

Supplementary File

Design, *in silico* and *in vitro* Evaluations of a Novel Nicotinamide Derivative as a VEGFR-2 inhibitor

Ibrahim H. Eissa^{a*}, Muhammad Abd ElGayed Bkrah^a, Reda G. Yousef^a, Hazem Elkady^a, Eslam B. Elkaeed^{b*}, Bshra A. Alsfook^c, Ibrahim M. Ibrahim^d, Ahmed M. Metwaly^{e,f*}, Dalal Z. Husein^g

^a Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt. muhammad.bkrah@azhar.edu.eg; redayousof@azhar.edu.eg Hazemelkady@azhar.edu.eg

^b Department of Pharmaceutical Sciences, College of Pharmacy, AlMaarefa University, Riyadh 13713, Saudi Arabia. ekaheed@um.edu.sa

^c Department of Pharmaceutical Sciences, College of Pharmacy, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia baalsfook@pnu.edu.sa;

^d Biophysics Department, Faculty of Science, Cairo University. Giza 12613, Egypt. ibrahim_mohamed@cu.edu.eg

^e Pharmacognosy and Medicinal Plants Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt.

^f Biopharmaceutical Products Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications (SRTA-City), Alexandria, Egypt.

^g Chemistry Department, Faculty of Science, New Valley University, El-Kharja 72511, Egypt. dalal_husein@sci.nvu.edu.eg

Content

Fig. S.1. Change in the eigenvalues with increasing the eigenvectors (blue). The cumulative variance retained in the eigenvectors (red)	
Fig.S.2. 1 st ten eigenvectors' distribution.	
Fig. S.3. Cosine content of the 1 st ten eigenvectors for the two trajectories.	
Method	Molecular Docking
	MD Simulations
	MM-GBSA
	DFT
	Essential dynamics
	ADMET studies
	Synthesis
	<i>In vitro</i> assays
Spectral Data	IR and Ms
	¹H and ¹³C NMR
Toxicity report	

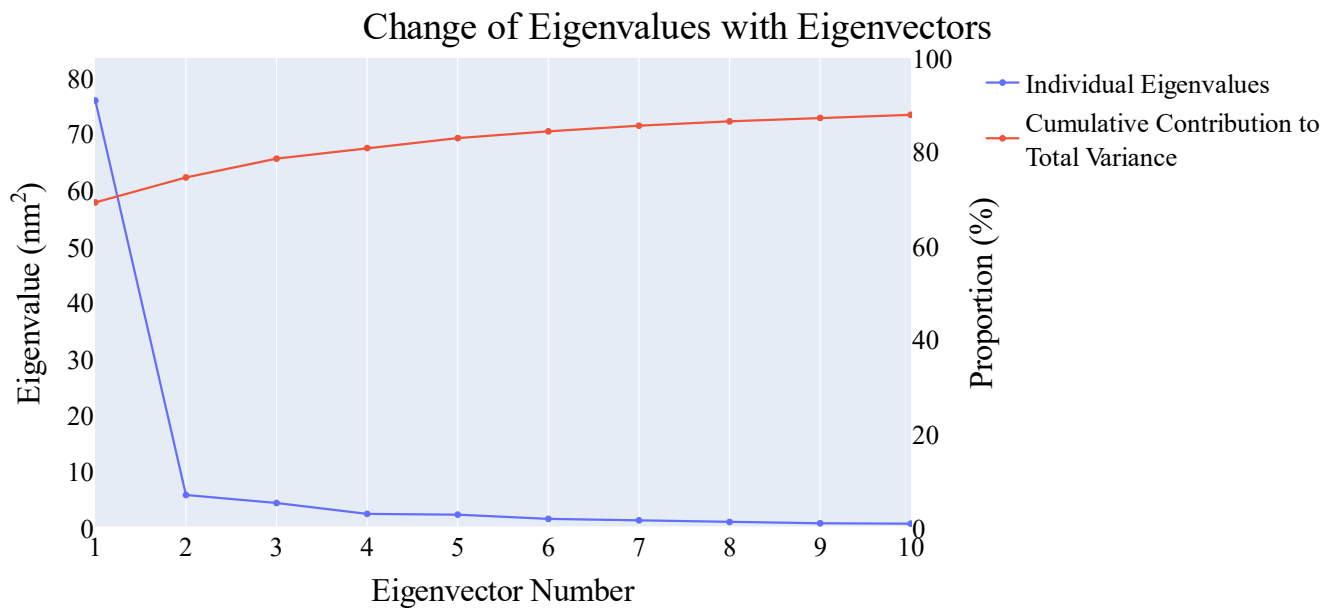


Fig. S.1. Change in the eigenvalues with increasing the eigenvectors (blue). The cumulative variance retained in the eigenvectors (red)

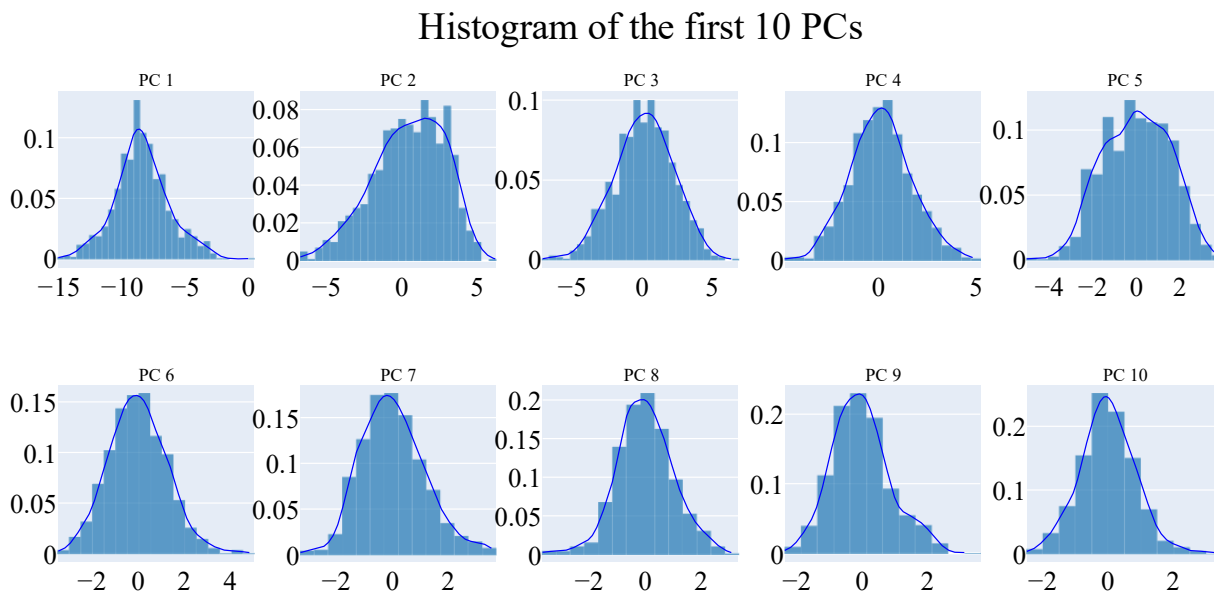


Fig.S.2. 1st ten eigenvectors' distribution.

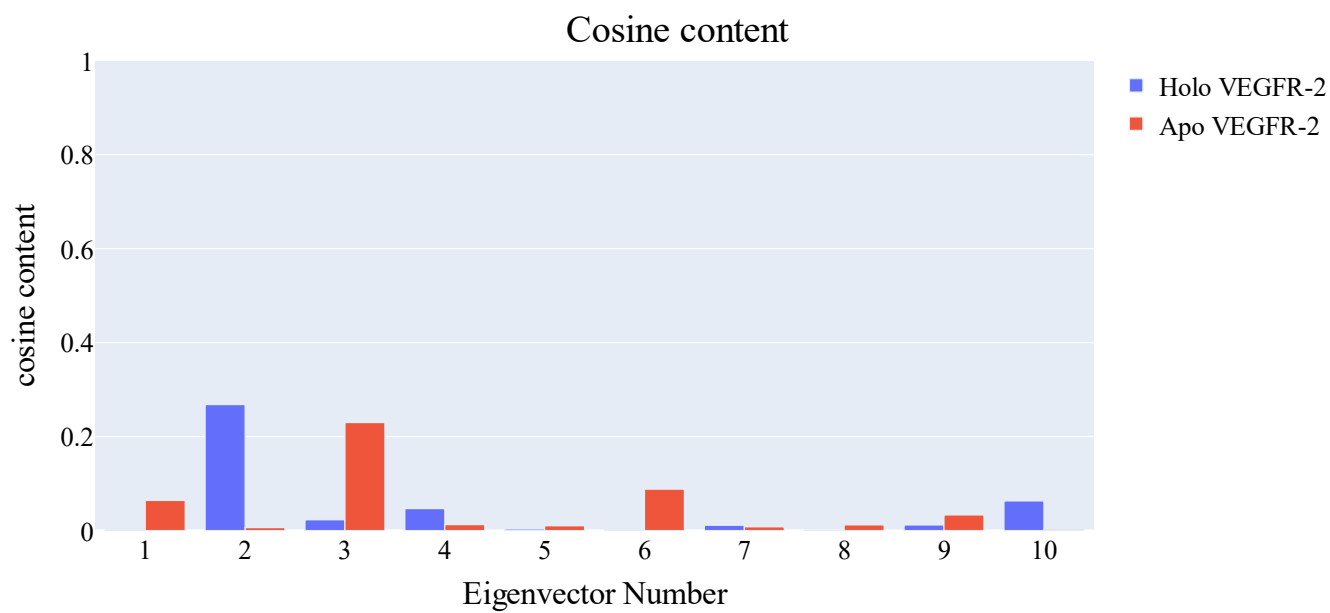
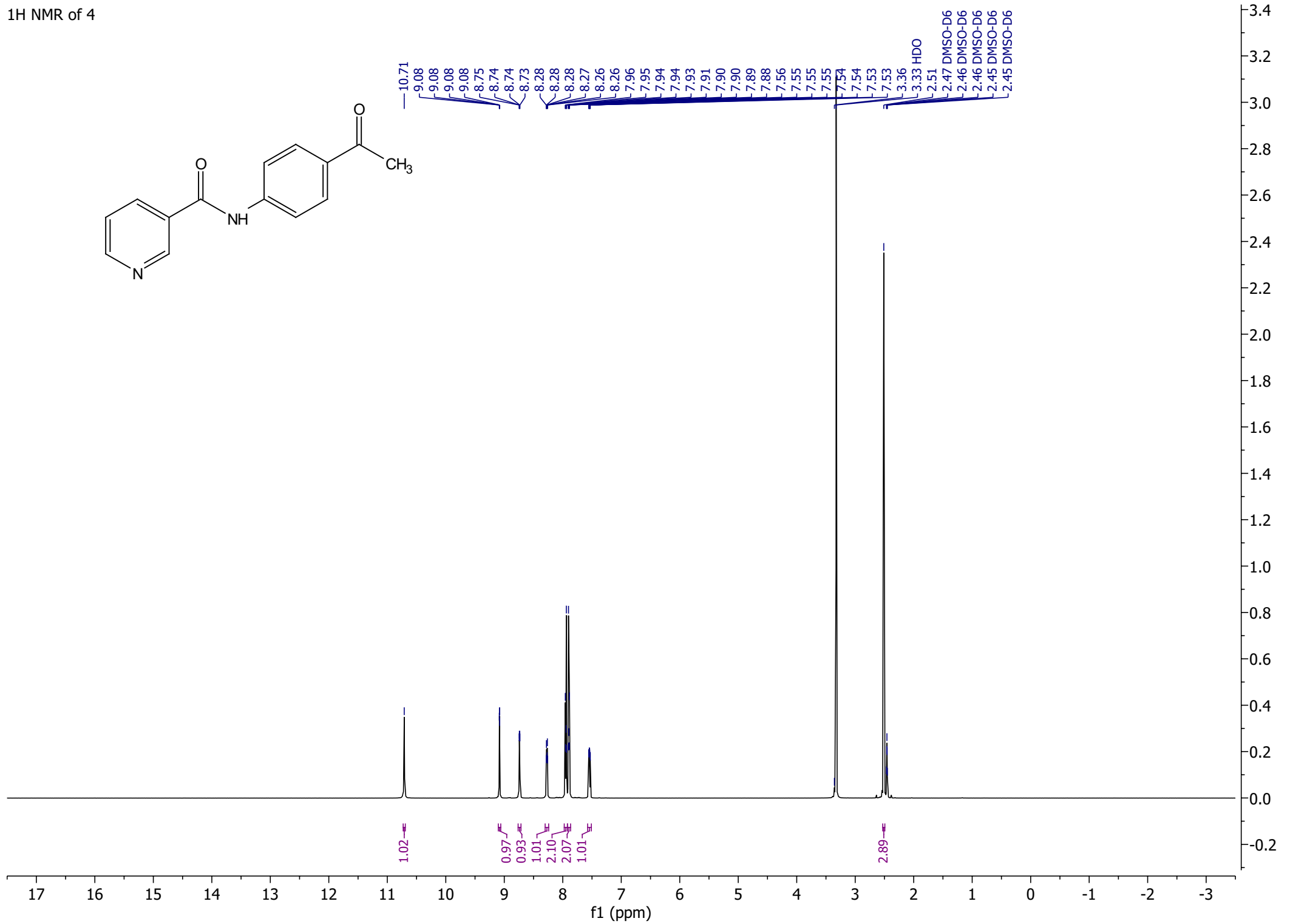
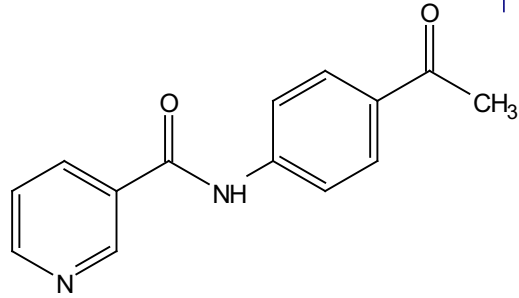


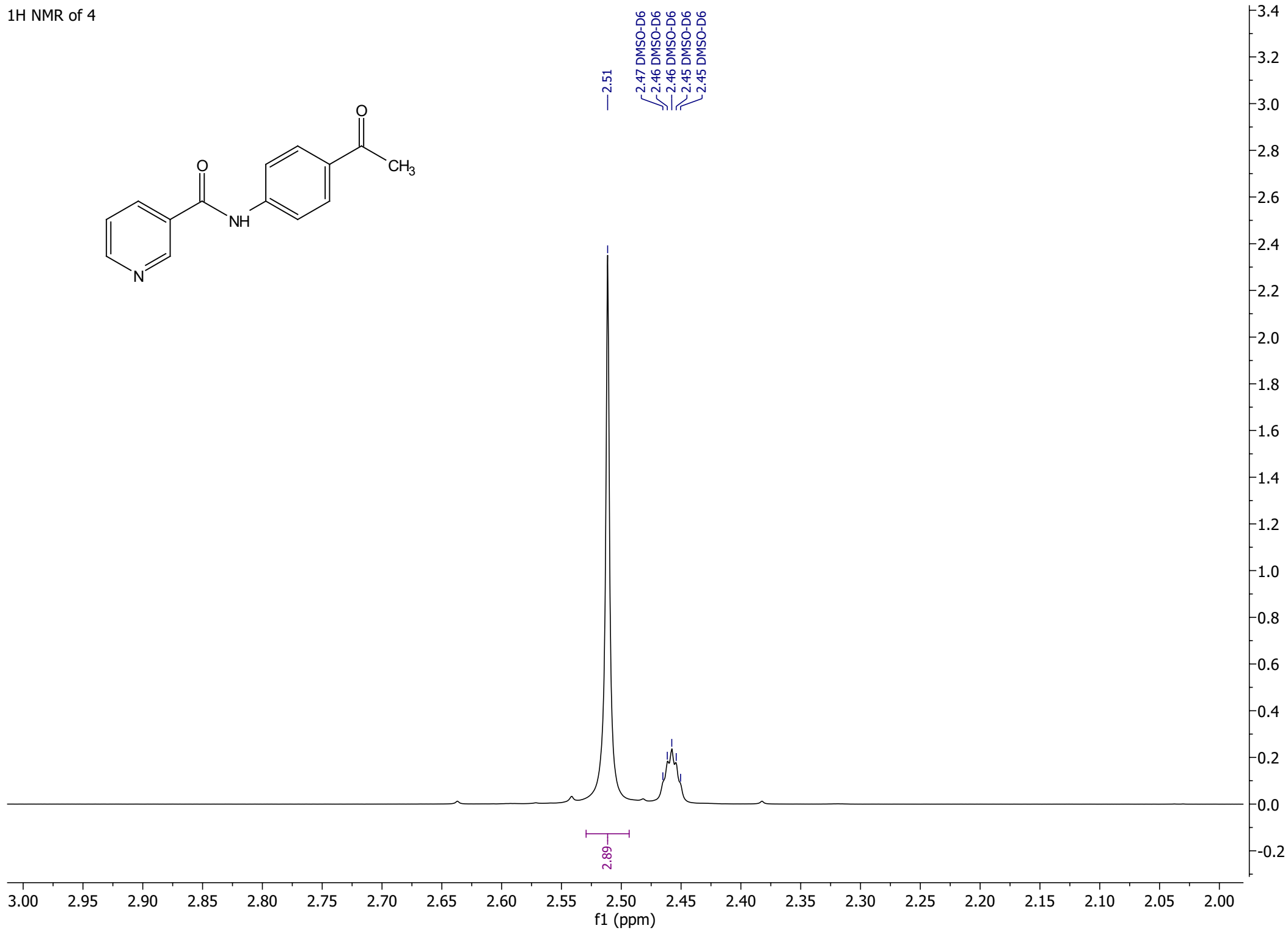
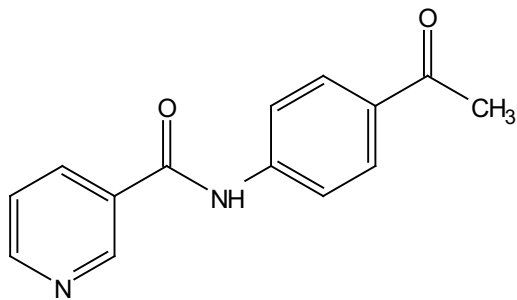
Fig. S.3. Cosine content of the 1st ten eigenvectors for the two trajectories.

