Top Features for ne	gative contribution	
Fingerprint			Score
FCFP_2	203677720	AND Enantiomer	-0.0829
FCFP_2	16	[*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0512
	10	[*][c](:[*]):[*]	
FCFP_2	0	AND Enantiomer OH [*]C([*])[*]	-0.0314



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

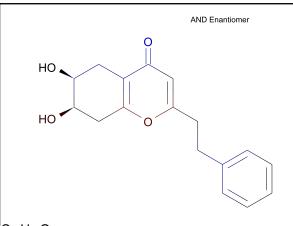
Structural Similar Compounds					
Name	BENZYL ACETATE	1-PHENYL-3-METHYL-5- PYRAZOLONE	PHTHALIC ANHYDRIDE		
Structure		O N N			
Actual Endpoint (-log C)	2.46903	2.88887	2.3413		
Predicted Endpoint (-log C)	3.22857	3.45184	3.37248		
Distance	0.296	0.348	0.371		
Reference	NCI/NTP TR-431	NCI/NTP TR-141	NCI/NTP TR-159		

Model Applicability

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

Feature Cont	Feature Contribution					
	Top Features	for negative contribution	on			
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_2	1872154524	[*]C(=O)[*]	-0.105			
		[*]C(=O)[*]		_		

FCFP_2	203677720	[*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829
FCFP_2	1	[*]=O	-0.0796



C₁₇H₁₈O₄

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	TOLAZAMIDE	CHLORPROPAMIDE	TOLBUTAMIDE		
Structure	T Z T T T T T T T T T T T T T T T T T T		O II S N H		
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.

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_2	-1143715940	AND Enentiomer HO HO [*]C(=[*])OC(=[*])[*]	0.095			

FCFP_2	3	AND Enantiomer HO HO [*]O	0.0737
	Top Features	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantiomer HO HO (*]C([*])CC(=[*])[*]	-0.111
FCFP_2	1872154524	AND Enantiomer	-0.105

FCFP_2

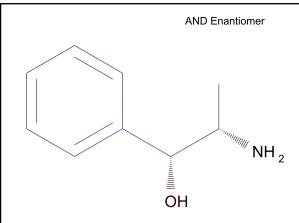
203677720

[*]C(=O)[*]

[*]C([*])[c](:[cH]:[*]):[cH]:[*]

AND Enantiomer

-0.0829



C₉H₁₃NO

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

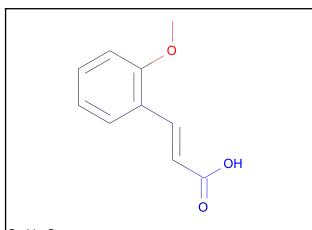
Structural Similar Compounds					
Name	EPHEDRINE SULFATE	DL-AMPHETAMINE SULFATE	ACETAMINOPHEN (4- HYDROXYACETANILIDE)		
Structure	OH N H	NH 2	HONH		
Actual Endpoint (-log C)	4.16695	4.43204	2.74809		
Predicted Endpoint (-log C)	3.11455	3.18009	2.67552		
Distance	0.229	0.326	0.417		
Reference	NCI/NTP 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.

Feature Contribution					
	Top features	for positive contribution	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	3	AND Enantiomer NH ₂ OH [*]O	0.0737		
Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		

FCFP_2	203677720	AND Enantiomer	-0.0829
		·,NH2	
		ŌН	
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	
FCFP_2	16	AND Enantiomer	-0.0512
		NH ₂	
		- ОН	
		[*][c](:[*]):[*]	
FCFP_2	þ	AND Enantiomer	-0.0314
		· NH ₂	
		он	
		[*]C([*])[*]	



 $C_{10}H_{10}O_3$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

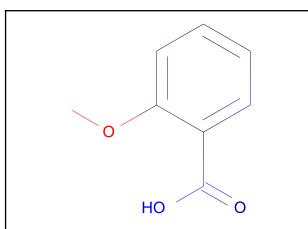
Structural Similar Compounds				
Name	EUGENOL	O-NITROANISOLE	MEXACARBATE	
Structure	OH OH	O NO	NH NH	
Actual Endpoint (-log C)	2.78402	3.18813	4.07253	
Predicted Endpoint (-log C)	2.98617	3.55234	3.7152	
Distance	0.369	0.369	0.401	
Reference	NCI/NTP TR-223	NCI/NTP TR-416	NCI/NTP TR-147	

Model Applicability

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

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
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	,	[*]O	-0.214
FCFP_2	-548632217	[*]C(=[*])O	-0.119
FCFP_2	1872154524	[*]C(=O)[*]	-0.105



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

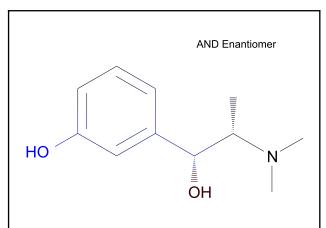
Structural Similar Compounds				
Name	O-NITROANISOLE	P-CRESIDINE	P-NITROBENZOIC ACID	
Structure	NO O	NH ₂	O OH	
Actual Endpoint (-log C)	3.18813	2.48408	2.90081	
Predicted Endpoint (-log C)	3.55234	3.53389	2.77599	
Distance	0.309	0.366	0.410	
Reference	NCI/NTP TR-416	NCI/NTP TR-142	NCI/NTP TR-442	

Model Applicability

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

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_2	136627117	I,JOC	0.173			
[*]OC						

FCFP_2	1036089772	но	0.0749
FCFP_2	332760439	[*]:[c](:[*])OC HO	0.0611
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	HO O	-0.214
FCFP_2	-548632217	HO O	-0.119
FCFP_2	1872154524	HO (*)	-0.105



C₁₁H₁₇NO₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	EPHEDRINE SULFATE	t-BUTYLHYDROQUINONE	ACETAMINOPHEN (4- HYDROXYACETANILIDE)		
Structure	N H	HO	HONH		
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		

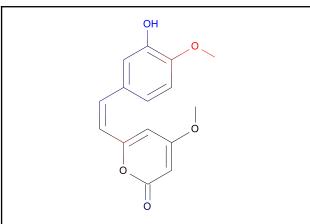
Model Applicability

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

Feature Cont	Feature Contribution					
	Top features	for positive contribution	า			
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_2	3	AND Enantiomer HO OH [*]O	0.0737			
	Top Features for negative contribution					

Top Features for negative contribution					
Fingerprint Bit/Smiles Feature Structure Score					

FCFP_2	7	AND Enantiomer HO OH [*]O	-0.214
FCFP_2	-549108873	AND Enantlomer OH [*]:[c](:[*])O	-0.127
FCFP_2	203677720	AND Enantiomer : OH [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