Spectral data

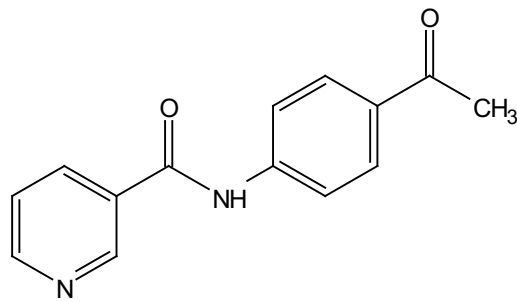
1H NMR of 4



1H NMR of 4



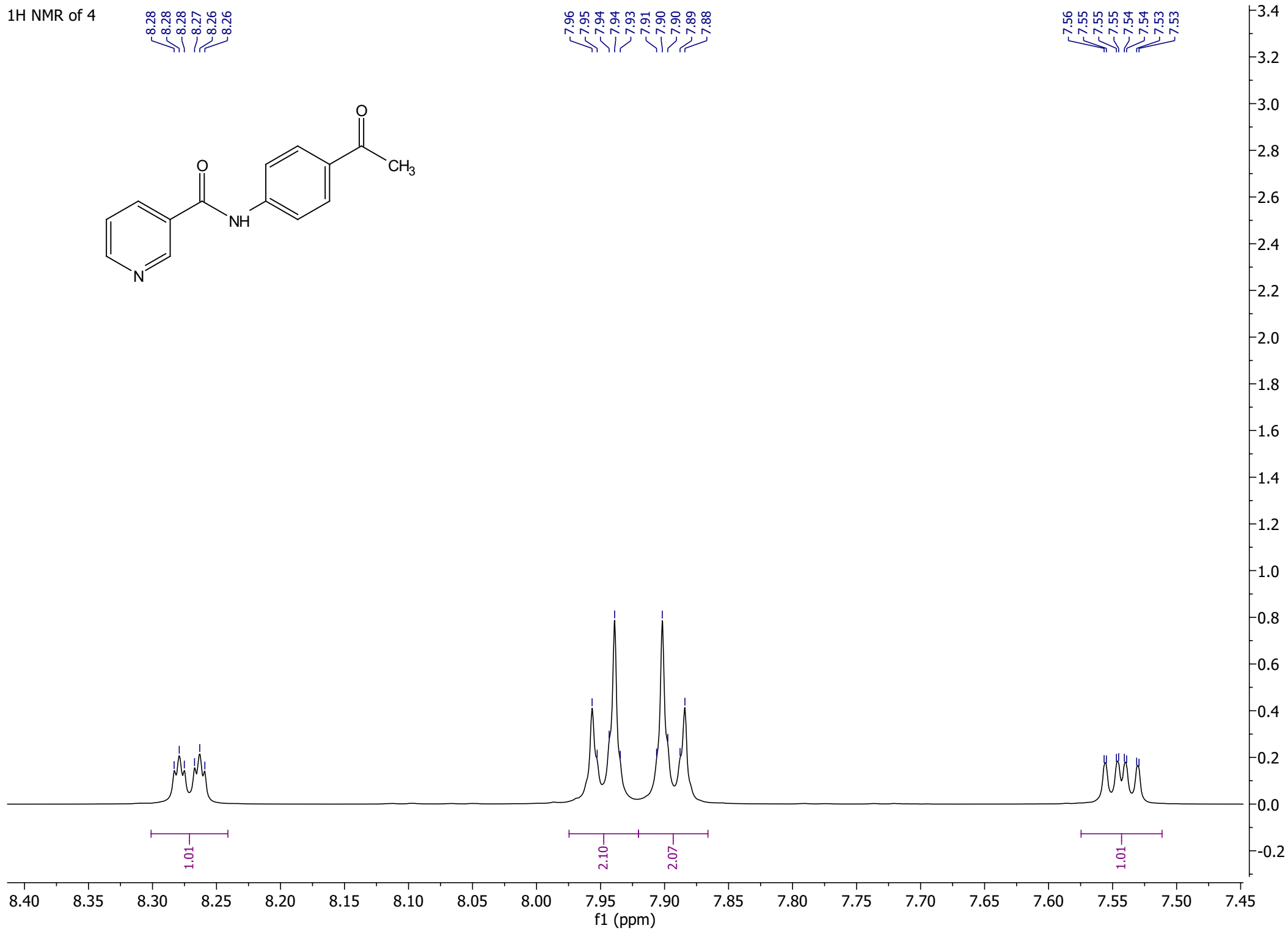
1H NMR of 4



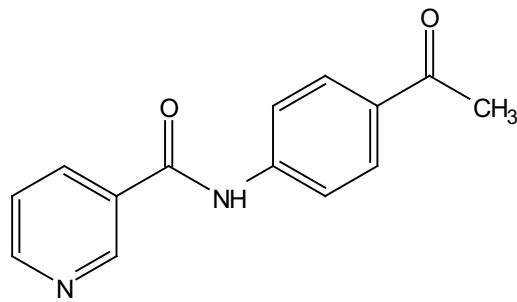
8.28
8.28
8.27
8.26
8.26

7.96
7.95
7.94
7.94
7.93
7.91
7.90
7.90
7.89
7.88

7.56
7.55
7.55
7.54
7.54
7.53
7.53

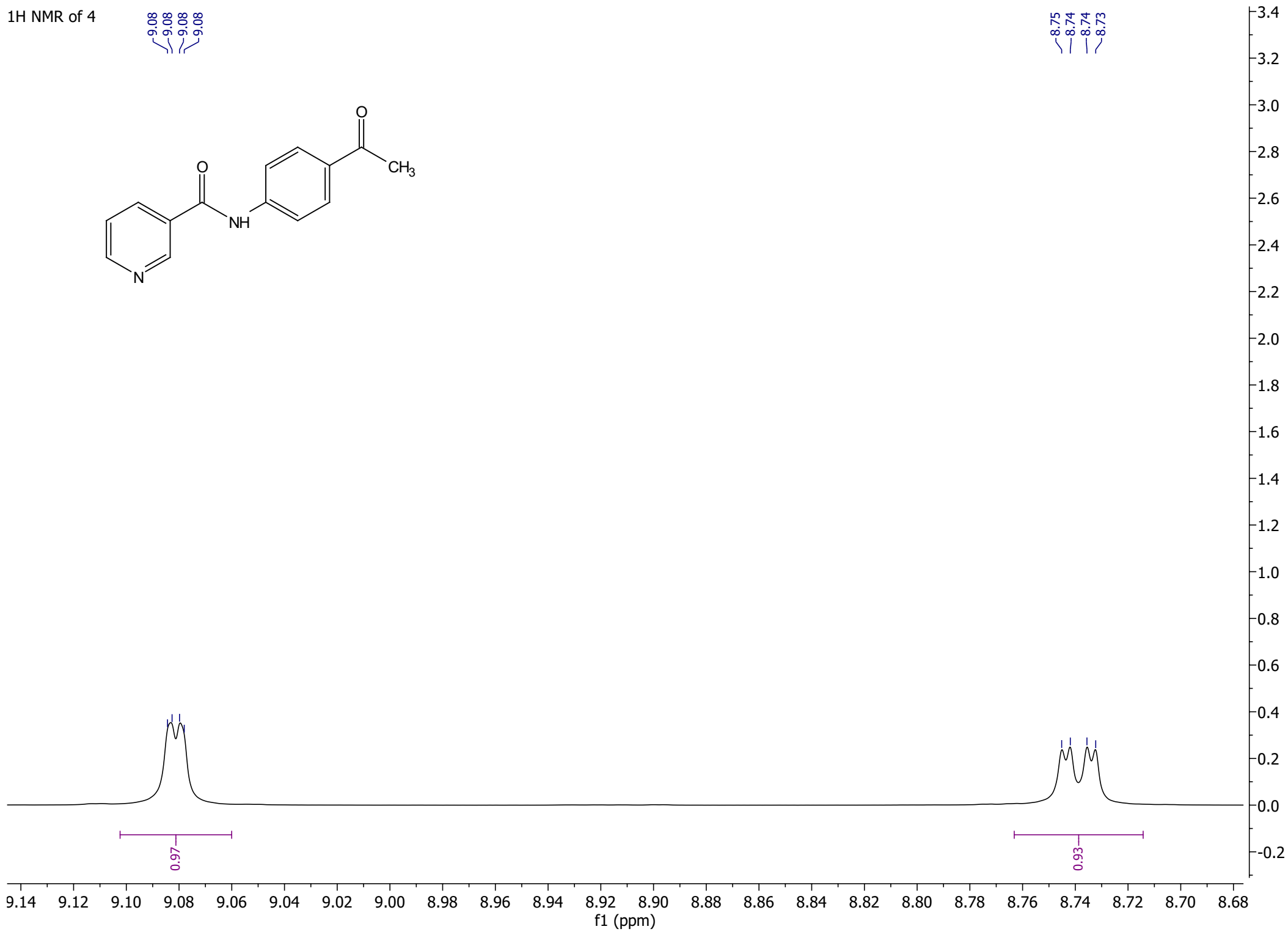


1H NMR of 4

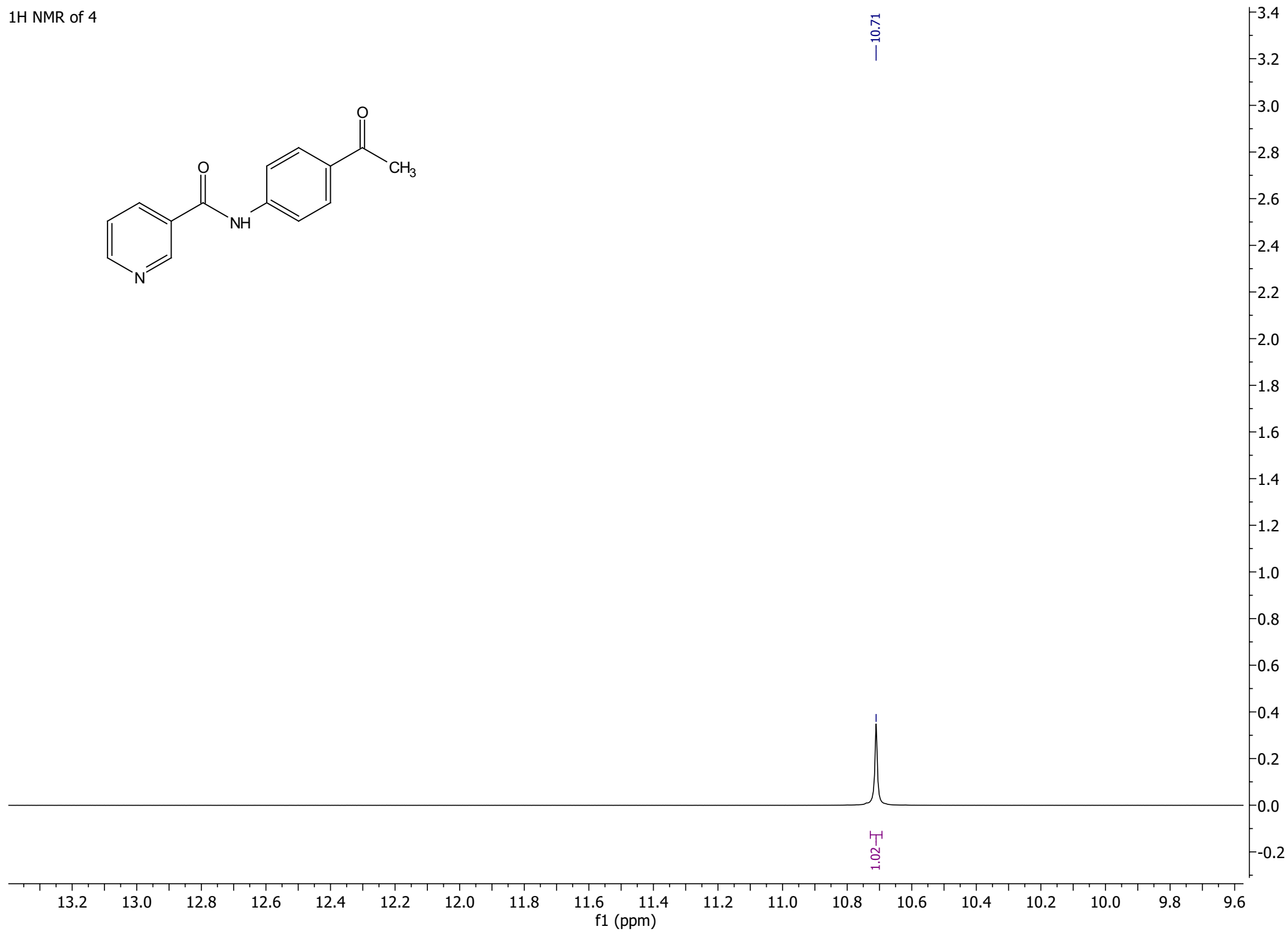
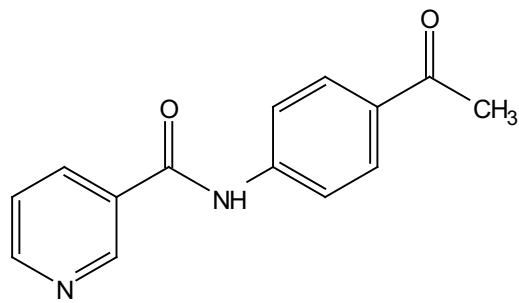


9.08
9.08
9.08
9.08

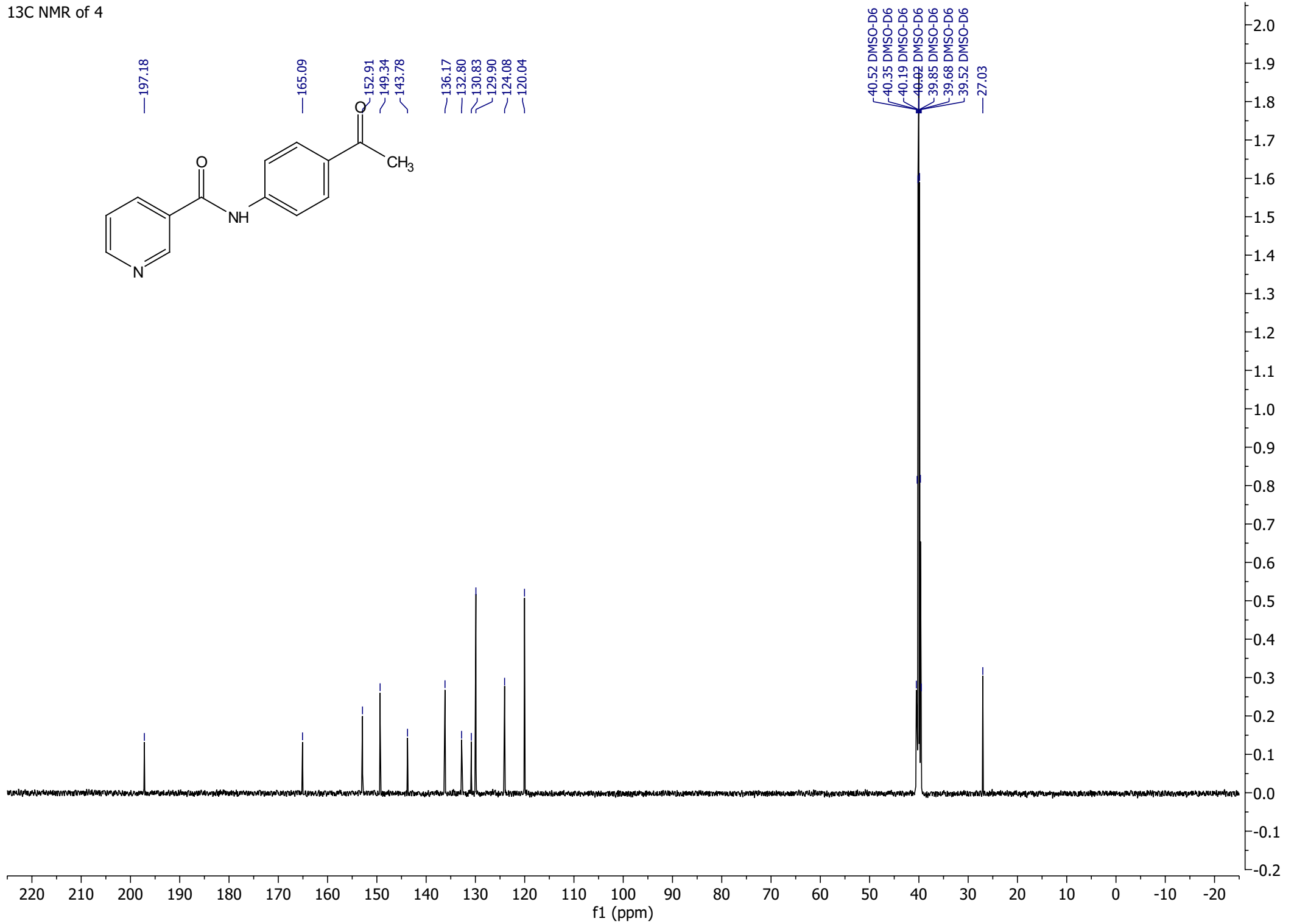
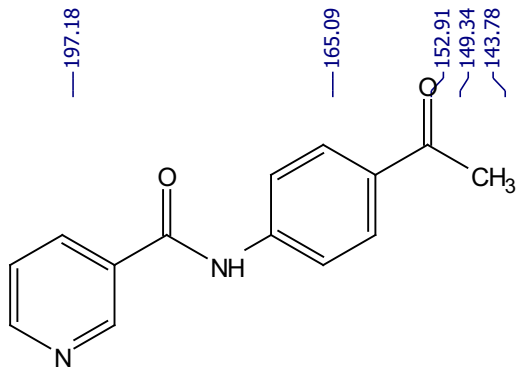
8.75
8.74
8.74
8.73

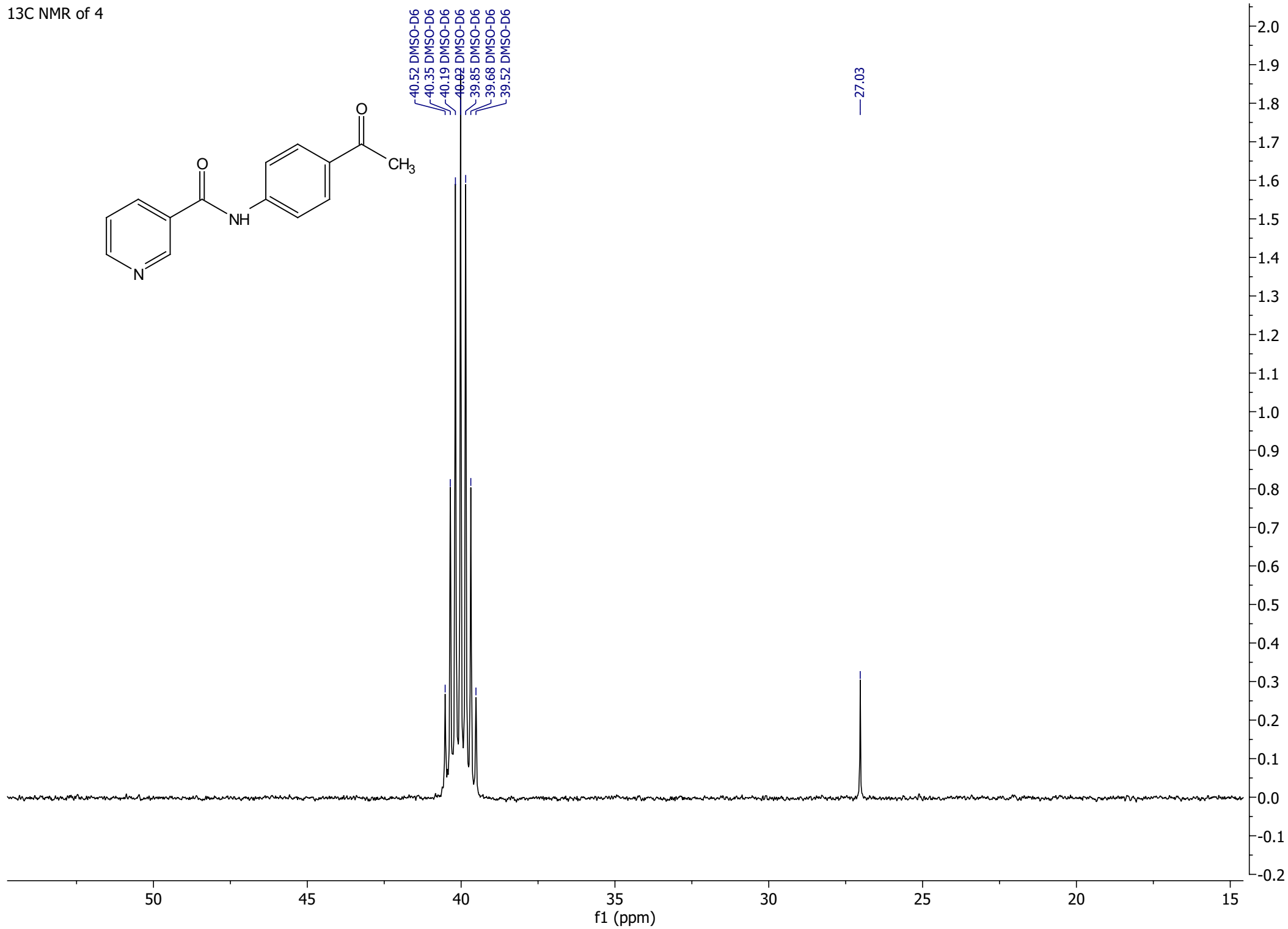
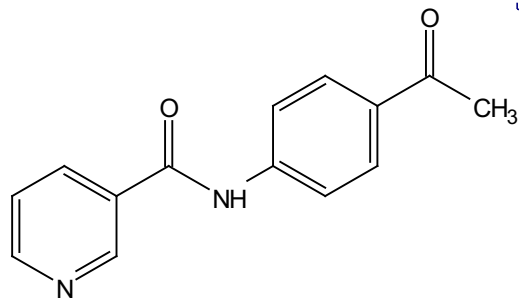


¹H NMR of 4

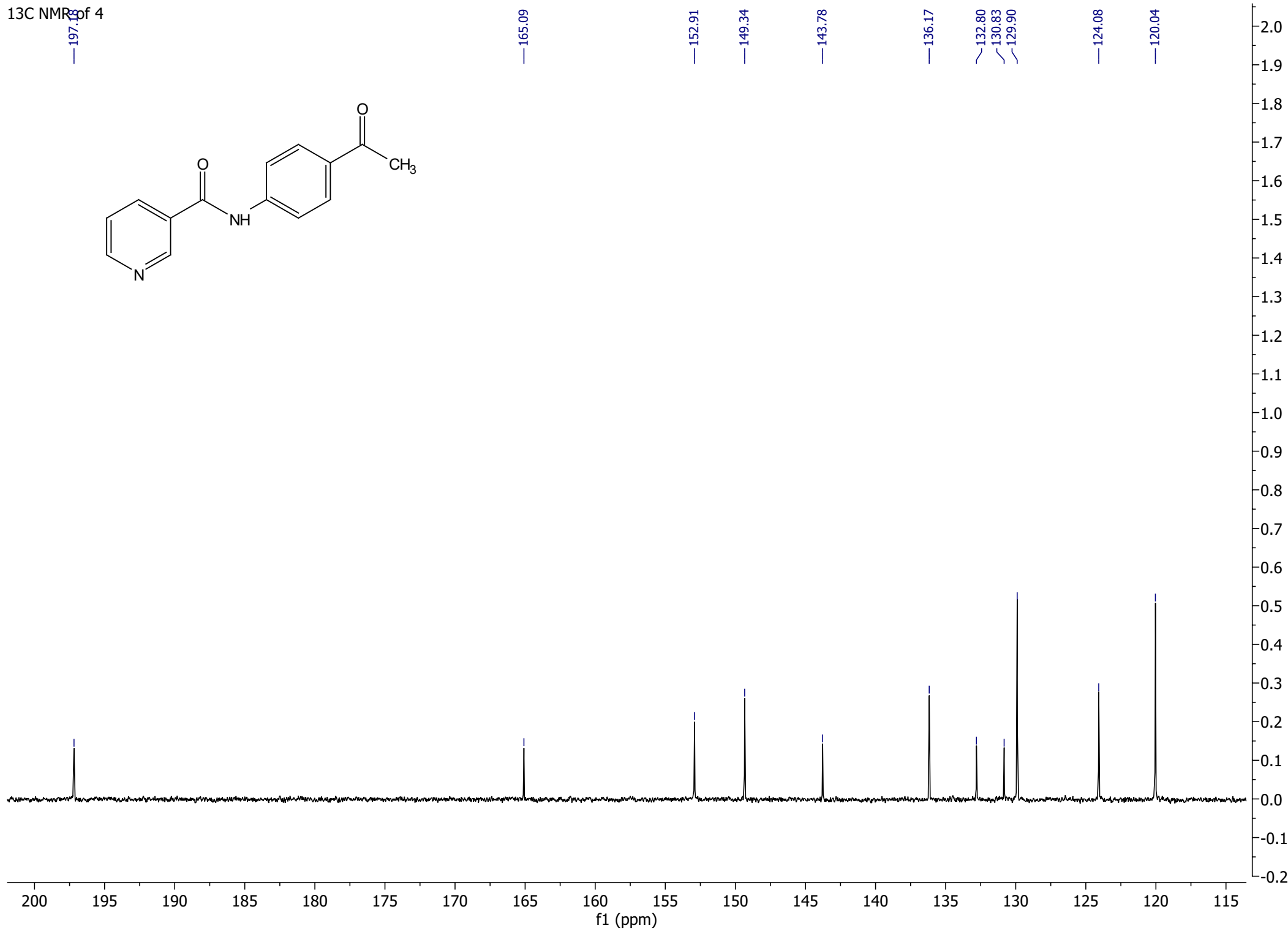
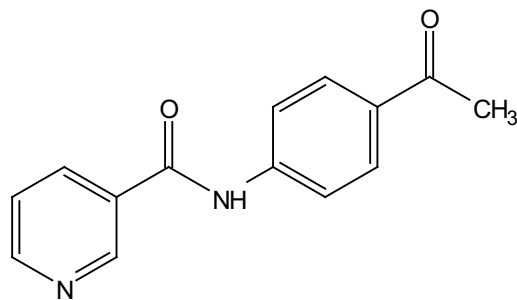


¹³C NMR of 4

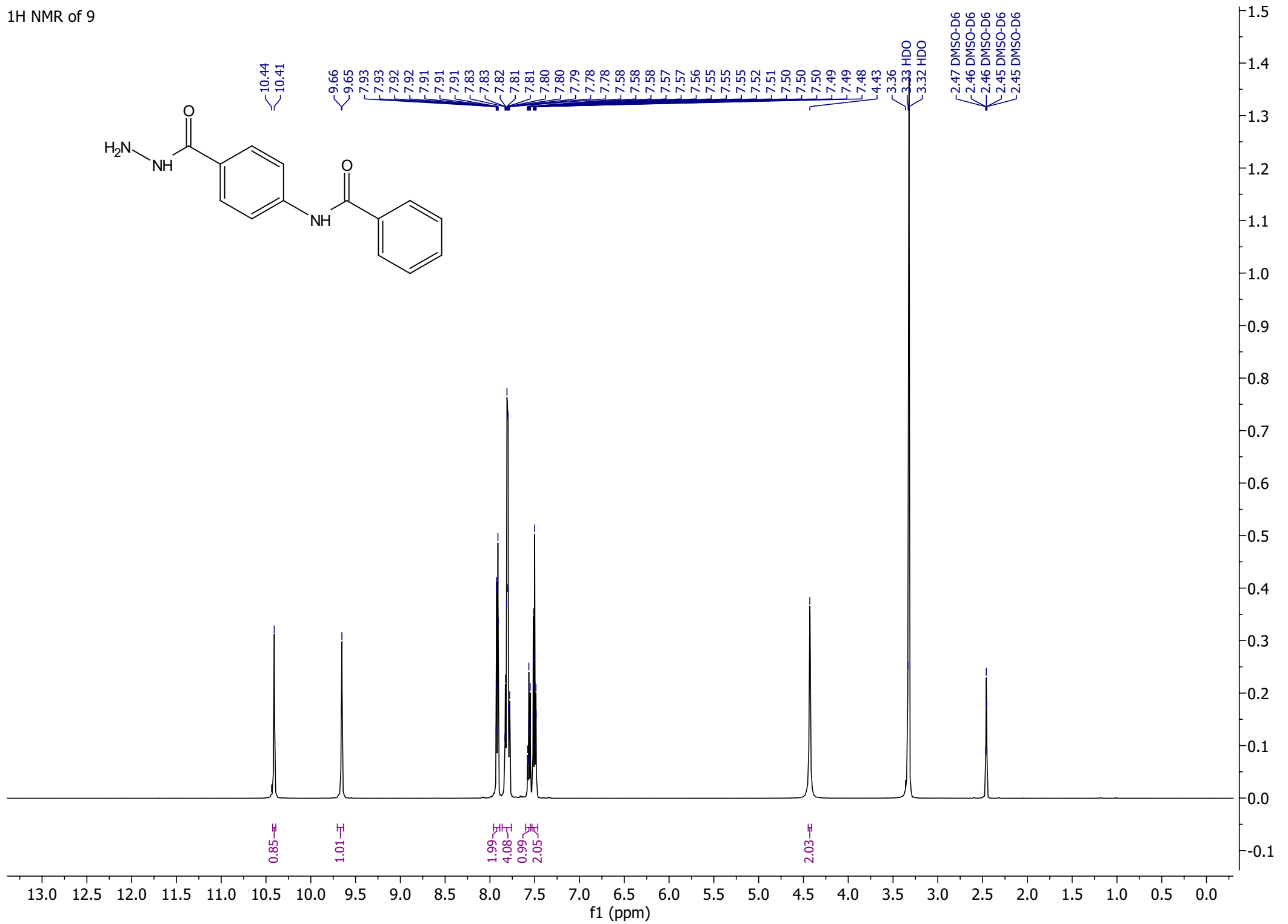
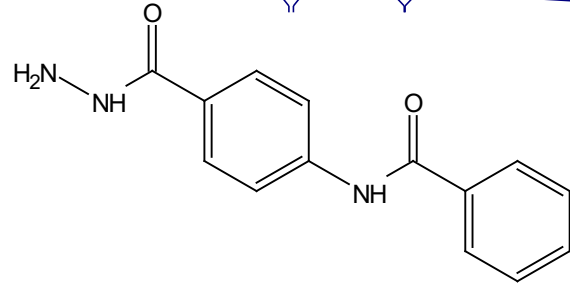




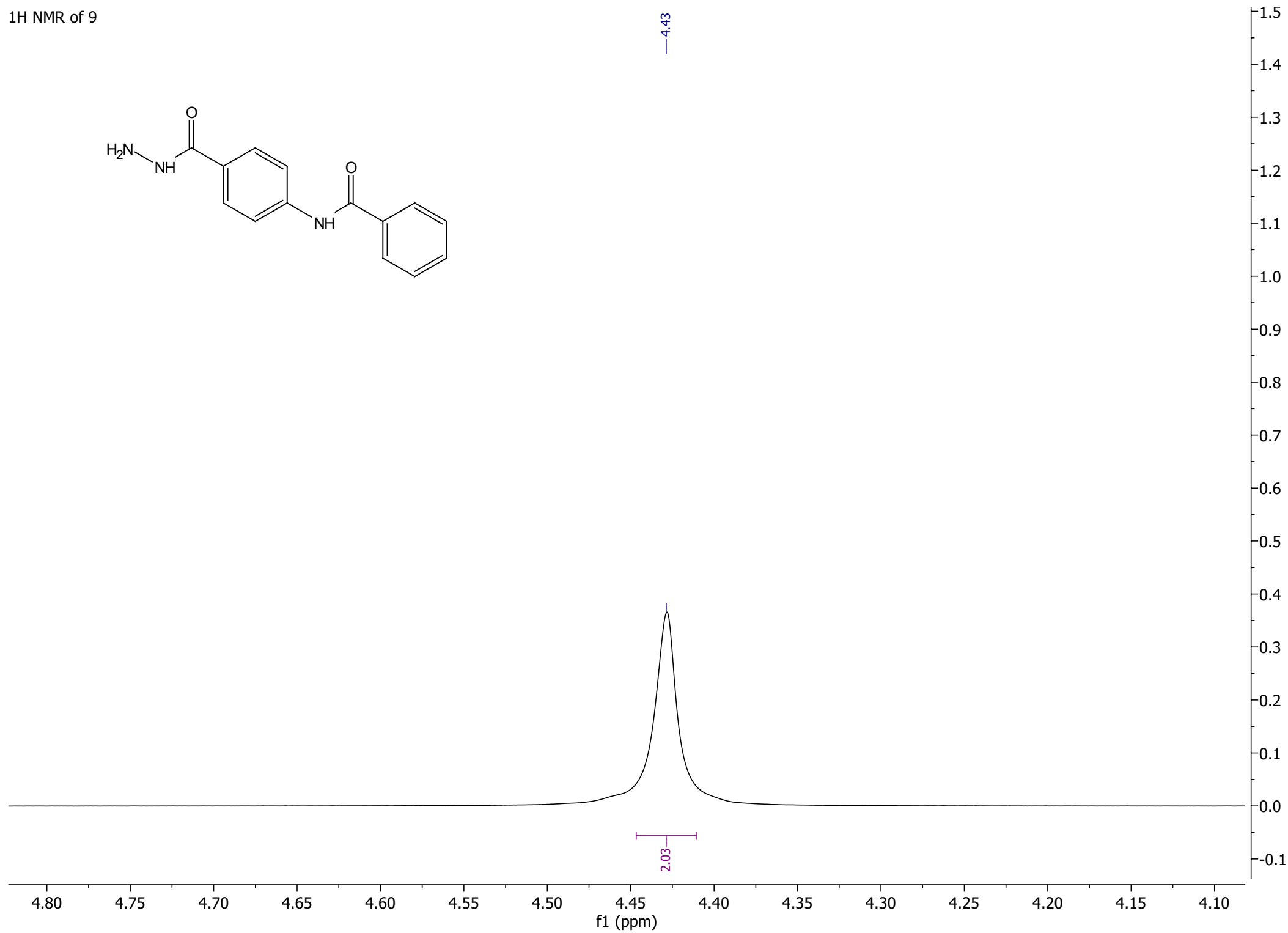
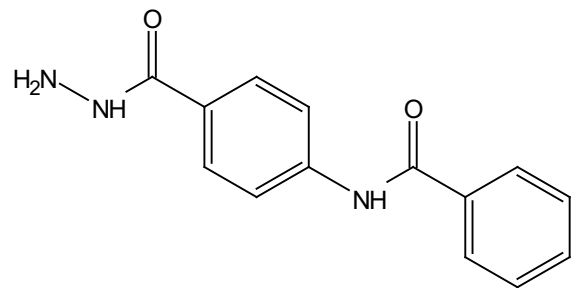
13C NMR of 4



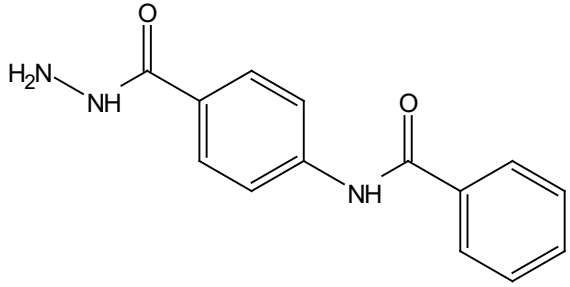
1H NMR of 9



¹H NMR of 9



¹H NMR of 9

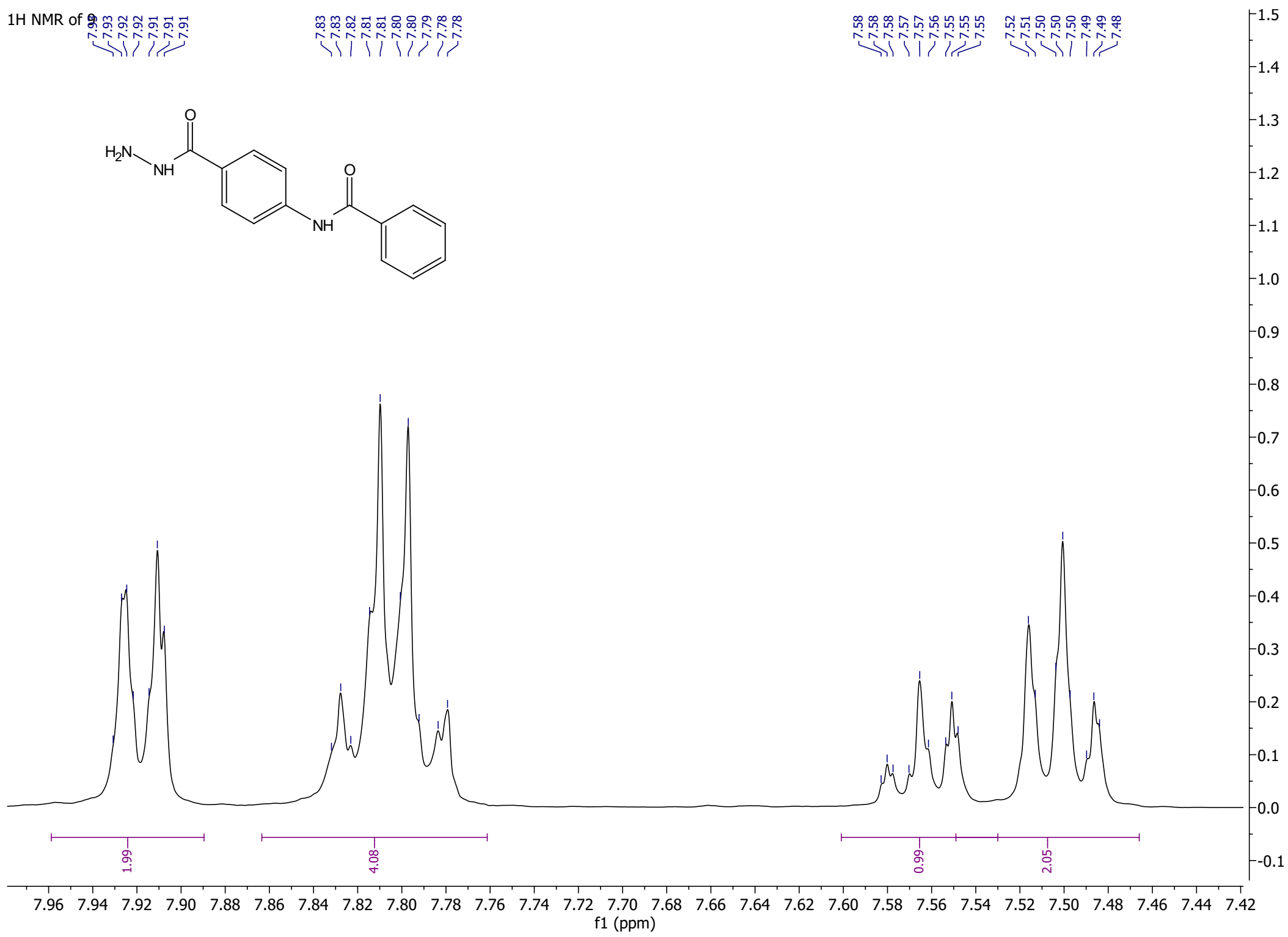


7.99
7.93
7.92
7.92
7.91
7.91

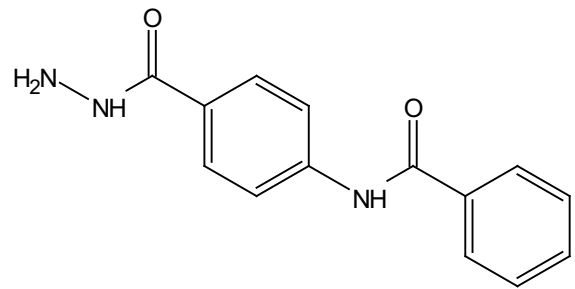
7.83
7.83
7.82
7.81
7.81
7.80
7.80
7.79
7.78
7.78