C₁₅H₁₄O₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

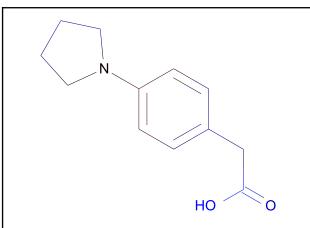
Structural Similar Compounds					
Name	FENTHION	DIMETHYL TEREPHTHALATE	3-NITRO-P- ACETOPHENETIDE		
Structure	S _M		O No -		
Actual Endpoint (-log C)	5.4903	2.93604	3.11767		
Predicted Endpoint (-log C)	4.18234	3.50861	3.42808		
Distance	0.486	0.505	0.527		
Reference	NCI/NTP TR-103	NCI/NTP TR-121	NCI/NTP TR-133		

Model Applicability

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

Feature Contribution						
	Top features	for positive contributio	n			
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_2	136627117	[*]OC	0.173			

FCFP_2	-1143715940	OH O	0.095
FCFP_2	1036089772	[*]:[c](:[*])OC	0.0749
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	/	[*]0	-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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

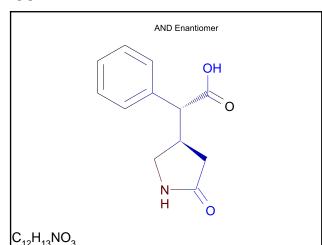
Name	MEXACARBATE	METHYLPHENIDATE HYDROCHLORIDE	EUGENOL
Structure	NH NH	N H O	OH OH
Actual Endpoint (-log C)	4.07253	3.66897	2.78402
Predicted Endpoint (-log C)	3.7152	3.4956	2.98617
Distance	0.383	0.419	0.431
Reference	NCI/NTP TR-147	NCI/NTP TR-439	NCI/NTP TR-223

Model Applicability

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

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	N HO O	-0.214
FCFP_2	-548632217	HO O	-0.119
FCFP_2	-1272798659	HOOO [*]C([*])CC(=[*])[*]	-0.111



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

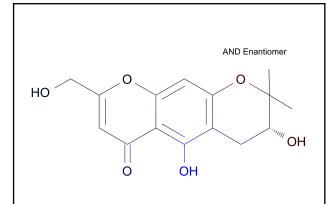
Structural Similar Compounds					
Name	PRIMIDONE (PRIMACLONE)	3-AMINO-4- ETHOXYACETANILIDE	O-ANTHRANILIC ACID		
Structure	HN N H	O Number of the North American North	O OH		
Actual Endpoint (-log C)	3.94103	2.45902	2.00682		
Predicted Endpoint (-log C)	3.24349	3.20661	2.55301		
Distance	0.358	0.392	0.442		
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.

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_2	-885550502	AND Enantiomer OH NH O [*]C([*])NC	0.115			

FCFP_2	3	AND Enantiomer OH N	0.0737
		[*]O	
	Top Features f	or negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	AND Enantiomer OH N N H O [*]O	-0.214
FCFP_2	-548632217	AND Enantiomer OH N N H O [*]C(=[*])O	-0.119
FCFP_2	-1272798659	AND Enantiomer OH N N H O [*]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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	PROPYL GALLATE	METHYLDOPA SESQUIHYDRATE	ACETOHEXAMIDE		
Structure	HO NO O	OH HO MAN HO	NH ONH		
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.

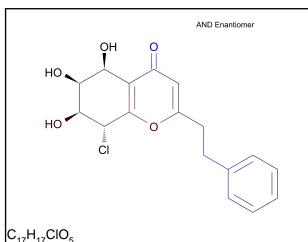
Feature Contribution					
	Top features	for positive contribution	า		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	1036089772	AND Enantiomer HO OH OH [*]:[c](:[*])OC	0.0749		

FCFP_2	3	AND Enantiomer HO OH OH [*]O	0.0737
FCFP_2	332760439	AND Enantiomer HO OH OH [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.0611
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	AND Enantiomer HO OH OH [*]O	-0.214
FCFP_2	-549108873	AND Enantiomer HO OH OH [*]:[c](:[*])O	-0.127
FCFP_2	1872154524	AND Enantiomer HO OH OH [*]C(=O)[*]	-0.105

F

┖

ᆮ



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

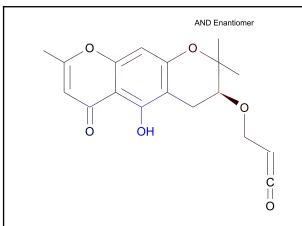
Structural Similar Compounds					
Name	ACETOHEXAMIDE	TOLAZAMIDE	PROPYL GALLATE		
Structure	NH O S S	T T T T T T T T T T T T T T T T T T T	HO W		
Actual Endpoint (-log C)	2.55683	2.84011	2.59435		
Predicted Endpoint (-log C)	3.62413	3.59315	2.18569		
Distance	0.513	0.538	0.559		
Reference	NCI/NTP TR-050	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.

	Feature Contribution					
Top features	for positive contribution	า				
Bit/Smiles	Feature Structure	Score				
-1143715940	AND Enantiomer OH OH OCI [*]C(=[*])OC(=[*])[*]	0.095				
	Bit/Smiles	Bit/Smiles Feature Structure -1143715940 AND Enantiomer	-1143715940 AND Enantiomer 0.095			

	3	HO : CI [*]O	0.0737
FCFP_2	71953198	AND Enantiomer HO CI [*]C([*])CI	0.058
	Top Features for ne	gative contribution	
	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantiomer HO OH O	-0.111
FCFP_2	1872154524	AND Enantiomer HO CI [*]C(=O)[*]	-0.105
FCFP_2	203677720	AND Enantiomer HO CI [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



C₁₈H₁₈O₆

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

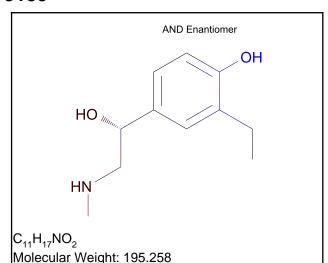
Structural Similar Compounds					
Name	TOLAZAMIDE	PROPYL GALLATE	ACETOHEXAMIDE		
Structure	H N H N N N N N N N N N N N N N N N N N	HO NO O	NH ONH ONH		
Actual Endpoint (-log C)	2.84011	2.59435	2.55683		
Predicted Endpoint (-log C)	3.59315	2.18569	3.62413		
Distance	0.476	0.510	0.525		
Reference	NCI/NTP TR-051	NCI/NTP TR-240	NCI/NTP TR-050		

Model Applicability

- 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	AND Enantiomer O OH C C O (*)C(=[*])OC(=[*])[*]	0.095			
	ļ	<u> </u>	<u>ļ</u>			

FCFP_2	1036089772	AND Enantiomer OH C C O [*]:[c](:[*])OC	0.0749
FCFP_2	332760439	AND Enantiomer OH [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.0611
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	/	AND Enantiomer OH C C O	-0.214
FCFP_2	-549108873	AND Enantiomer OH C C O [*]:[c](:[*])O	-0.127
FCFP_2	1872154524	AND Enantiomer OH C C O [*]C(=O)[*]	-0.105



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

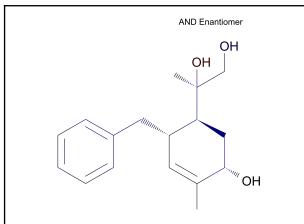
Name	EPHEDRINE SULFATE	t-BUTYLHYDROQUINONE	PRIMIDONE (PRIMACLONE)
Structure	OH N H	HO	HN N H
Actual Endpoint (-log C)	4.16695	2.44253	3.94103
Predicted Endpoint (-log C)	3.11455	2.62758	3.24349
Distance	0.414	0.521	0.546
Reference	NCI/NTP TR-307	NCI/NTP TR-459	NCI/NTP TR-476

Model Applicability

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

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	-885550502	AND Enantiomer OH HO. [*]C([*])NC	0.115	

FCFP_2	3		0.0737
		OH	
		HO,	
		J	
		HN	
		•	
		[*]O	
	Top Features for ne		
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	AND Enantiomer	-0.214
		OH	
		но,	
		HN	
		[*]O	
FCFP_2	-549108873	AND Enantiomer	-0.127
		HO,.	
		HN [*]	
		[*]:[c](:[*])O	
FCFP_2	203677720		-0.0829
_		ОН	
		но .	
		HN,	
		[*]C([*])[c](:[cH]:[*	
]):[cH]:[*]	



 $|C_{17}H_{24}O_3|$

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	O S N H	CI N H	HN	
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.

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	3	AND Enantiomer OH OH OH (*]O	0.0737	
	Top Features	for negative contribution	on	
Fingerprint	Bit/Smiles	Feature Structure	Score	

FCFP_2	-1272798659	AND Enantiomer OH OH OH (S) OH (S	-0.111
FCFP_2	203677720	AND Enentiomer OH OH (C) OH (C) (C) (C) (C) (C) (C) (C) (C	-0.0829
FCFP_2	16	AND Enantiomer OH OH (*][c](:[*]):[*]	-0.0512

AND Enantiomer	
O NH N S N	
N O	

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	C.I.PIGMENT RED 23	RESERPINE	SALICYLAZOSULFAPYRI DINE	
Structure	OF THE STATE OF TH	O The Control of the	N N N N N N N N N N N N N N N N N N N	
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	

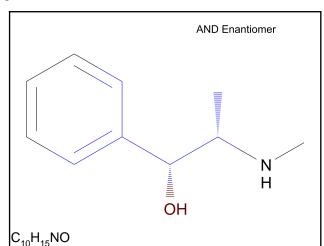
Model Applicability

- 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

Top features for positive contribution			
Feature Structure	Score		
	•		

FCFP_2	136627117	AND Enantiomer AND Enantiomer NS N S N S N S N S N S N S N S N S N	0.173
FCFP_2	-885550502	AND Enantiomer AND Enantiomer NS N O O O O O O O O O O O O O O O O O	0.115
FCFP_2	1036089772	AND Enantiomer AND Enantiomer NS N S N (*):[c](:[*])OC	0.0749
	Top Features for	or negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantiomer AND Enantiomer N S N O O O O O O O O O O O O O O O O O	-0.111
FCFP_2	1872154524	AND Enantiomer NS NS N (*]C(=O)[*]	-0.105

FCFP_2	203677720	AND Enantiomer	-0.0829
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	



ALogP: 1.234 Rotatable Bonds: 3

Molecular Weight: 165.232

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccuratė.

Structural Similar Compounds					
Name	ALPHA-METHYLBENZYL ALCOHOL	BENZYL ALCOHOL	M-CRESIDINE		
Structure	OH	ОН	H ₂ N ₁		
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		

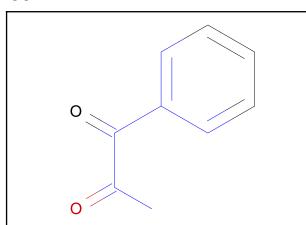
Model Applicability

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

- All properties and OPS components are within expected ranges. 1.
- Unknown FCFP 2 feature: 136686699: [*]NC

Feature Cont	ribution				
	Top features	for positive contribution	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	3	AND Enantiomer OH H	0.104		
		[*]0			
	Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score		

FCFP_2	136597326	AND Enantiomer OH [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantiomer	-0.406
FCFP_2	0	AND Enantiomer OH [*]C([*])[*]	-0.29



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

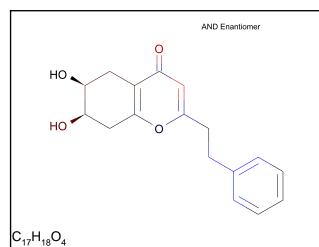
Structural Similar Compounds					
Name	BETA-NITROSTYRENE	BENZALDEHYDE	COUMARIN		
Structure	N I O	0			
Actual Endpoint (-log C)	2.6965	2.38236	3.76684		
Predicted Endpoint (-log C)	2.69255	2.25838	3.6624		
Distance	0.348	0.400	0.444		
Reference	NCI/NTP TR-170	NCI/NTP TR-378	NCI/NTP TR-422		

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1549192822: [*]C(=[*])C(=O)[c](:[*]):[*]

Feature Cont	Feature Contribution					
	Top features	for positive contribution	n			
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



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

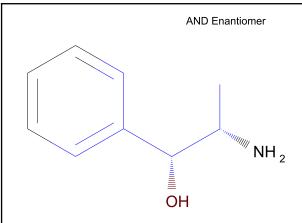
Structural Similar Compounds					
Name	PROBENECID	2-AMINO-5- NITROPHENOL	2-AMINO-4- NITROPHENOL		
Structure	OH O	O N NH 2 OH	O N NH 2 OH		
Actual Endpoint (-log C)	2.85333	2.88684	2.78993		
Predicted Endpoint (-log C)	2.4258	3.28491	3.28491		
Distance	0.714	0.734	0.734		
Reference	NCI/NTP TR-395	NCI/NTP TR-334	NCI/NTP TR-339		

Model Applicability

- 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	า		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	1	AND Enantiomer HO HO [*]=O	0.511		

FCFP_2	451847724	AND Enantiomer HO HO [*]C(=CC(=[*))(*))(*)	0.225
FCFP_2	3	AND Enantiomer HO HO [*]O	0.104
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	HO	-0.406
FCFP_2	1872154524	AND Enantiomer HO HO C (*]C(=O)[*]	-0.307
FCFP_2	0	AND Enantiomer HO HO ([*]C([*])[*]	-0.29