7.58
7.58
7.58
7.57
7.57
7.56
7.55
7.55

7.52
7.51
7.50
7.50
7.50
7.49
7.49
7.48

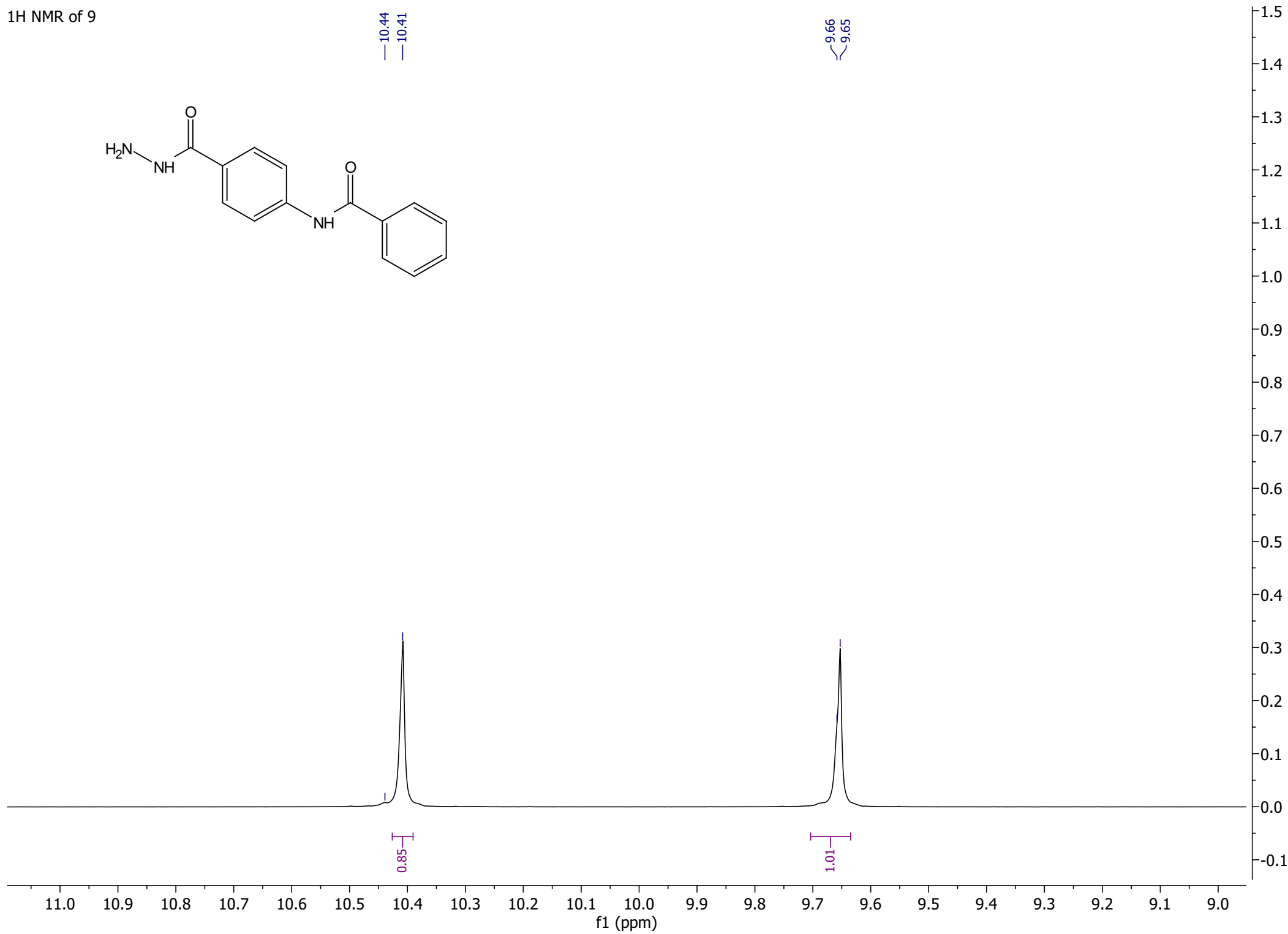


¹H NMR of 9

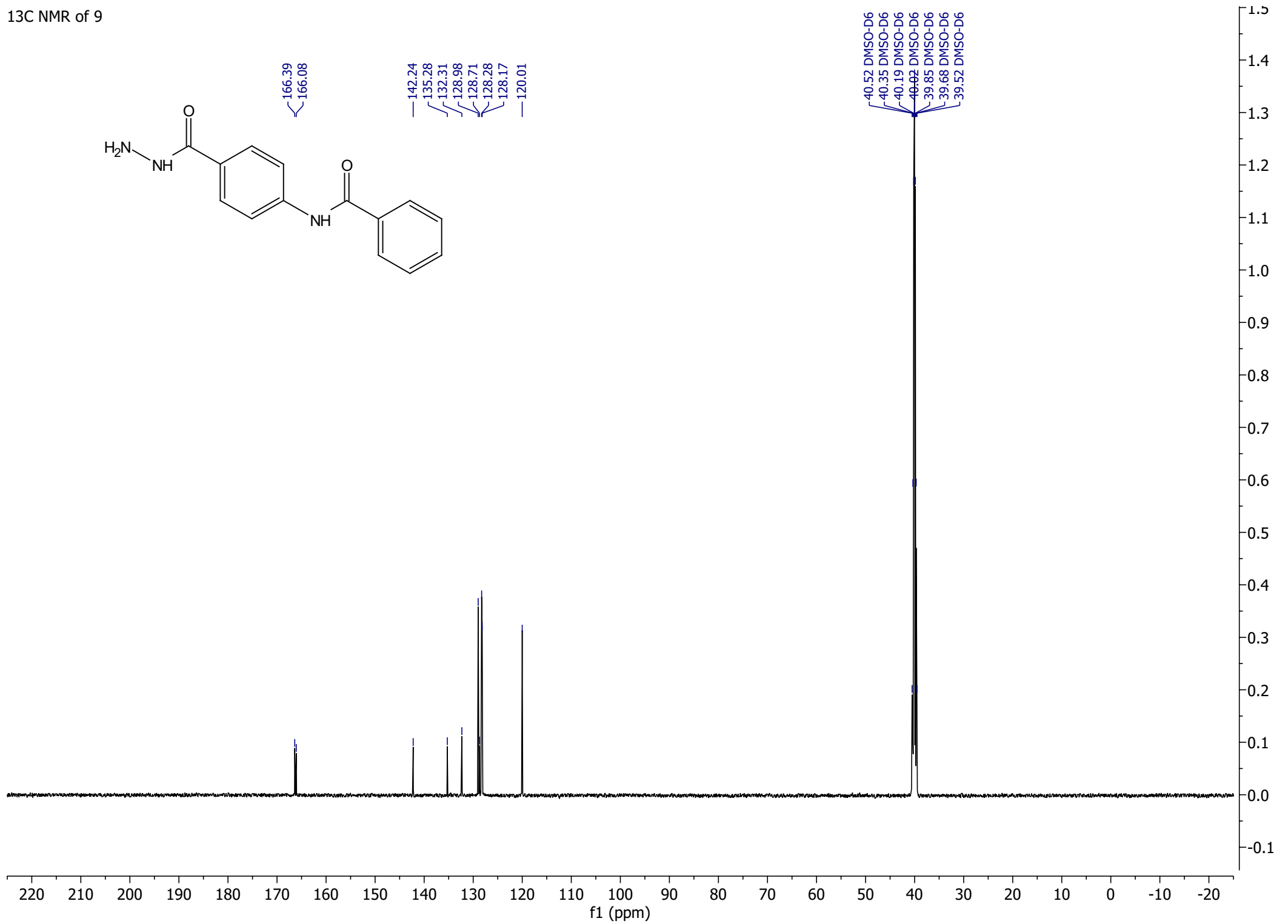
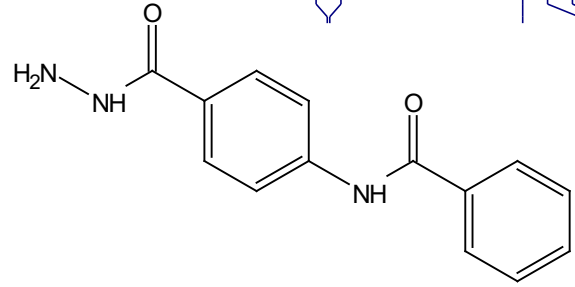


— 10.44
— 10.41

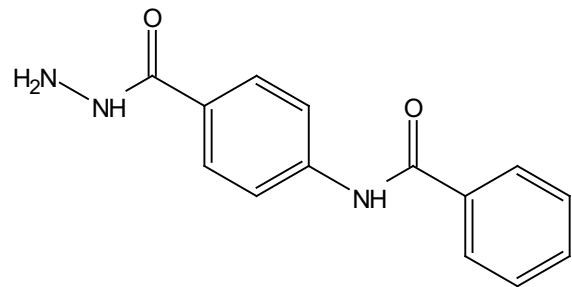
— 9.66
— 9.65



¹³C NMR of 9

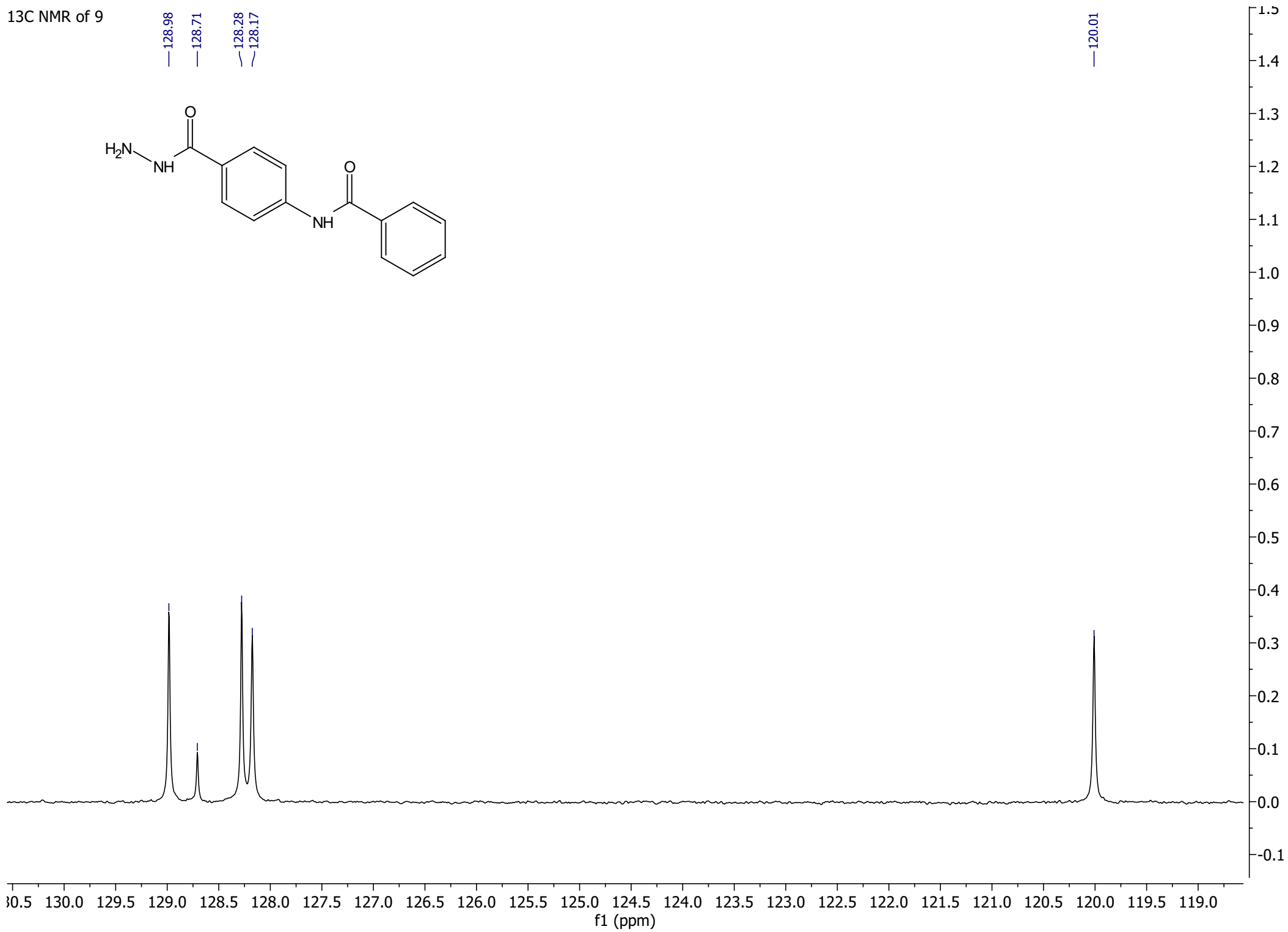


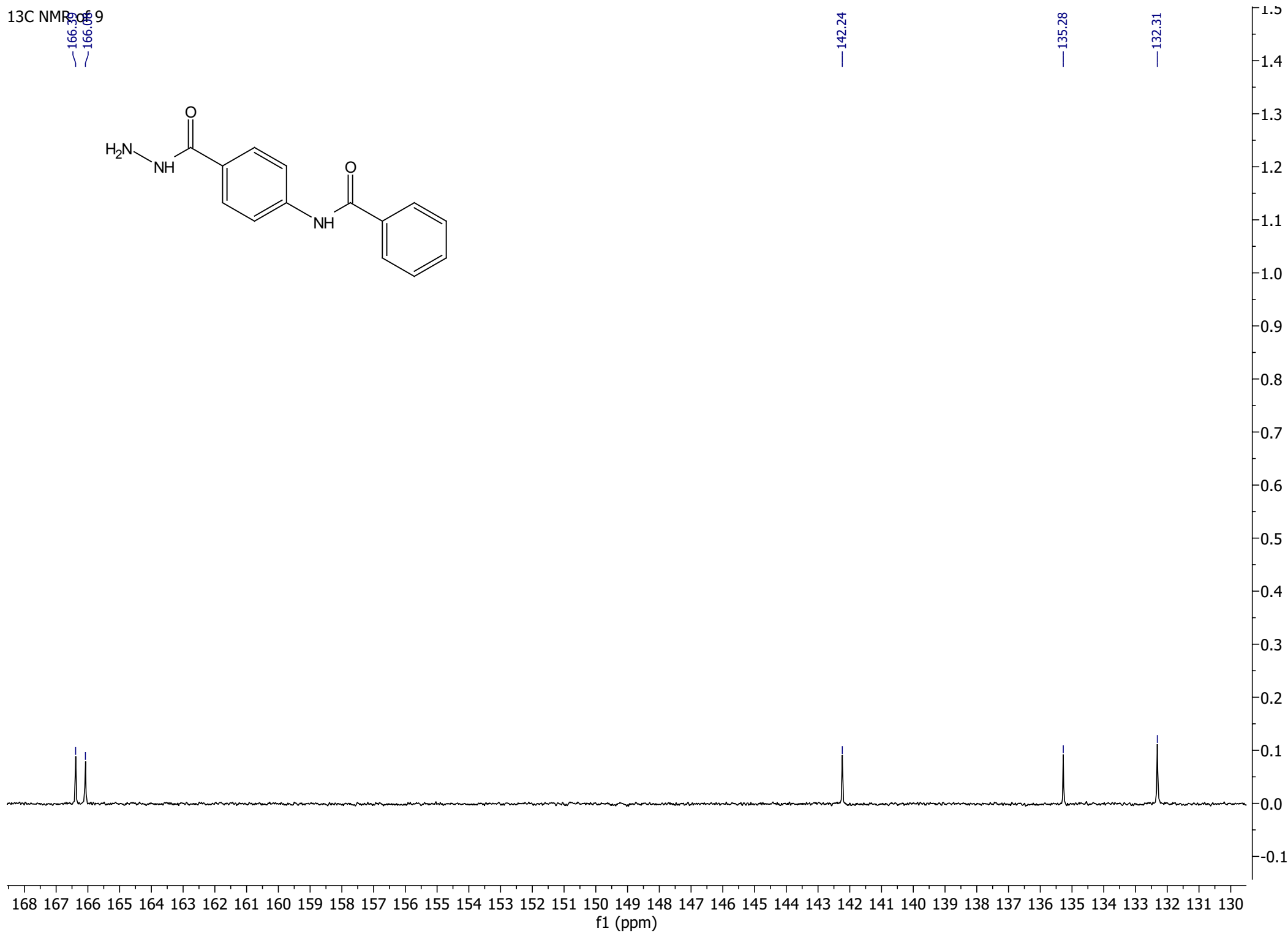
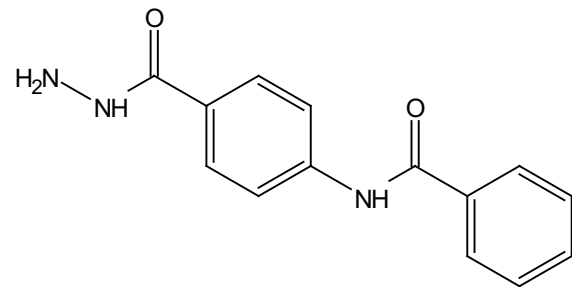
¹³C NMR of 9



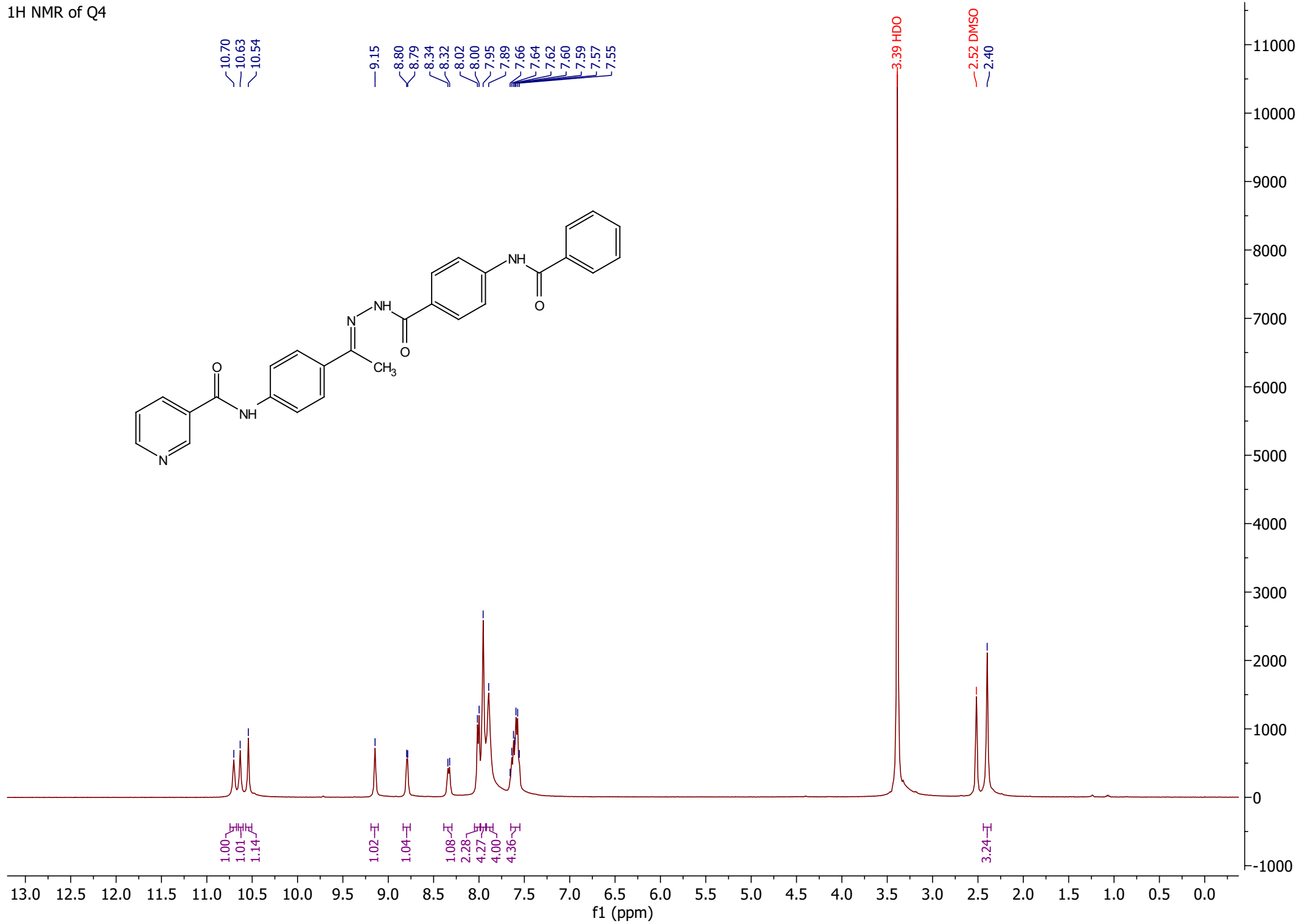
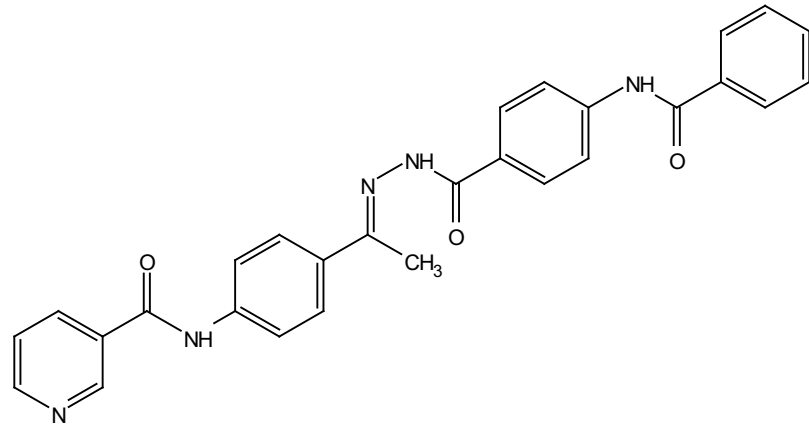
128.98
128.71
128.28
128.17

120.01

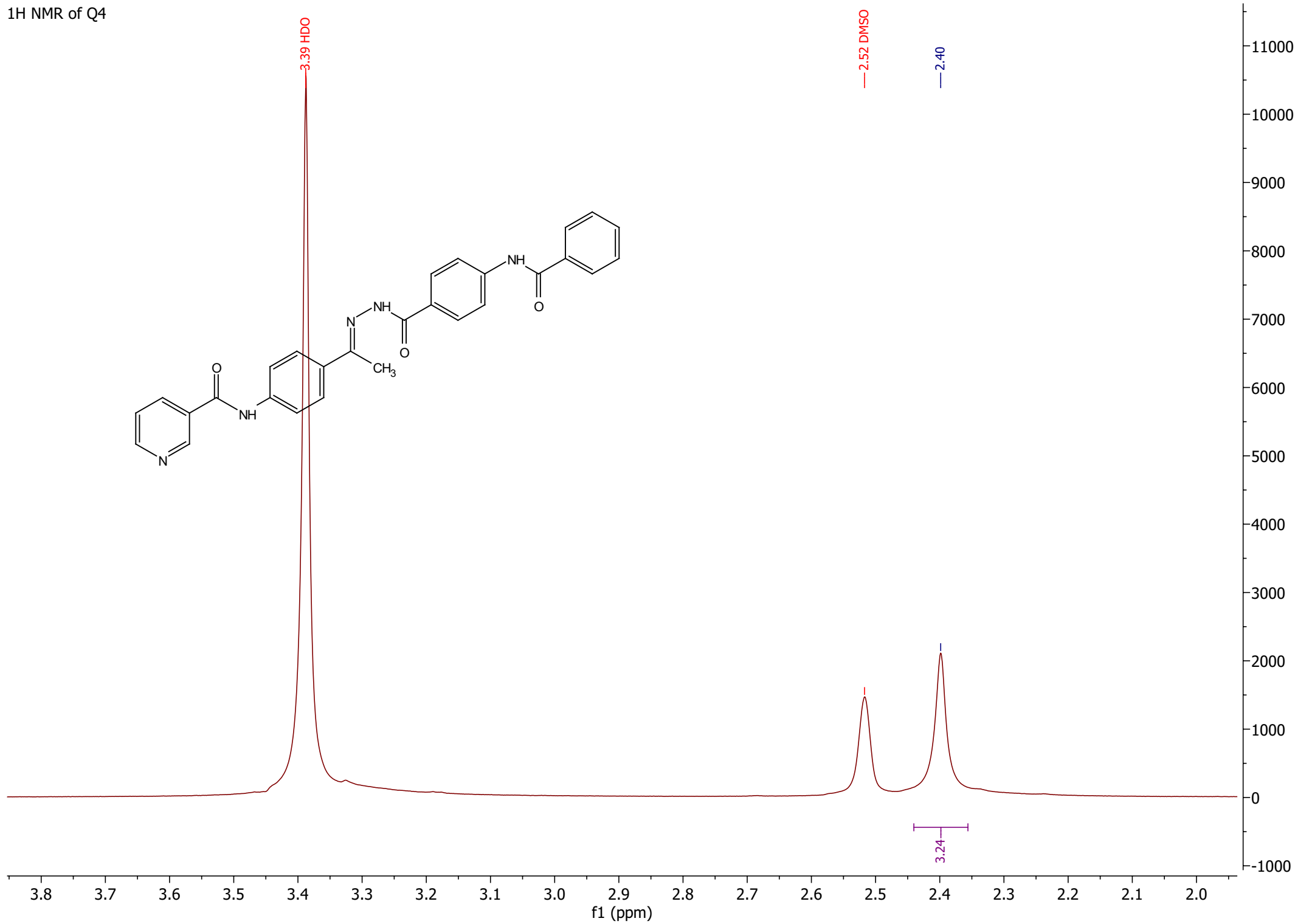
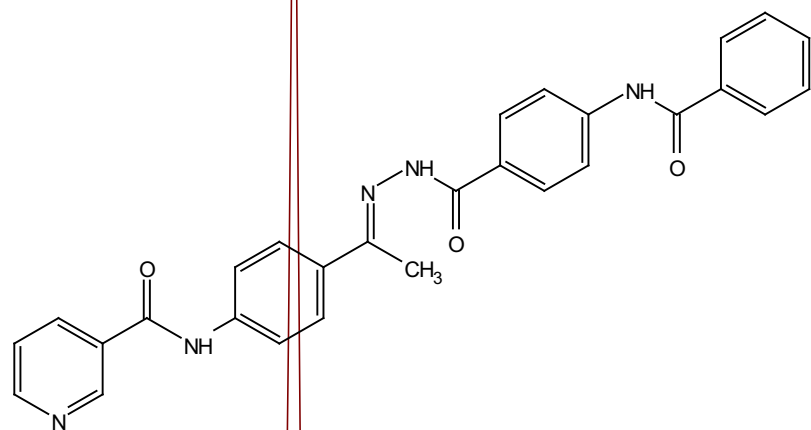




1H NMR of Q4

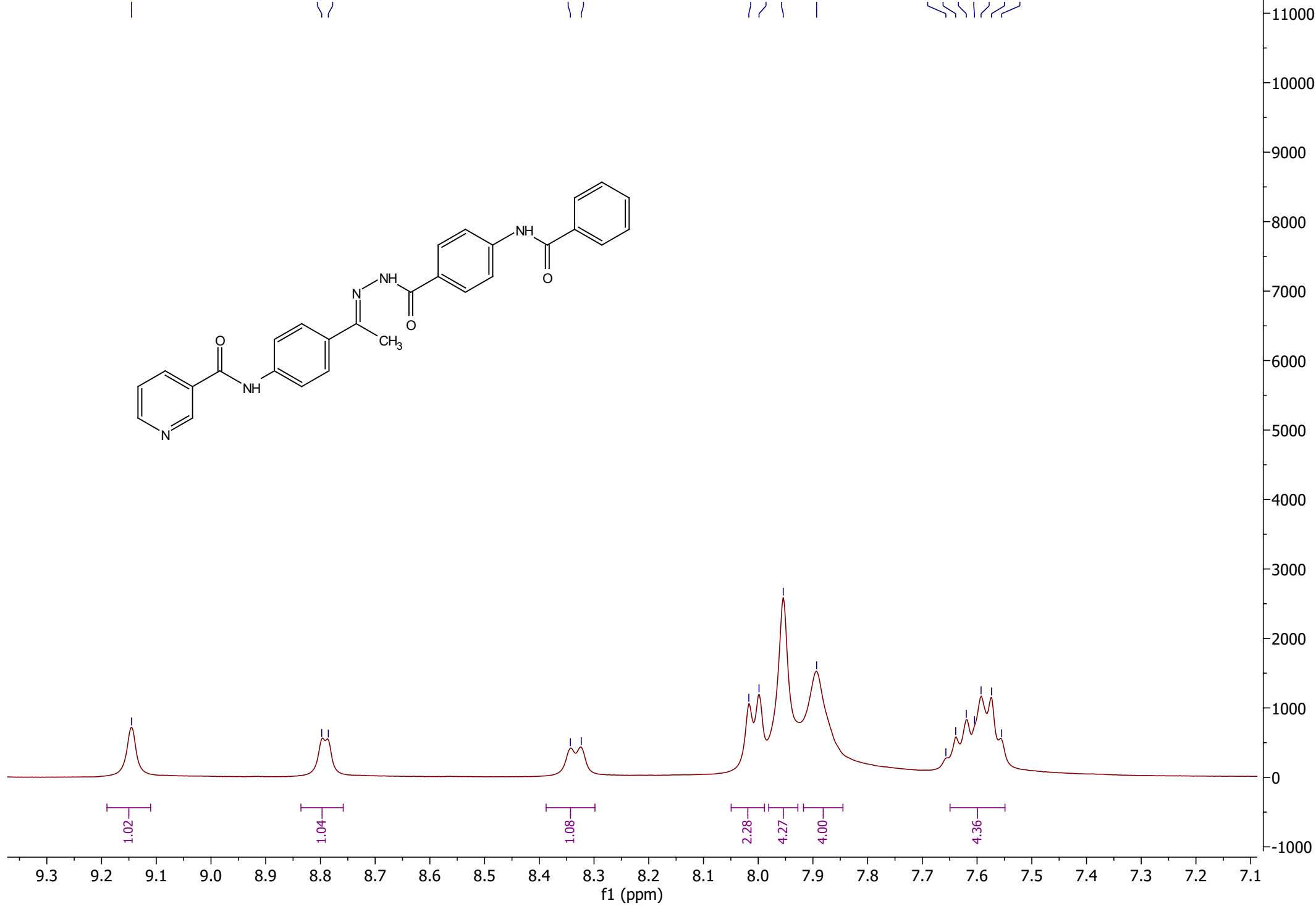
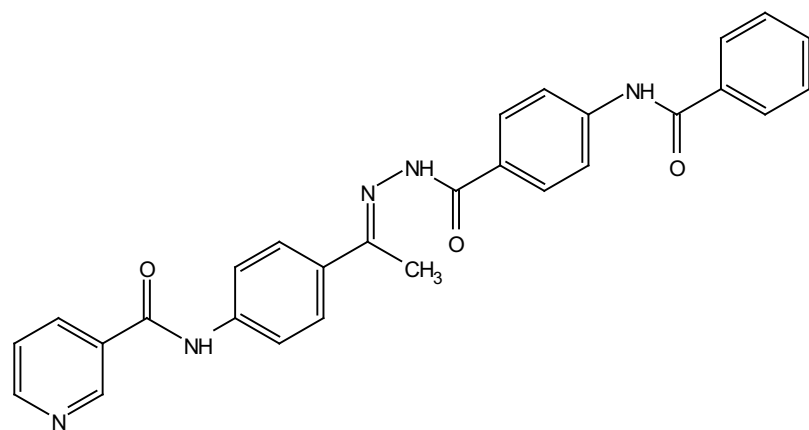


1H NMR of Q4

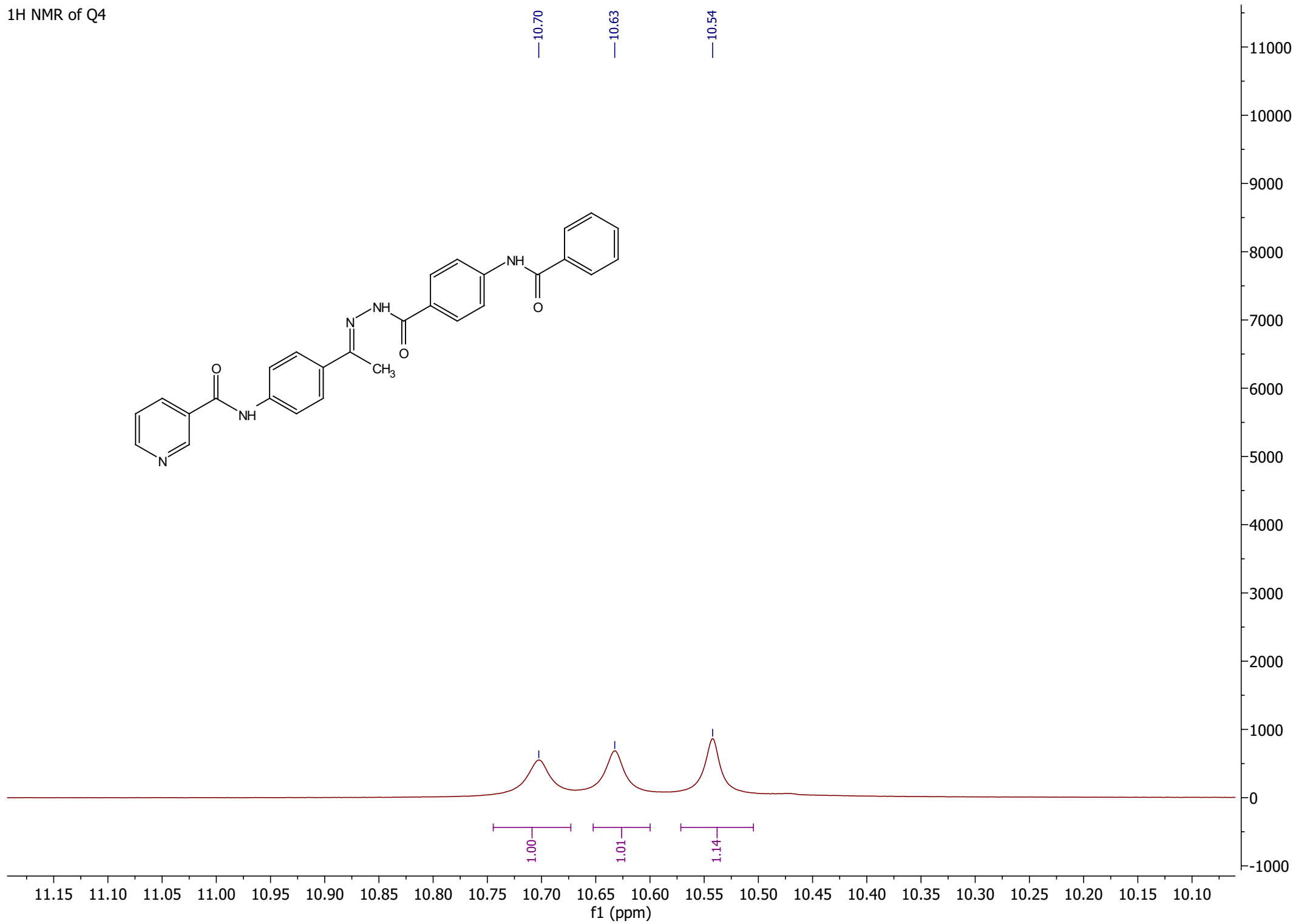
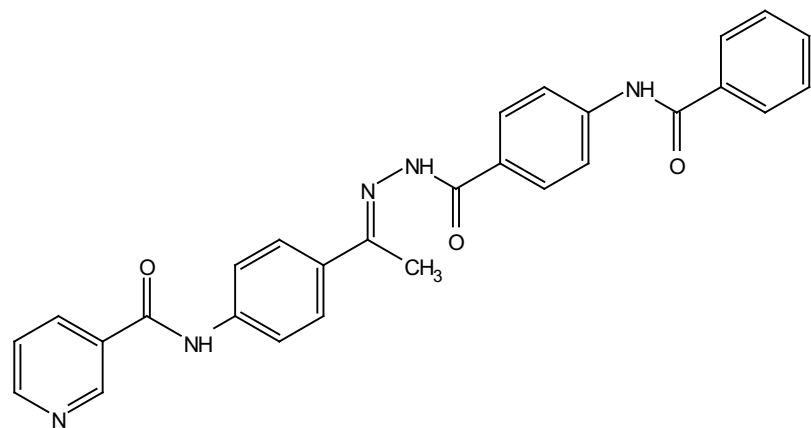