C₉H₁₃NO

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

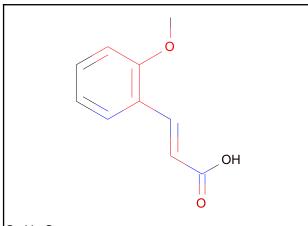
Structural Similar Compounds					
Name	ALPHA-METHYLBENZYL ALCOHOL	BENZYL ALCOHOL	HYDROQUINONE		
Structure	OH	OH	НО		
Actual Endpoint (-log C)	2.21189	2.43193	3.34286		
Predicted Endpoint (-log C)	1.71063	2.09547	3.36172		
Distance	0.405	0.548	0.593		
Reference	NCI/NTP TR-369	NCI/NTP TR-343	NCI/NTP TR-366		

Model Applicability

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

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	3	AND Enantiomer NH ₂ OH	0.104	
	Top Features	for negative contribution	on	
Fingerprint	Bit/Smiles	Feature Structure	Score	

FCFP_2	136597326	AND Enantiomer NH ₂ OH [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantiomer OH [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	0	AND Enantiomer NH ₂ OH [*]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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	M-CRESIDINE	BETA-NITROSTYRENE	COUMARIN
Structure	H ₂ N ₁₁		
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

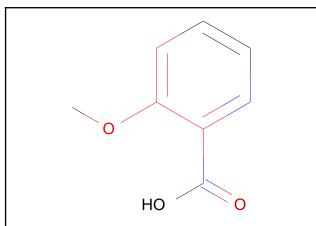
Model Applicability

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	Г*]=О	0.511
FCFP_2	451847724	[*]C(=CC(=[*])[*])[*]	0.225
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score -0.406
FCFP_2	203677720	[*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.400
FCFP_2	1872154524	[*]C(=O)[*]	-0.307
FCFP_2	0	[*]C([*])[*]	-0.29



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

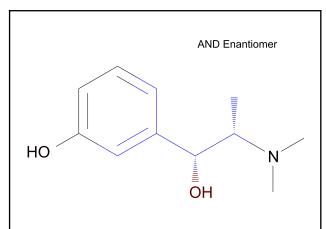
Structural Similar Compounds				
Name	M-CRESIDINE	BETA-NITROSTYRENE	COUMARIN	
Structure	H ₂ N ₁	N I O		
Actual Endpoint (-log C)	5.93318	2.6965	3.76684	
Predicted Endpoint (-log C)	3.87056	2.69255	3.6624	
Distance	0.397	0.543	0.552	
Reference	NCI/NTP TR-105	NCI/NTP TR-170	NCI/NTP TR-422	

Model Applicability

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

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	HO [*]=0	0.511
FCFP_2	136627117	HO O	0.0304
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	[*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	HO O [*]C(=O)[*]	-0.307
FCFP_2	0	HO O	-0.29



C₁₁H₁₇NO₂

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

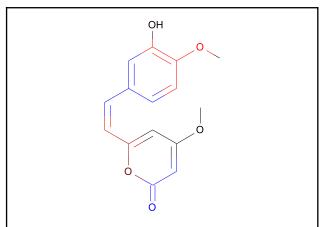
Name	4-HEXYLRESORCINOL	HYDROQUINONE	RESORCINOL
Structure	OH OH	НО	НО
Actual Endpoint (-log C)	3.1915	3.34286	2.99262
Predicted Endpoint (-log C)	2.16134	3.36172	3.36172
Distance	0.592	0.617	0.617
Reference	NCI/NTP TR-330	NCI/NTP TR-366	NCI/NTP TR-403

Model Applicability

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

Feature Cont	Feature Contribution				
	Top features	for positive contributio	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	3	AND Enantiomer HO OH	0.104		
		[*]0			

FCFP_2	7	AND Enantiomer	0.0144
		HO N	
		он '	
		O[*]	
FCFP_2	74595001	AND Enantiomer	0.000246
		HO, OH	
		[*]:[cH]:[c](O):[cH]:	
		[*]	
	Top Features for ne		
		Feature Structure	Score
FCFP_2	136597326	AND Enantiomer	-0.489
		HO OH	
		OH	
		[*]C([*])C	
FCFP_2	203677720	AND Enantiomer	-0.406
		но ОН	
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	
]):[cH]:[*]	
FCFP_2	0	AND Enantiomer	-0.29
		HO OH	
		[*]C([*])[*]	



C₁₅H₁₄O₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

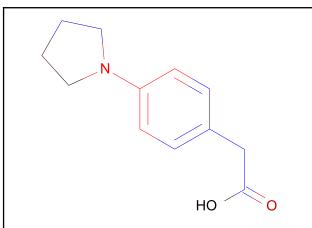
Structural Similar Compounds				
Name	PROBENECID	DIALLYL PHTHALATE	DIGLYCIDYL RESORCINOL ETHER (DGRE)	
Structure	OH OON N			
Actual Endpoint (-log C)	2.85333	3.3914	4.99464	
Predicted Endpoint (-log C)	2.4258	3.50093	4.37491	
Distance	0.642	0.674	0.688	
Reference	NCI/NTP TR-395	NCI/NTP TR-284	NCI/NTP TR-257	

Model Applicability

- 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			
		Wr 1)-r 1				

FCFP_2	1	OH O\ (*)=O	0.511
FCFP_2	451847724	[*]C(=CC(=[*])[*])[*]	0.225
		for negative contributio	
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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

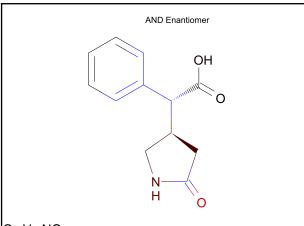
Structural Similar Compounds				
Name	3,4-DIHYDROCOUMARIN	M-CRESIDINE	BETA-NITROSTYRENE	
Structure	0 0	H ₂ N ₁	N = 0	
Actual Endpoint (-log C)	2.69361	5.93318	2.6965	
Predicted Endpoint (-log C)	3.51534	3.87056	2.69255	
Distance	0.567	0.580	0.606	
Reference	NCI/NTP TR-423	NCI/NTP TR-105	NCI/NTP TR-170	

Model Applicability

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

Feature Contribution				
	Top features	for positive contribution	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	332760439	[*]O[c](:[cH]:[*]):[c]([*]):[*]	0.672	

FCFP_2 FCFP_2	-1272798659	[*]=O	0.511
		[*]C([*])CC(=[*])[*]	
Fingerprint	Bit/Smiles	or negative contribution 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 nonnormal 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	O N NH ₂	O N NH 2	OH O S O O	
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	

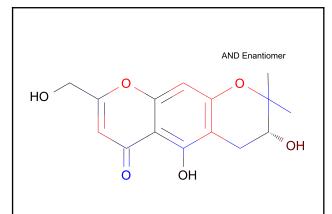
Model Applicability

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

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	1	AND Enantiomer OH NH O [*]=O	0.511	

FCFP_2	3	AND Enantiomer	0.104
		OH OH	
		$\langle \cdot \rangle$	
		H O	
		[*]O	
FCFP_2	-1272798659	AND Enantiomer	0.0703
		OH	
		N— H O	
		[*]C([*])CC(=[*])[*]	
	Top Features for ne	gative contribution	
			Score
FCFP_2	203677720	AND Enantiomer	-0.406
		OH OH	
		N-W _O	
		[*]C([*])[c](:[cH]:[*	
]):[cH]:[*]	
FCFP_2	1872154524	^	-0.307
		OH OH	
		人 。	
		N—N H	
		[*]C(=O)[*]	
FCFP_2	0	AND Enantiomer	-0.29
		OH OH	
		\rightarrow	
		N—— H O	
		[*]C([*])[*]	

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



C₁₅H₁₆O₆

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 nonnormal 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	OH HN MH 2	OH OH	O N NH 2 OH	
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

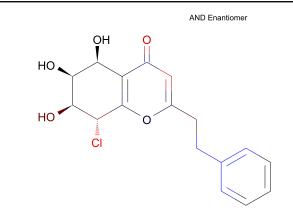
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: 436915834: [*]C\C(=C\[*])\O[*]
- 3. Unknown FCFP_2 feature: -1549192822: [*]C(=[*])C(=O)[c](:[*]):[*]

Feature Contribution				
Top features	for positive contribution	n		
Bit/Smiles	Feature Structure	Score		
332760439	AND Enantiomer HO OH OH [*]O[c](:[cH]:[*]):[c]([*]):[*]	0.672		
	Top features Bit/Smiles	Top features for positive contribution Bit/Smiles Feature Structure 332760439 AND Enantiomer HO OH OH OH OH OH	Top features for positive contribution Bit/Smiles Feature Structure Score 332760439 AND Enantioner OH [*]O[c](:[cH]:[*]):[c	

FCFP_2	1	AND Enantiomer HO OH OH [*]=O	0.511
FCFP_2	451847724	AND Enantiomer HO OH OH [*]C(=CC(=[*])[*])[*]	0.225
	Top Features for ne		
	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	HO OH OH	-0.489
FCFP_2	203677720	AND Enantiomer HO OH OH [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	AND Enantiomer HO OH OH [*]C(=O)[*]	-0.307

F

┖



C₁₇H₁₇CIO₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

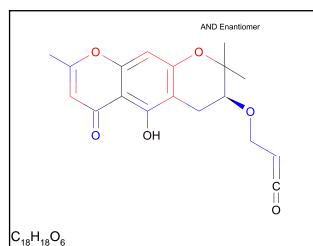
Structural Similar Compounds				
Name	HC RED 3	PENICILLIN VK	OCHRATOXIN	
Structure	OH HN NH O NH O	O NH O Tres	OH OH OH OH	
Actual Endpoint (-log C)	2.59592	2.54455	6.28396	
Predicted Endpoint (-log C)	3.285	3.9702	5.12358	
Distance	0.690	0.721	0.757	
Reference	NCI/NTP TR-281	NCI/NTP TR-336	NCI/NTP TR-358	

Model Applicability

- 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	AND Enantiomer HO OH O HO CI [*]CI	0.526		

FCFP_2	1	AND Enantiomer HO CI [*]=O	0.511
FCFP_2	451847724	AND Enantiomer HO	0.225
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	AND Enantiomer HO Ci [*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	AND Enantiomer HO CI [*]C(=O)[*]	-0.307
FCFP_2	0	AND Enantiomer HO CI [*]C([*])[*]	-0.29



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	O H	OH O S O O	O N NH 2 OH	
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	

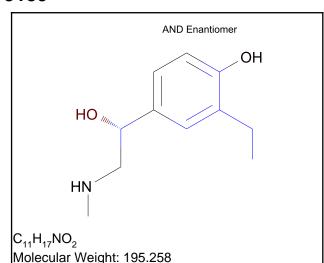
Model Applicability

Facture Cantribution

- 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 Cont	-eature Contribution				
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	332760439	AND Enantiomer OH [*]O[c](:[cH]:[*]):[c	0.672		
		[*]O[c](:[cH]:[*]):[c]([*]):[*]			

	1.		L 1
FCFP_2		AND Enantiomer OH CH OH	0.511
		[*]=O	
FCFP_2	451847724	AND Enantiomer OH C OH	0.225
		[*]C(=CC(=[*])[*])[*]	
	Top Features f	or negative contribution	ı
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantiomer OH C C O [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantomer OH C C (*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	AND Enantiomer OH CH CH CH CH CH CH CH CH CH	-0.307



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

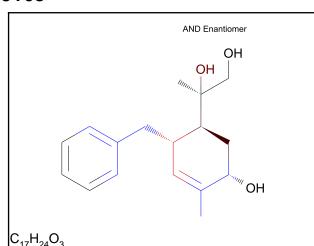
Name	4-HEXYLRESORCINOL	2,4-DIAMINOPHENOL DIHYDROCHLORIDE	RESORCINOL
Structure	OH OH	NH 2 NH 2 OH	НО
Actual Endpoint (-log C)	3.1915	3.69598	2.99262
Predicted Endpoint (-log C)	2.16134	3.37438	3.36172
Distance	0.600	0.717	0.735
Reference	NCI/NTP TR-330	NCI/NTP TR-401	NCI/NTP TR-403

Model Applicability

- 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	AND Enantiomer OH HN [*]O	0.104		
		1 11-			

FCFP_2	7	AND Enantiomer OH HO,	0.0144
FCFP_2	74595001	[*]O AND Enantiomer HO	0.000246
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	HO,	-0.489
FCFP_2	203677720	AND Enantiomer OH HO	-0.406
FCFP_2	0	AND Enantiomer OH HO,	-0.29



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	4-HEXYLRESORCINOL	HC RED 3	LITHOCHLOLIC ACID	
Structure	OH OH	OH HN MN NH 2	OH WATER TO THE T	
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

- 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 contributio	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	451847724	AND Enantiomer OH OH OH O(H) O(0.225		