1H NMR of Q4

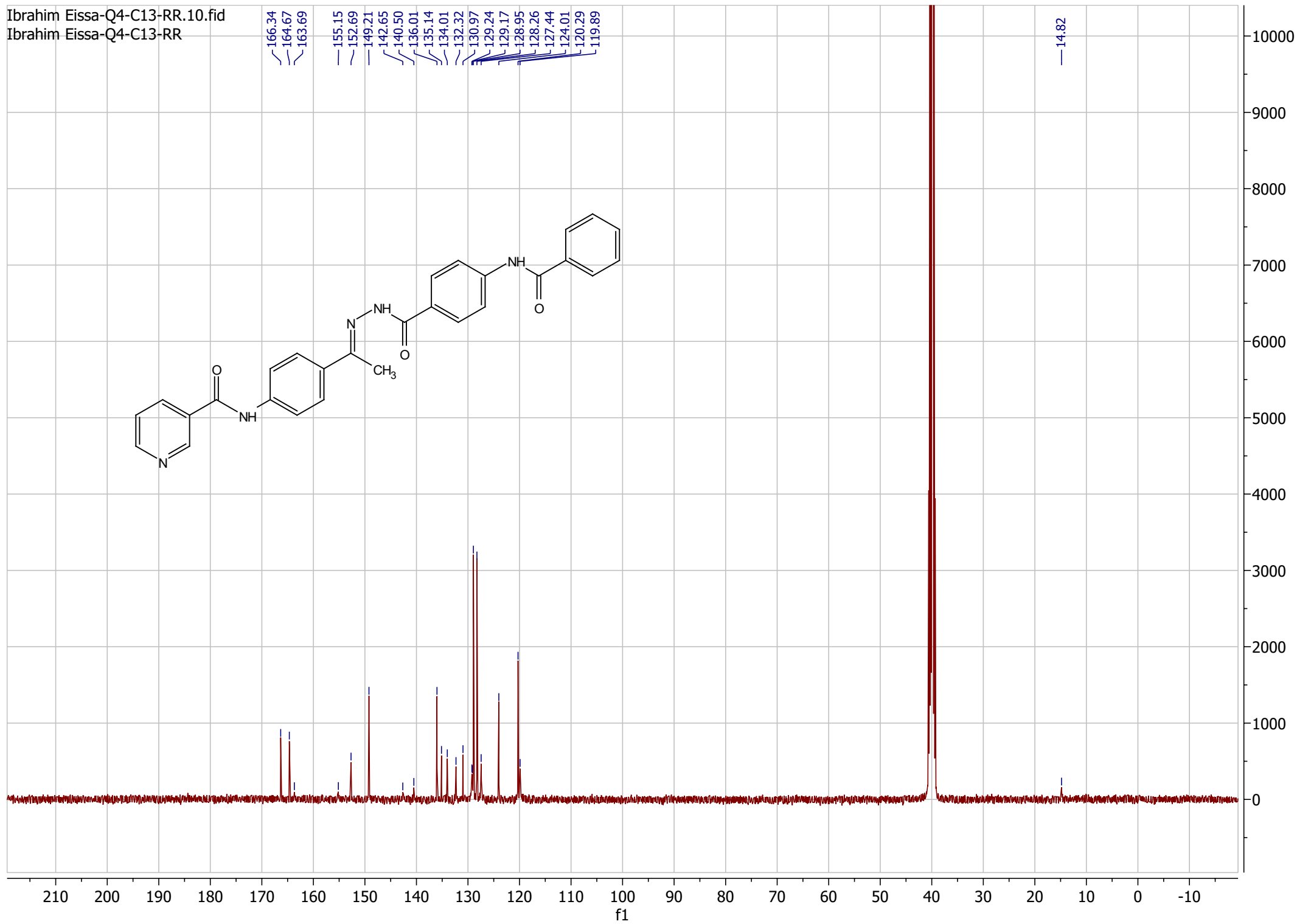
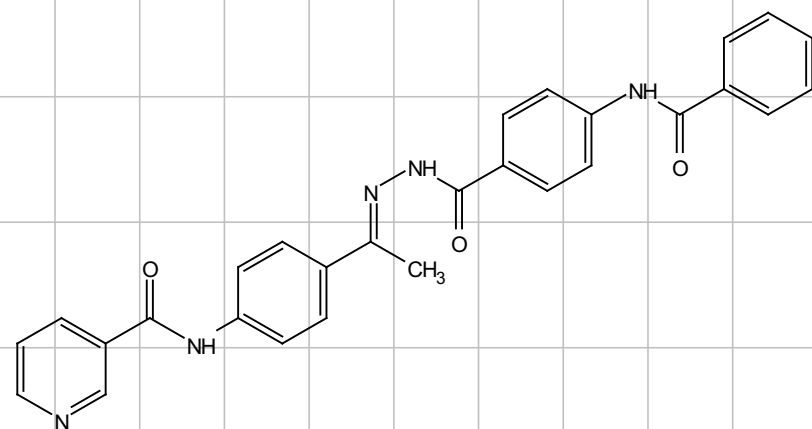
9.15
8.80
8.79
8.34
8.32
8.02
8.00
7.95
7.89
7.66
7.64
7.62
7.60
7.59
7.57
7.55



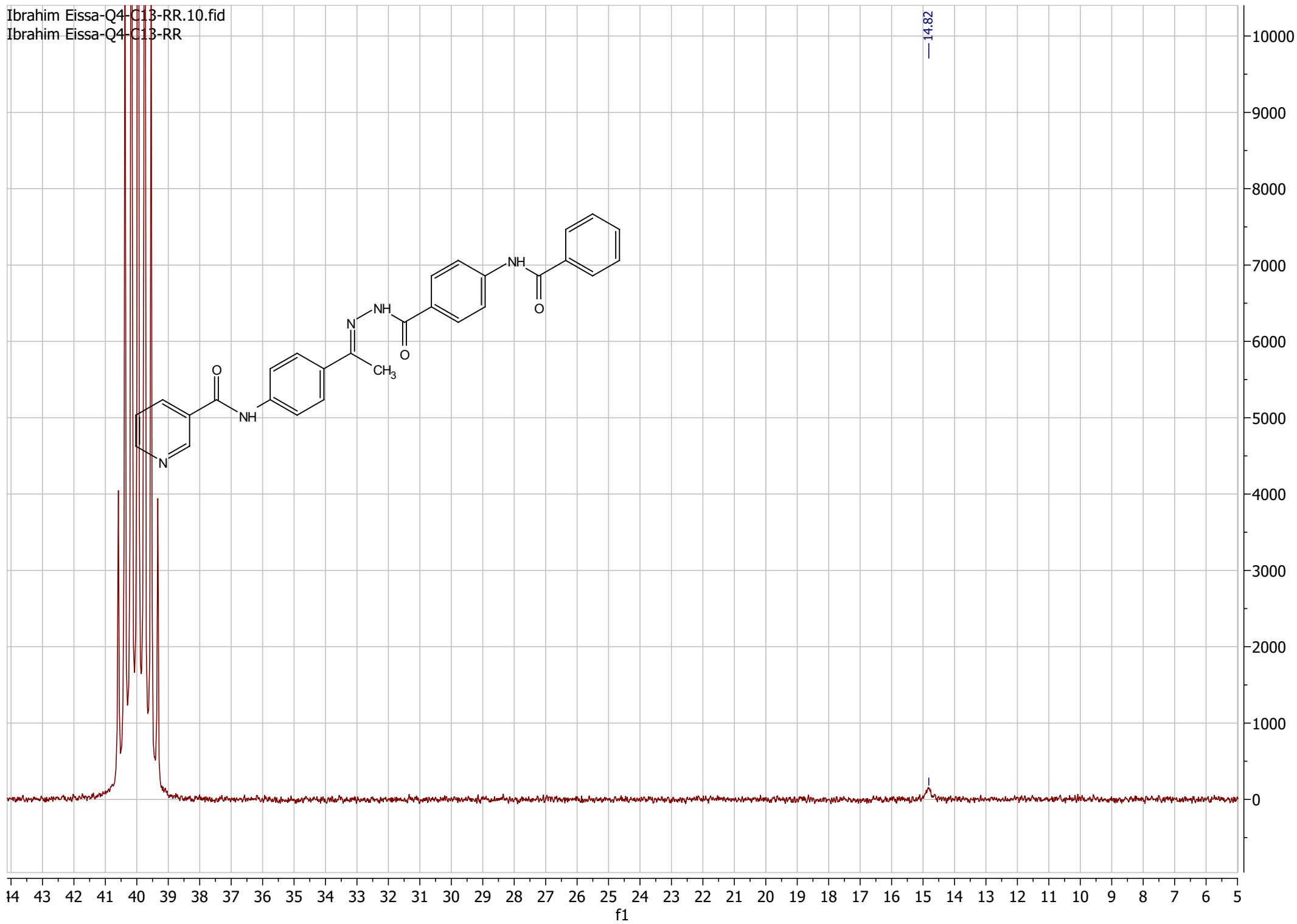
1H NMR of Q4



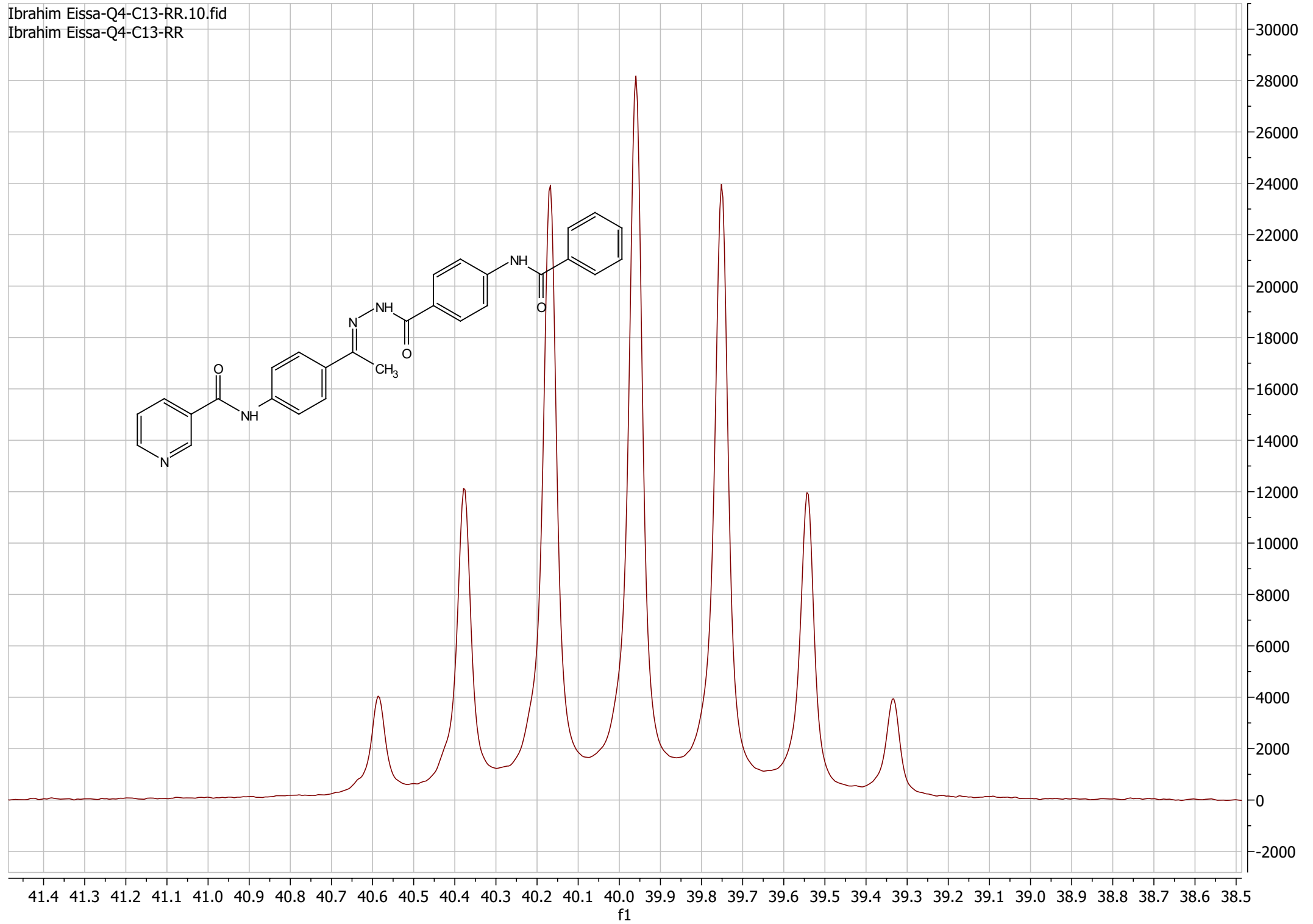
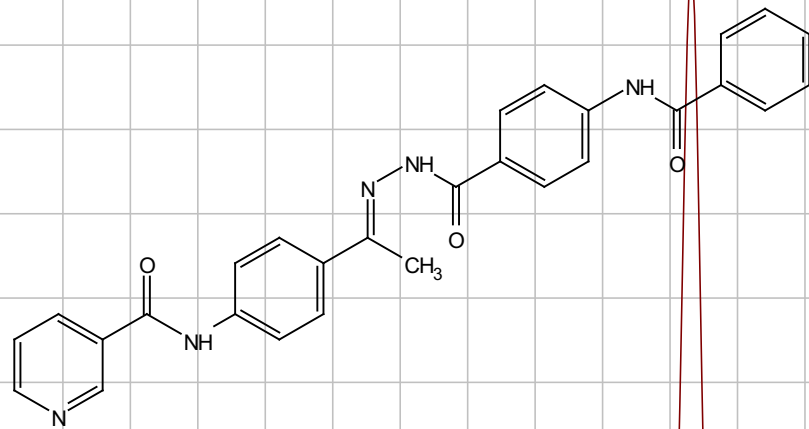
Ibrahim Eissa-Q4-C13-RR.10.fid
Ibrahim Eissa-Q4-C13-RR