FCFP_2	3	ОН	0.104
		·	
		ОН	
FCFP_2	-1272798659	[*]O AND Enantiomer	0.0703
-CFF_2	-1272790009	OH OH	0.0703
		. ОН	
		[*]C([*])CC(=[*])[*]	
	Top Features for ne		
ingerprint		Feature Structure	Score
FCFP_2	136597326	AND Enantiomer	-0.489
		· Р	
		У. ОН	
		[*]C([*])C	
FCFP_2	203677720	AND Enantiomer OH OH I	-0.406
		. OH	
		[*]C([*])[c](:[cH]:[*]):[cH]:[*]	
FCFP_2	0	AND Enantiomer	-0.29
		·. OH	
		ОН ОН	
		[*]C([*])[*]	

F

F

ᆮ

Simeprevir

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage

AND Enantiomer
O NH
N S
ĤN
N N N N N N N N N N N N N N N N N N N

 $C_{38}H_{47}N_5O_7S_2$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	OCHRATOXIN	PENICILLIN VK	AMPICILLIN TRIHYDRATE		
Structure	OH MANH HO WHAT HAVE I	O NH THE SOLUTION OH	H ₂ N ₄ , NH ONH OH OH		
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		

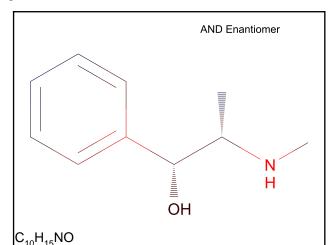
Model Applicability

- 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	AND Enantiomer	0.672
		[*]O[c](:[cH]:[*]):[c]([*]):[*]	
FCFP_2	1	AND Enantiomer	0.511
FCFP_2	451847724	[*]=O AND Enantiomer	0.225
	To Footon	[*]C(=CC(=[*1)[*])[*]	
Fingerprint	Bit/Smiles	for negative contribution Feature Structure	Score
FCFP_2	136597326	AND Enantiomer AND Enantiomer NS N S N O N O N O N O N O N O N O N O	-0.489
FCFP_2	203677720	AND Enantiomer AND Enantiomer (*]C([*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	565998553	AND Enantiomer AND Enantiomer	-0.348
		[*]OC(=O)C=[*]	



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 nonnormal 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	OH N H	NH NH	NH	
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

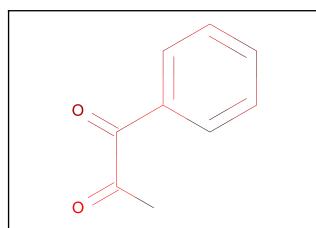
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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	AND Enantiomer	0.281	
		[*][c](:[*]):[*]		

ECFP_6	-1897341097	AND Enantiomer	0.216
		OH H	
		[*]N[*]	
ECFP_6	1571214559	AND Enantiomer	0.19
		 Р Н	
		[*]1:[cH]:[cH]:[cH]:[cH]:1	
Top Features for negative contribution			
Fingerprint		Feature Structure	Score
ECFP_6	2014710090	AND Enantiomer	-0.225
		OH H	
		[*10/[*1\[-14.[-1.1].[-	
		[*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:1	
ECFP_6	734603939	AND Enantiomer	-0.201
		S H	
		[*]C	
FCFP_6	3	AND Enantiomer	-0.107
		DH	
		[*]0	



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.

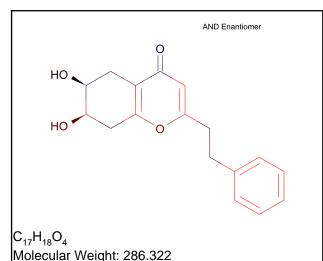
Structural Similar Compounds				
Name	ACETOPHENONE	METHYLBENZOATE	PHENYL ACETATE	
Structure	0	0	0	
Actual Endpoint (-log C)	2.169	2.063	1.922	
Predicted Endpoint (-log C)	2.12357	1.70515	2.03223	
Distance	0.334	0.363	0.390	
Reference	GTPZAB 26(8);53;82	FAVUAI 18;69;86	AIHAAP 30;470;69	

Model Applicability

- 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	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	642810091	[*][c](:[*]):[*]	0.281		

ECFP_6	1571214559	[*]1:[cH]:[cH]:[cH]:[cH]:1	0.19
FCFP_6	-1549192822	[*]C(=[*])C(=O)[o](:[*]):[*]	0.168
		egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	734603939	[*]C	-0.201
ECFP_6	1997021792	[*]:[cH]:[cH]:[*]	-0.0912
FCFP_6	0		-0.0791



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 nonnormal 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	HO m	N H N N N N N N N N N N N N N N N N N N	HN O	
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

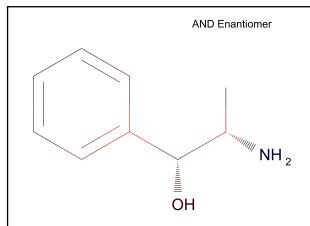
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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]:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	

ECFP_6	642810091	AND Enantiomer	0.281
		HO	
		но	
		[*][c](:[*]):[*]	
ECFP_6	-560785749	AND Enantiomer	0.259
		HO HO	
ECFP_6	1571214559	[*]C(=[*])OC(=[*])[*] AND Enantiomer	0.19
ECFP_0	137 12 14339	HO.	0.19
		HO O	
		[*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	
	Top Features for ne	egative contribution	
Fingerprint			Score
ECFP_6	2106656448	AND Enantiomer	-0.352
		HO	
		[*]C(=O)[*]	
ECFP_6	683445015	AND Enantiomer	-0.266
		HO	
		[*]0[*]	

ECFP_6	-1795525632	AND Enantiomer	-0.176
		HO HO	
		[*]CCC(=[*])[*]	



C₉H₁₃NO

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	3-AMINO-2- METHYLBENZYL ALCOHOL	EPHEDRINE	GLYCOLANILIDE	
Structure	H ₂ N OH	OH N H	NH	
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

- 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 Cont	Feature Contribution					
	Top features	for positive contributio	n			
Fingerprint	Bit/Smiles	Feature Structure	Score			
ECFP_6	642810091	AND Enantiomer NH ₂ OH [*][c](:[*]):[*]	0.281			

ECFP_6	1571214559	AND Enantiomer	0.19
		NH ₂	
		ŌН	
		[*]1:[cH]:[cH]:[
		cH]:[cH]:1	
FCFP_6	1070061035	AND Enantiomer	0.0996
		·, NH ₂	
		ОН	
		[*]C([*])O	
	Top Features f	or negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2014710090	AND Enantiomer	-0.225
		., _{NH2}	
		ŌН	
		[*]C([*])[o]1:[cH]:[c H]:[cH]:[cH]:[cH]:1	
ECFP_6	734603939	AND Enantiomer	-0.201
		., NH ₂	
		[*]C	
FCFP_6	3	AND Enantiomer	-0.107
		·, NH ₂	
		[*]0	

Molecular Weight: 178.185

ALogP: 1.911 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

 $|C_{10}H_{10}O_3|$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

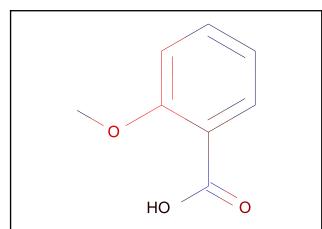
Structural Similar Compounds					
Name	CINNAMIC ACID; 2- ACETOXY-	p- METHOXYPHENYLACETI C ACID	ETHYL SALICYLATE		
Structure	O OH	OH OH	OH O		
Actual Endpoint (-log C)	1.816	2.03	2.1		
Predicted Endpoint (-log C)	2.18026	2.31007	1.7354		
Distance	0.290	0.325	0.337		
Reference	JPETAB 99;450;50	FCTXAV 14;685;76	FCTXAV 16;751;78		

Model Applicability

- 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 contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	642810091	[*][c](:[*]):[*]	0.281	

FCFP_6	136627117	[*]OC	0.17
ECFP_6	-1074141656	[*]=O	0.142
		or negative contribution	n
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



C₈H₈O₃

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

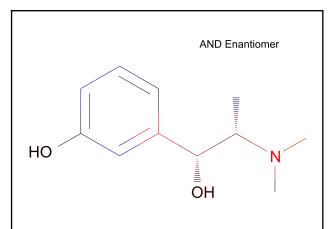
Structural Similar Compounds				
Name	METHYL SALICYLATE	VANILLIN	p- METHOXYPHENYLACETI C ACID	
Structure	OH O	O OH	OH OH	
Actual Endpoint (-log C)	2.234	1.984	2.03	
Predicted Endpoint (-log C)	1.90826	1.65955	2.31007	
Distance	0.226	0.278	0.321	
Reference	FCTXAV 2;327;64	FCTXAV 2;327;64	FCTXAV 14;685;76	

Model Applicability

- 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	HO O	0.17
ECFP_6	-1074141656	HO O	0.142
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	734603939	[*]C	-0.201
FCFP_6	1036089772	HO O	-0.136
ECFP_6	864909220	HOOO [*]OC	-0.119



C₁₁H₁₇NO₂

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 nonnormal 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	OH N H	OH OH	HONNH		
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

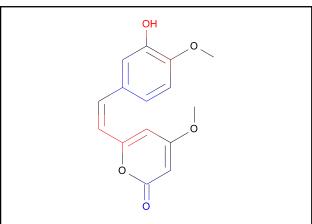
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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	AND Enantiomer	0.322
		но	
		[*]N([*])C	
ECFP_6	642810091	AND Enantiomer	0.281
		HO OH	
		[*][c](:[*]):[*]	
ECFP_6	2019062761	AND Enantiomer	0.138
		HO OH	
		[*]:[c](:[*])O	
	Top Features	for negative contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	946589555	AND Enantiomer	-0.204
		ОН	
		O[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	
ECFP_6	734603939	AND Enantiomer	-0.201
		HO OH I	
		[*]C	

FCFP_6	3	AND Enantiomer	-0.107
		HO OH :	
		[*]0	



C₁₅H₁₄O₅

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	O OH	O N H	N H	
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

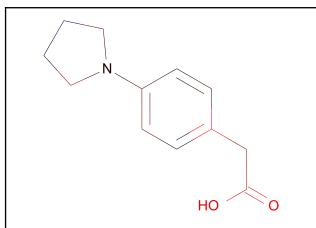
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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(=[,])[,] OH	0.259
FCFP_6	436915834	[*]C\C(=C\[*])\O[*]	0.184
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	[*]C(=O)[*]	-0.352
ECFP_6	683445015	[*]O[*]	-0.266

ECFP_6	2077607946	OH O	-0.252
		[*]O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	



 $C_{12}H_{15}NO_2$

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			PHENOXYACETIC ACID; p-CHLORO	
Structure	HO	O OH	CI	
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	

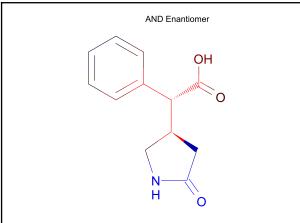
Model Applicability

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 0.281

FCFP_6	675769755	HOOO	0.155
ECFP_6	-1074141656	[*]:[c](:[*])N1C[*][*]C1	0.142
_			
		но [~] о [*]=О	
		for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	670515721	(*]N([*])[*]	-0.108
ECFP_6	2025485523	HO O HO O HO O HO O O HO O O O	-0.0801
FCFP_6	0	[*]C([*])[*]	-0.0791
		L 3 (L 3/L 3	



 $C_{12}H_{13}NO_3$

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds	
-------------------------------------	--

Name	PHENOBARBITAL	PRIMIDONE	SALACETAMIDE
Structure	HN O	HN O	OH O
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

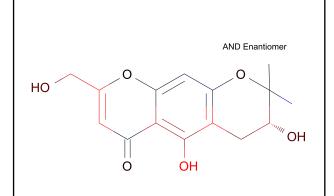
Model Applicability

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.
- $2. \qquad \text{Unknown ECFP_2 feature: 1603312431: } [*]C([*])[C@@H](C(=[*])[*])[c](:[*]):[*]$
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1186303932	AND Enantiomer OH NH O [*]C([*])[C@@H][C(=[*])(*)]/G[-([*)]:*]	0.375	

642810091	AND Enantlomer OH N N (*][c](:[*]):[*]	0.281
1571214559	AND Enantiomer OH OH N OH CH CH CH CH CH CH CH CH CH	0.19
	Feature Structure	Score
	OH N—O H O [*]C(=O)[*]	-0.352
2014710090	AND Enantomer OH NH O [*]C([*])[c]1:[cH]:[c H]:[cH]:[cH]:[1	-0.225
566058135	AND Enantiomer OH N O=C1C[*][*]N1	-0.216
	Top Features f Bit/Smiles 2106656448 2014710090	1571214559



C₁₅H₁₆O₆

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 nonnormal 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	HO O O O	OH N N N N N N N N N N N N N N N N N N N	HO NO O	
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

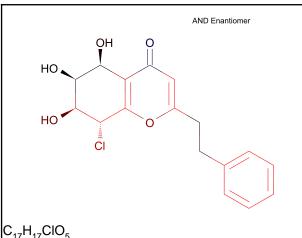
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
 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([*])[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	