Ibrahim Eissa-Q4-C13-RR.10.fid
Ibrahim Eissa-Q4-C13-RR

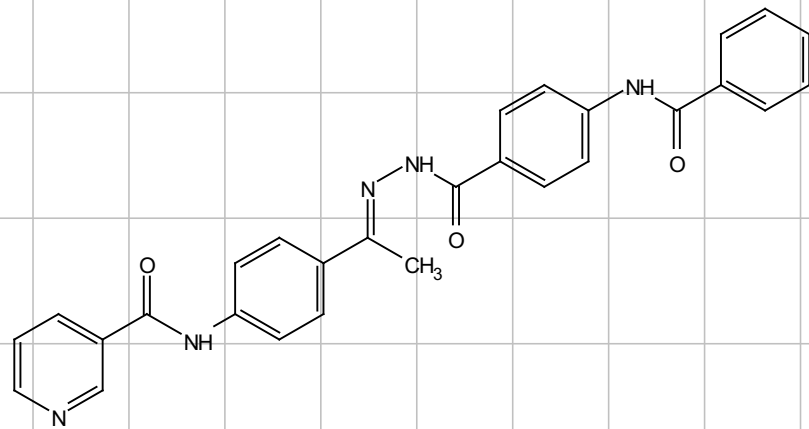


Ibrahim Eissa-Q4-C13-RR.10.fid
Ibrahim Eissa-Q4-C13-RR

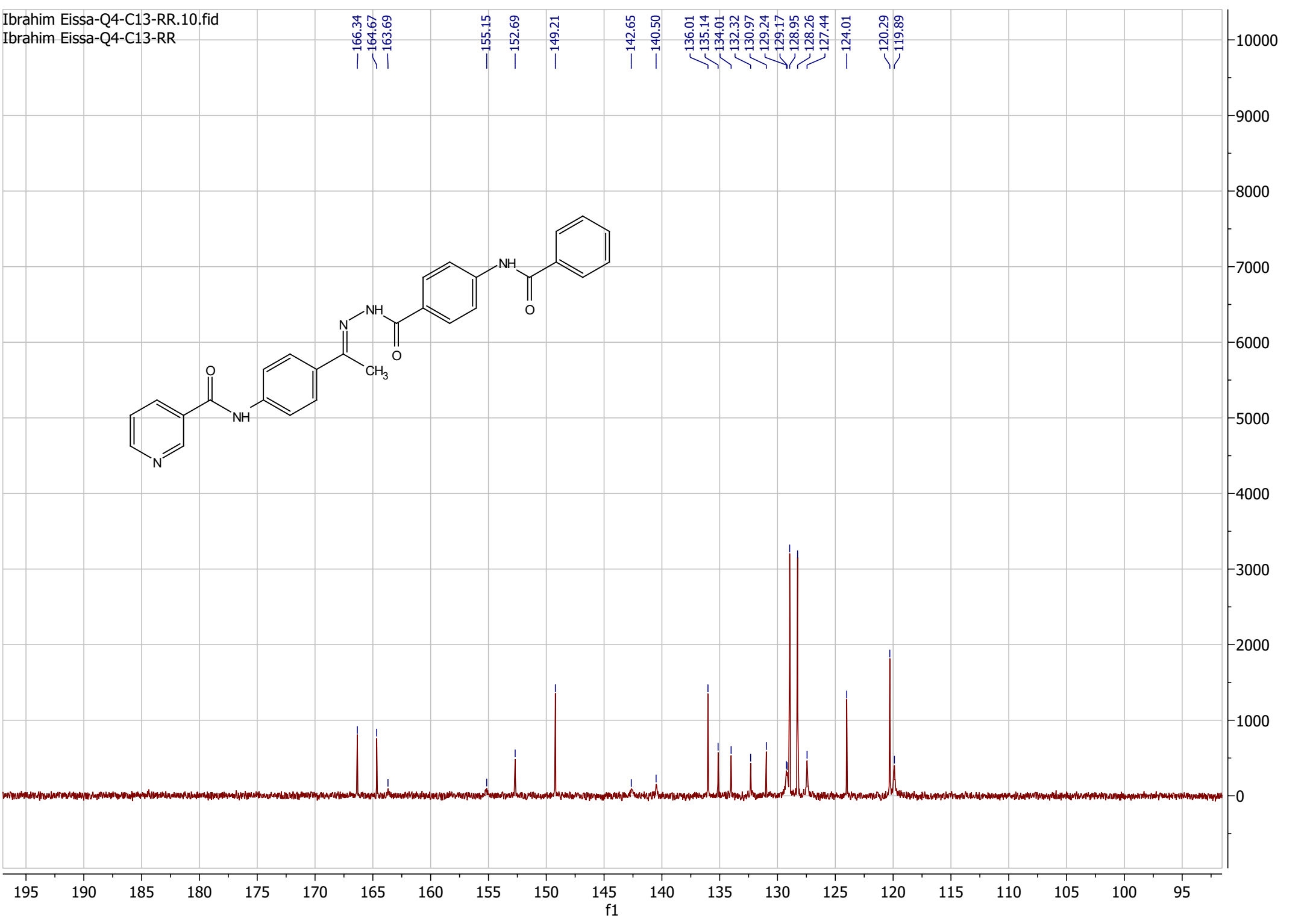


Ibrahim Eissa-Q4-C13-RR.10.fid

Ibrahim Eissa-Q4-C13-RR

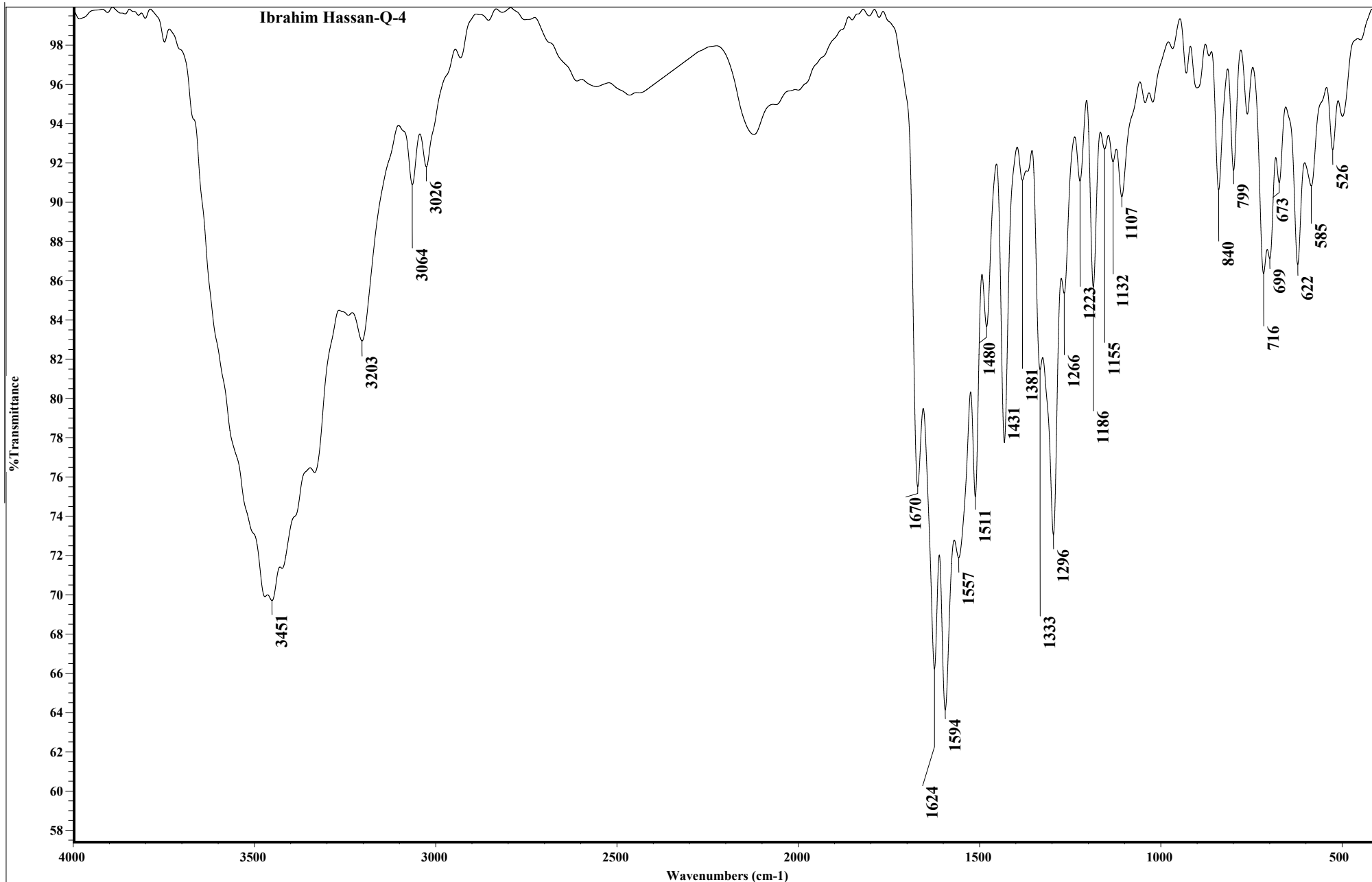


- 166.34
- 164.67
- 163.69
- 155.15
- 152.69
- 149.21
- 142.65
- 140.50
- 136.01
- 135.14
- 134.01
- 132.32
- 130.97
- 129.24
- 129.17
- 128.95
- 128.26
- 127.44
- 124.01
- 120.29
- 119.89



195 190 185 180 175 170 165 160 155 150 145 140 135 130 125 120 115 110 105 100 95

f1



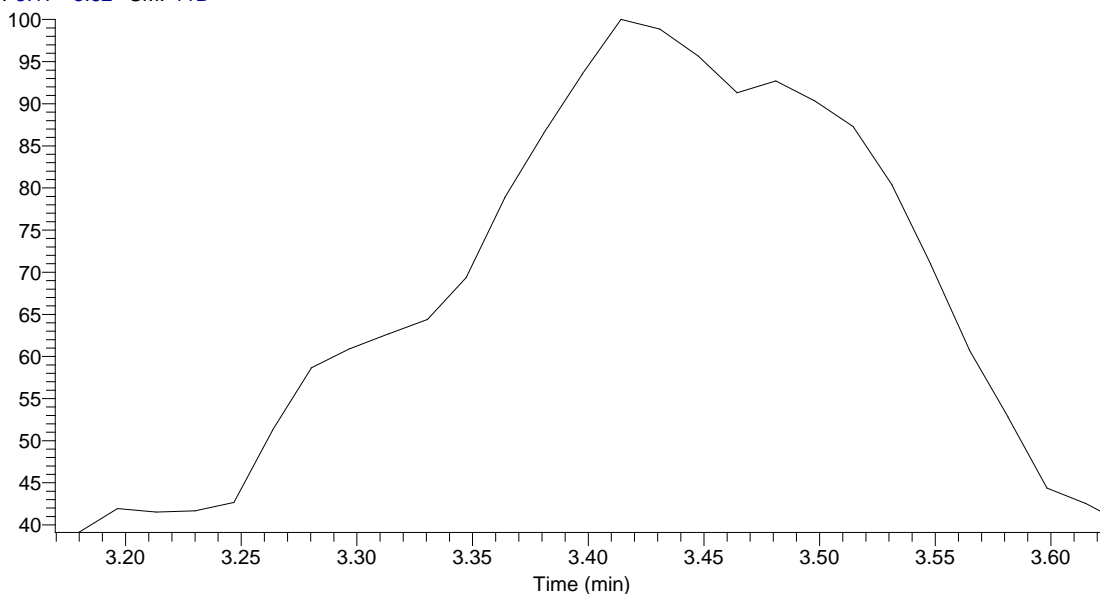
Number of sample scans: 32
Number of background scans: 32
Resolution: 8.000
Sample gain: 1.0
Optical velocity: 0.4747
Aperture: 150.00

ThermoFisher
SCIENTIFIC

Mon Jun 20 12:01:20 2022 (GMT+02:00)

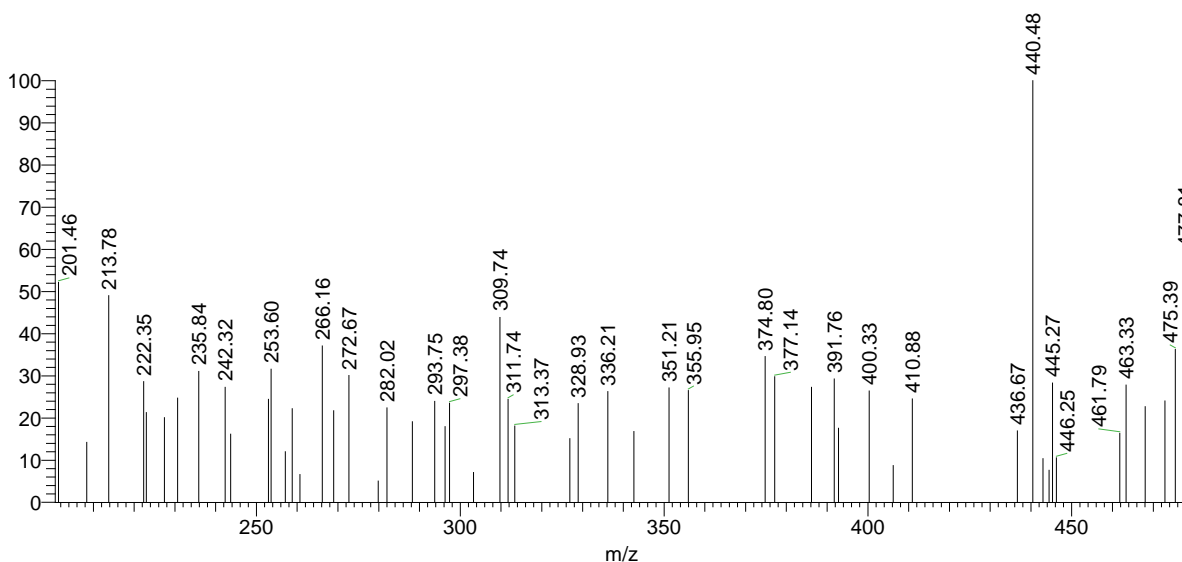
Mansoura University
Faculty of Science
Spectral Analysis Unit
unitofspectra@gmail.com

RT: 3.17 - 3.62 SM: 11B



NL:
9.42E3
TIC MS
IBRAHEIM-
HASSN-
EISA-
Q4_202302
25193831

IBRAHEIM-HASSN-EISA-Q4_20230225193831 #203-205 RT: 3.41-3.45 AV: 3 SB: 26 1.21-1.34, 0.87-1.14 NL: 1.45E2
T: + c EI Full ms [40.00-1000.00]



m/z	Intensity	Relative
-----	-----------	----------

201.46	75.6	52.15
--------	------	-------

208.37	20.6	14.24
--------	------	-------

213.78	71.1	49.04
--------	------	-------

222.35	41.5	28.62
--------	------	-------

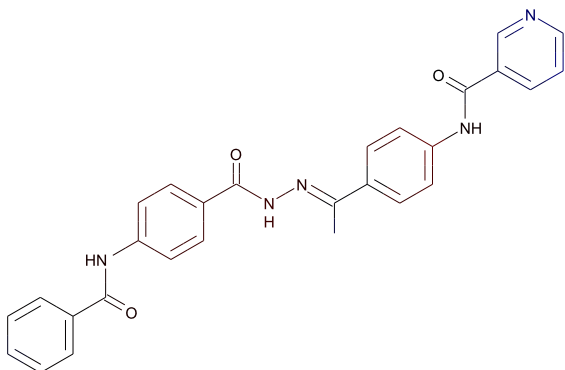
223.00	30.8	21.27
--------	------	-------

227.42	29.1	20.08
--------	------	-------

230.67	35.9	24.72
235.84	45.1	31.08
242.32	39.5	27.27
243.69	23.5	16.19
252.94	35.4	24.43
253.60	45.8	31.60
257.07	17.4	12.00
258.77	32.2	22.19
260.70	9.5	6.58
266.16	53.8	37.10
268.96	31.5	21.73
272.67	43.6	30.06
279.88	7.3	5.03
282.02	32.5	22.40
288.27	27.7	19.12
293.75	34.7	23.96
296.27	26.0	17.96
297.38	34.1	23.51
303.23	10.3	7.10
309.74	63.6	43.87
311.74	35.4	24.43
313.37	26.3	18.11
326.89	21.9	15.08
328.93	34.0	23.41
336.21	38.1	26.27
342.61	24.4	16.80
351.21	39.3	27.12

355.95	38.4	26.51
374.80	50.1	34.57
377.14	43.2	29.82
386.19	39.5	27.26
391.76	42.4	29.25
392.80	25.5	17.56
400.33	38.3	26.42
406.19	12.6	8.71
410.88	35.6	24.52
436.67	24.5	16.92
440.48	145.0	100.00
442.96	15.1	10.39
444.46	11.0	7.60
445.27	41.0	28.30
446.25	15.3	10.57
461.79	23.7	16.38
463.33	40.4	27.83
468.02	32.9	22.71
472.87	34.9	24.06
475.39	52.6	36.26
477.81	85.8	59.1

Toxicity Report



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Non-Mutagen

Probability: 0.714

Enrichment: 1.28

Bayesian Score: -1.18

Mahalanobis Distance: 8.29

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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	GLYBURIDE	6724-53-4	GLIPIZIDE
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.593	0.616	0.632
Reference	PDR 1994	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	PDR 1994

Model Applicability

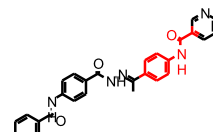
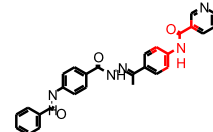
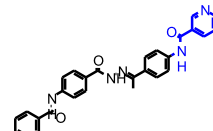
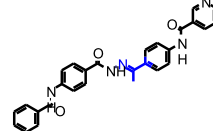
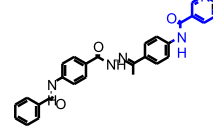
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- All properties and OPS components are within expected ranges.

Feature Contribution

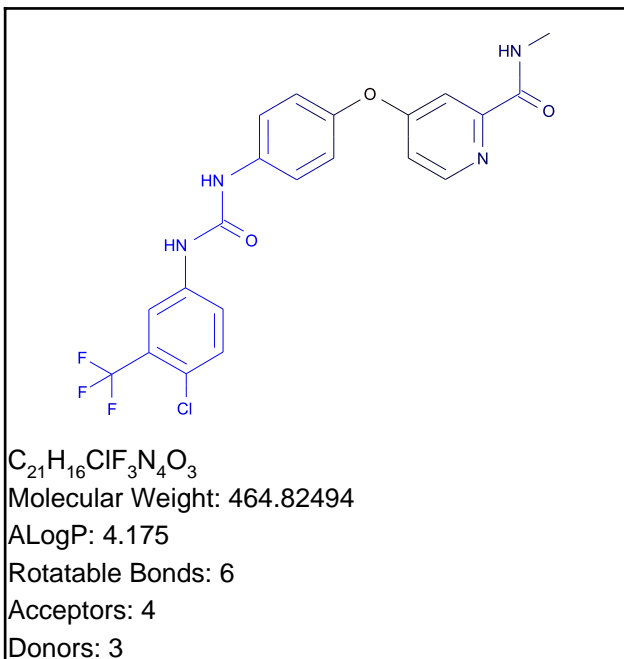
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	555539852	 [*][cH]:[c](:[cH]:[*])C(=O)N[c]1:[cH]:[c H]:[*]:[cH]:[cH]:1	0.447	22 out of 24

SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c] (NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	2096901122	 [*]:[cH]:[c](NC(=O)[c]):[*]:[*]:[cH]:[*]]	0.429	33 out of 37
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1165971455	 [*]NC(=O)[c]1:[cH]:[c H]:[cH]:n:[cH]:1	-0.762	0 out of 2
SCFP_12	-331724199	 [*]N=C(/C)[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	903335088	 [*]NC(=O)[c]1:[cH]:[c H]:[*]:n:[cH]:1	-0.762	0 out of 2

Sorafenib

TOPKAT_Ames_Mutagenicity



Model Prediction

Prediction: Non-Mutagen

Probability: 0.0531

Enrichment: 0.0951

Bayesian Score: -19.7

Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 2.73e-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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

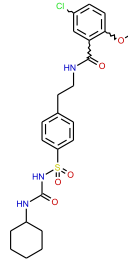
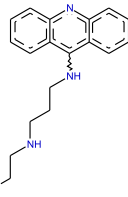
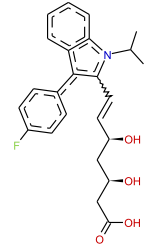
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	GLYBURIDE	38914-96-4	93957-54-1
Structure			
Actual Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Distance	0.590	0.592	0.600
Reference	PDR 1994	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	US Environmental Protection Agency at http://www.epa.gov/NCCT/dsstox/sdf_isscan_external.html