ECFP_6	642810091		0.281
		AND Enantiomer	
		HOTOTO	
		O OH ,OH	
		[*][c](:[*]):[*]	
		[][-](-[])-[]	
ECFP_6	-560785749		0.259
		AND Enantiomer	
		но	
		ö óн	
		[*]C(=[*])OC(=[*])[*]	
FCFP_6	436915834		0.184
		AND Enantiomer	
		HO CONTRACTOR	
		0 он он	
		[*]C\C(=C\[*])\O[*]	
	Top Features for ne	egative contribution	
Fingerprint		Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer	-0.352
		HO~~°	
		OH .OH	
		O OH	
		[*]C(=O)[*]	
ECFP_6	683445015		-0.266
_		AND Enantiomer	
		HO TO	
		O OH .OH	
		[*]O[*]	

ECFP_6	734603939	-0.201
		AND Enantiomer HO OH OH
		[*]C



10₁₇11₁₇010₅

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	OXAMNIQUINE	CETOPHENICOL	PROPYL GALLATE		
Structure	OH Mu N N H H H H N	CI H OH N	HO W		
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		

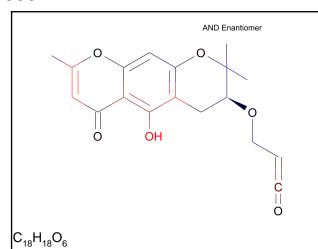
Model Applicability

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](CI)C(=[*])[*]
- 5. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score FCFP_6 71953198 0.392

ECFP_6	642810091	AND Enantiomer HO OH O	0.281
ECFP_6	-560785749	AND Enantiomer HO	0.259
		or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer HO CI [*]C(=O)[*]	-0.352
ECFP_6	683445015	AND Enantiomer HO CI [*]O[*]	-0.266
ECFP_6	-817402818	AND Enantiomer HO CI [*]CI	-0.263



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	N- [[(HEXAHYDROCYCLOPE NTA[c]PYRROL-2(1H)- YL)AMINO]CARBONYL]-4- METHYL- BENZENESULFONAMIDE	PHOSPHORIC ACID; 2- CHLORO-1-(2;4- DICHLOROPHENYL)VINY L ESTER	MYCOPHENOLIC ACID	
Structure	NH ONH ON NH	CI OH OH	HO PARTY OF THE OPEN CONTRACTOR OF THE OPEN C	
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	

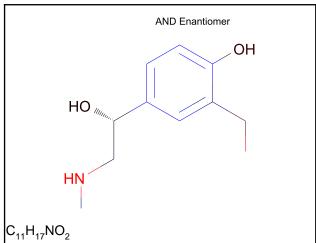
Model Applicability

- 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	AND Enantiomer O OH C	0.281
ECFP_6	-560785749	[*][c](:[*]):[*] AND Enantioner	0.259
2011_0		OH C:O	0.200
FCFP_6	436915834	[*]C(=[*])OC(=[*])[*] AND Enantiomer O OH C C C	0.184
		[*]C\C(=C\[*])\O[*]	
Fingerprint		egative contribution Feature Structure	Score
ECFP_6	2106656448	AND Enantiomer	-0.352
2011_0	2100000110	OH CE	0.002
ECFP_6	683445015	[*]C(=O)[*] AND Enantiomer O OH C C C C C	-0.266
		[*]0[*]	

ECFP_6	734603939	AND Enantiomer	-0.201
		O OH	
		Ö	
		[*]C	



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	ETILEFRINE	PHENYLEPHRINE	BENZYL ALCOHOL; 3;4- DIHYDROXYalpha [(METHYLAMINO)METHYL]-; .HCI (HCI STRIPPED)		
Structure	HO	HO NH	HO MH		
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

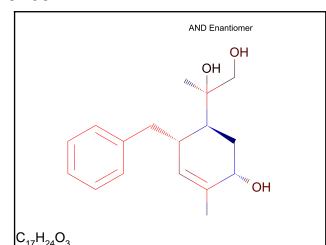
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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	0.281
		OH	
		но,	
		HN	
		[*][c](:[*]):[*]	
ECFP_6	-1897341097	AND Enantiomer	0.216
		OH	
		HO _r .	
		НЙ	
		[*]N[*]	
ECFP_6	864287155	AND Enantiomer	0.188
		HO,	
		HŅ	
		I I	
		[*]NC	
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	946589555	AND Enantiomer	-0.204
		но .	
		HN I	
		O[c]1:[cH]:[*]:[cH]:[
		cH]:[cH]:1	
ECFP_6	734603939	AND Enantiomer	-0.201
		HO,.	
		HN	
		[*]C	

FCFP_6	-451251206	AND Enantiomer OH	-0.143



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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	CARTEOLOL .HCI (HCI STRIPPED)	2-PHENYL-1;2- BUTANEDIOL-1- CARBAMATE	AMBROXOL		
Structure	HN OH H	OH H ₂ N	Br s NH 2		
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

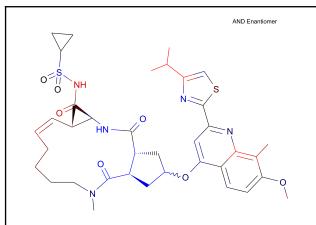
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 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]:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	642810091	AND Enantlomer OH OH OH (*][c](:[*]):[*]	0.281		

ECFP_6	1571214559	AND Enantiomer OH	0.19
		ОН.	
		[*]1:[cH]:[cH]:[cH]:[cH]:1	
ECFP_6	1095683433	AND Enantiomer OH OH	0.123
		ОН.	
		[*]C[c]1:[cH]:[*]:[cH]:[cH]:[dH]:1	
	Top Features f	or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	734603939	AND Enantiomer OH OH	-0.201
		OH OH	
ECFP_6	-801490360	[*]C AND Enantiomer	-0.189
		OH •••	
		OH	
		[*]C([*])CC([*])[*]	
FCFP_6	3	AND Enantiomer OH OH	-0.107
		OH	
		[*]0	

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: 0.209

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 nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	ANAPREL	RESERPINE	PHOSPHOROTHIOIC ACID; O;O'-(SULFONYLDI- p-PHENYLENE) O;O;O';O'- TETRAMETHYL ESTER		
Structure		O THE STATE OF THE			
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 Applicability

- 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 Cont	ribution		
	Top features	for positive contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	1186303932	AND Enantiomer AND Enantiomer	0.375
		[*]C([*))(C@@H)(C(=[*	
ECFP_6	642810091	AND Enantiomer AND Enantiomer (*)[c](:[*]):[*]	0.281
ECFP_6	-1897341097	AND Enantiomer AND Enantiomer NS N N N (*]N[*]	0.216
	Ton Features for no	egative contribution	
Fingerprint	Bit/Smiles		Score
ECFP_6	2106656448	AND Enantiomer AND Enantiomer NS N S N (F) C (=O)[*]	-0.352
ECFP_6	2077607946	AND Enantiomer AND Enantiomer (*)O[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-0.252

ECFP_6	655739385	AND Enantiomer -0.239	
		N, H, C, O,	
		[*]:n:[*]	