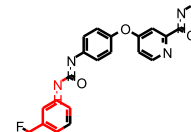
Model Applicability

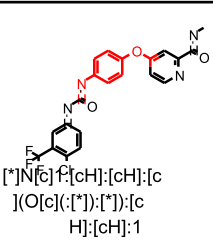
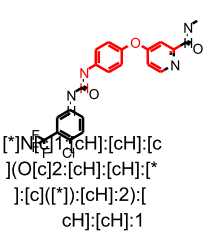
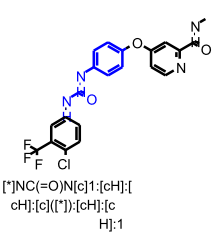
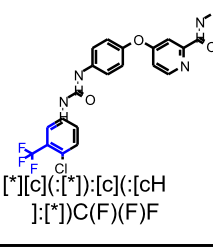
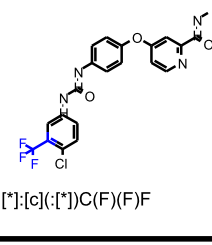
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

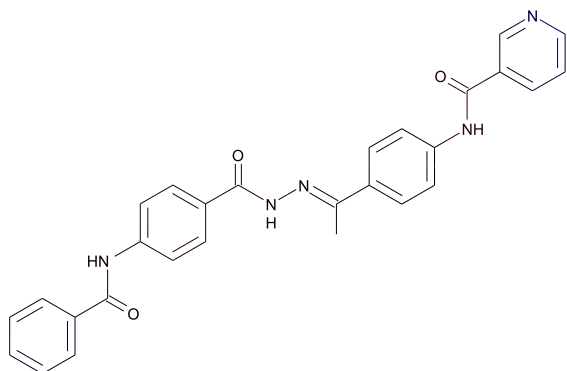
- All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	347281112	 <chem>[*]N(c1c:[cH]:[*]:[c]([*]):[c]([*]:[c]([*]:1)C([*])([*])[*])</chem>	0.337	18 out of 22

SCFP_12	1208843554	 [*]N(c)F[cH]:[cH]:[c (O[c](:[*]:[*]):[c H]:[cH]:1	0.337	6 out of 7
SCFP_12	-1943080297	 [*]N(c)F[cH]:[cH]:[c (O[c]2:[cH]:[cH]:[*]:[c]([*]):[cH]:2):[cH]:[cH]:1	0.304	5 out of 6
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	816802409	 [*]NC(=O)N(c)1:[cH]:[cH]:[c]([*]):[cH]:[c H]:1	-1.82	0 out of 9
SCFP_12	-1903175541	 [*][c]([*]):[c]:[cH]:[*]C(F)(F)F	-1.51	3 out of 30
SCFP_12	-300280774	 [*]:[c]([*])C(F)(F)F	-1.51	3 out of 30



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Non-Toxic

Probability: 0.506

Enrichment: 0.963

Bayesian Score: -1.12

Mahalanobis Distance: 9.15

Mahalanobis Distance p-value: 0.136

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ochratoxin a	Amsacrine	Citreoviridin
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.634	0.666	0.680
Reference	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 7(2):214-20; 1986	Food Chem Toxicol 24(12):1315-20; 1986

Model Applicability

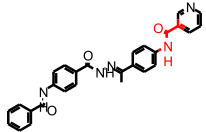
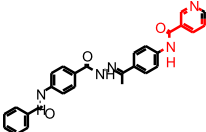
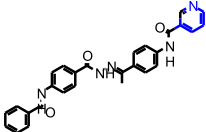
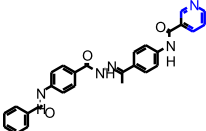
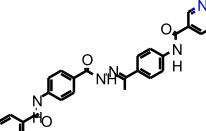
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- OPS PC17 out of range. Value: 2.9151. Training min, max, SD, explained variance: -2.7025, 2.8536, 1.067, 0.0167.

Feature Contribution

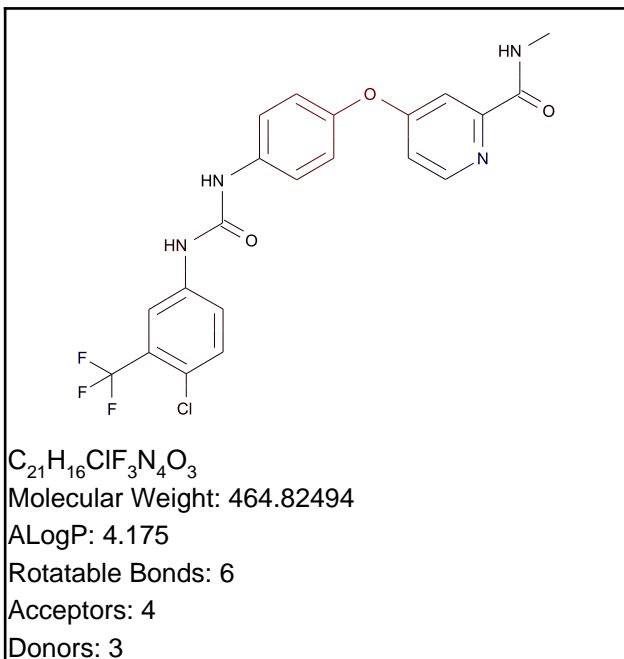
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	282594097	 <chem>[*]NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1</chem>	0.441	3 out of 3

SCFP_6	1257084377	 <chem>[*]NC(=O)[c](:[*]);[*]]</chem>	0.362	14 out of 18
SCFP_6	903335088	 <chem>[*]NC(=O)[c]1:[cH]:[c H]:[*]:n:[cH]:1</chem>	0.271	1 out of 1
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-758850909	 <chem>[*][c]1:[*]:n:[cH]:[c H]:[cH]:1</chem>	-0.646	2 out of 9
SCFP_6	-937094999	 <chem>[*]1:[cH]:[cH]:[cH]:n :[cH]:1</chem>	-0.358	3 out of 9
SCFP_6	-496201075	 <chem>[*]:[cH]:[cH]:n:[*]</chem>	-0.289	8 out of 21

Sorafenib

TOPKAT_Developmental_Toxicity_Potential



Model Prediction

Prediction: Toxic

Probability: 0.592

Enrichment: 1.13

Bayesian Score: 1.15

Mahalanobis Distance: 12.6

Mahalanobis Distance p-value: 2.07e-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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

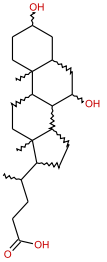
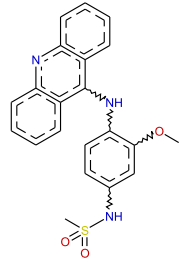
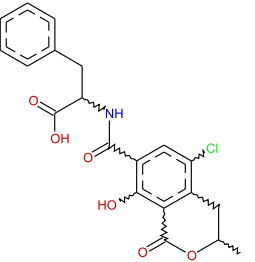
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Chenodioli	Amsacrine	Ochratoxin a
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.631	0.637	0.644
Reference	Arch Int Pharm 246:149-158; 1980	Fundam Appl Toxicol 7(2):214-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976

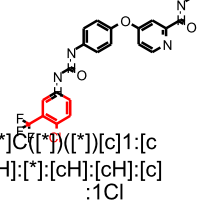
Model Applicability

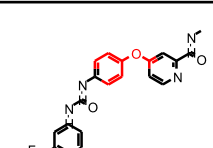
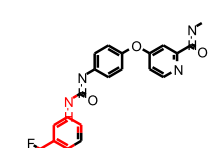
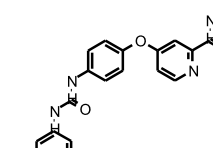
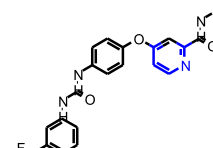
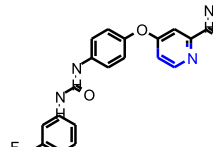
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

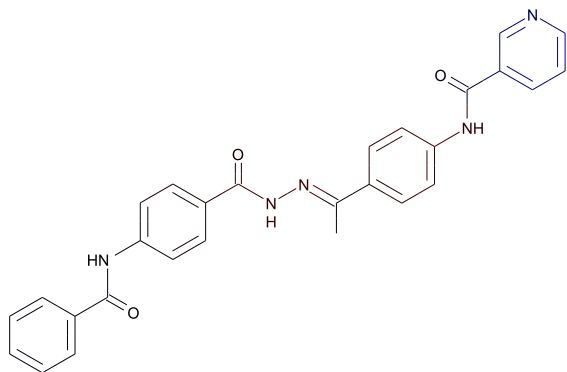
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1559190850	 [*]C([*]) ([*]) [c]1:[c]H:[*]:[cH]:[cH]:[c]::1Cl	0.441	3 out of 3

SCFP_6	-488587948	 [*]:[c]([*])O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.381	2 out of 2
SCFP_6	-347281112	 [*]N(c1:[cH]:[*]:[c]([*]):[c]([*])C([*])([*])[*])	0.381	2 out of 2
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1794974220	 [*]C([*])([*])F	-0.55	2 out of 8
SCFP_6	-937094999	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.358	3 out of 9
SCFP_6	-496201075	 [*]:[cH]:[cH]:n:[*]	-0.289	8 out of 21


 $C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.247

Enrichment: 0.771

Bayesian Score: -0.867

Mahalanobis Distance: 13

Mahalanobis Distance p-value: 0.000533

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Glimepride	Glyburide	Glipizide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.549	0.588	0.634
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

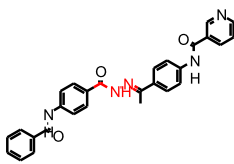
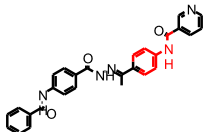
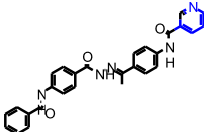
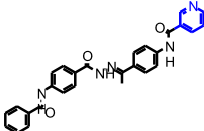
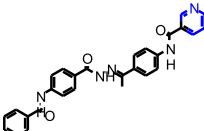
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

Feature Contribution

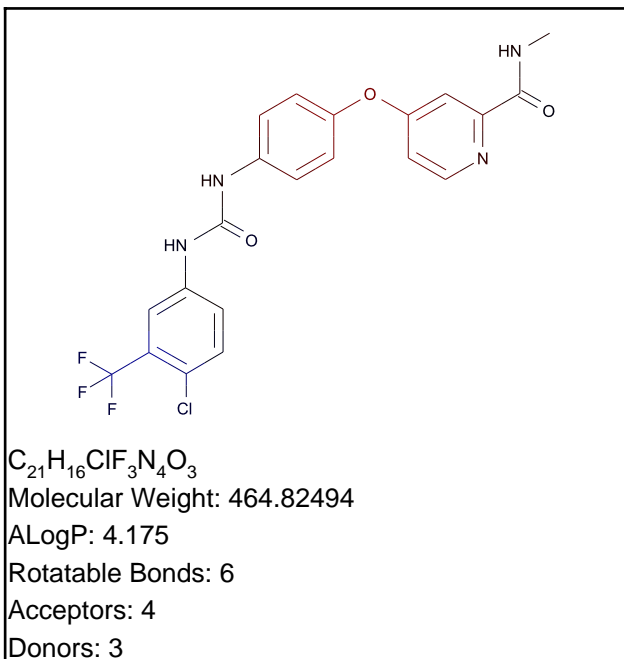
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1087070950	 [*]N=[*]	0.724	10 out of 14

ECFP_6	544048674	 <chem>[*]C(=[*])NN=[*]</chem>	0.617	2 out of 2
ECFP_6	738938915	 <chem>[*]C(=[*])N[c]1:[cH]:[cH]:[cH]:[cH]:1</chem>	0.617	2 out of 2
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	2013347047	 <chem>[*][c]1:[*]:n:[cH]:[cH]:[cH]:1</chem>	-0.805	0 out of 4
ECFP_6	-1818873508	 <chem>[*][c]1:[cH]:[cH]:[cH]:[cH]:n:[cH]:1</chem>	-0.482	0 out of 2
ECFP_6	764951226	 <chem>[*]1:[cH]:[cH]:[cH]:n:[cH]:1</chem>	-0.482	0 out of 2

Sorafenib

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



Model Prediction

Prediction: **Carcinogen**

Probability: 0.257

Enrichment: 0.801

Bayesian Score: -0.321

Mahalanobis Distance: 14.9

Mahalanobis Distance p-value: 4.21e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

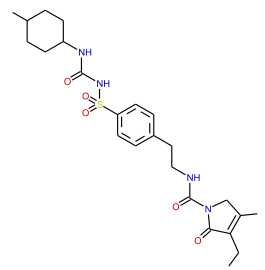
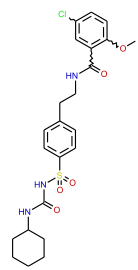
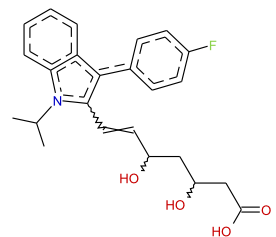
Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
 Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Glimepiride	Glyburide	Fluvastatin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.605	0.615	0.625
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

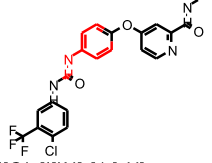
Model Applicability

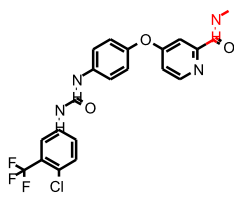
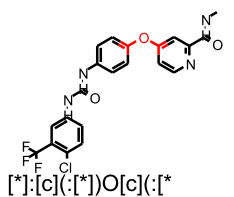
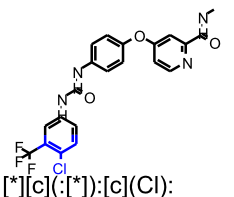
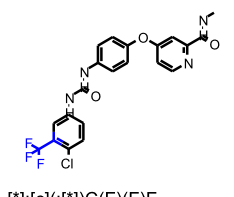
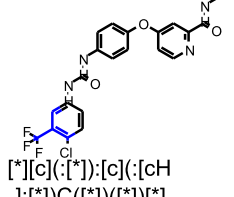
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

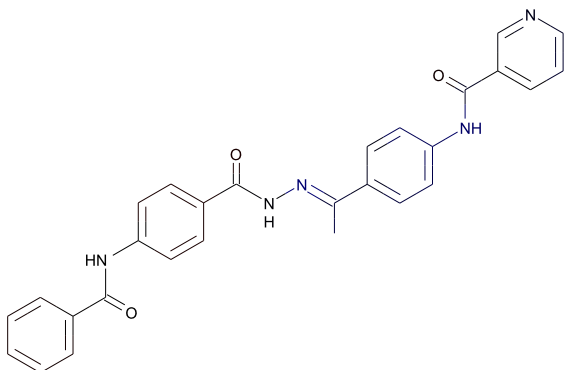
- OPS PC20 out of range. Value: -3.3309. Training min, max, SD, explained variance: -3.1862, 4.4571, 1.28, 0.0167.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	738938915	 <chem>[*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1</chem>	0.617	2 out of 2

ECFP_6	1338334141	 <chem>[*]C(=[*])NC</chem>	0.442	2 out of 3
ECFP_6	1305253718	 <chem>[*]:[c](:[*])O[c](:[*])[*]</chem>	0.424	1 out of 1
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1335691903	 <chem>[*][c](:[*]):[c](Cl):[cH]:[*]</chem>	-0.669	3 out of 22
ECFP_6	1952889961	 <chem>[*]:[c](:[*])C(F)(F)F</chem>	-0.657	0 out of 3
ECFP_6	1336678434	 <chem>[*][c](:[*]):[c](:[cH]1:[*])C([*])([*])[*]</chem>	-0.657	0 out of 3



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.259

Enrichment: 0.633

Bayesian Score: -5.16

Mahalanobis Distance: 12.9

Mahalanobis Distance p-value: 6.51e-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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Glimepiride	Labetalol	Primidolol
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.535	0.761	0.770
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

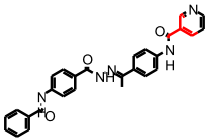
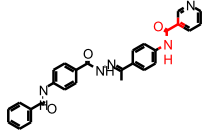
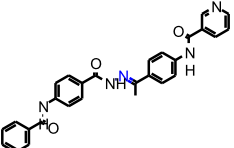
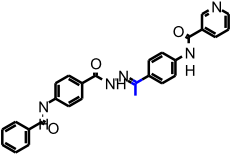
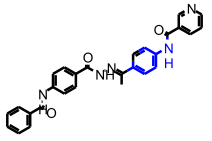
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

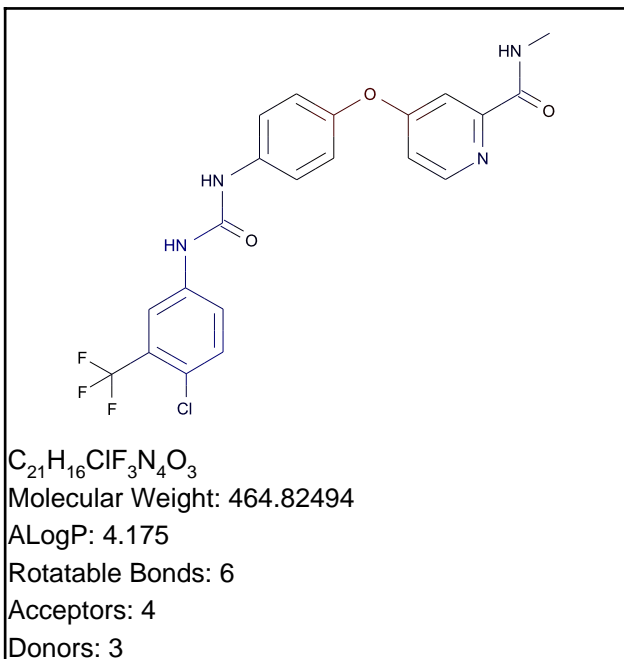
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:1	0.501	2 out of 2

ECFP_4	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.403	6 out of 9
ECFP_4	1430169877	 [*]NC(=O)[c](:[*]):[*]]	0.299	2 out of 3
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	-1087070950	 [*]N=[*]	-1.63	0 out of 10
ECFP_4	866218936	 [*]C(=[*])C	-1.24	0 out of 6
ECFP_4	888054369	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.8	0 out of 3

Sorafenib



Model Prediction

Prediction: Single-Carcinogen

Probability: 0.283

Enrichment: 0.691

Bayesian Score: -3.89

Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.00221

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

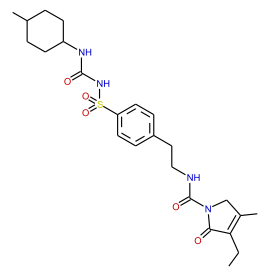
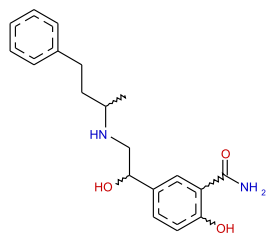
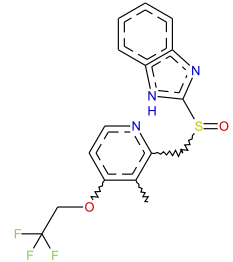
Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

TOPKAT_Mouse_Female_FDA_Single_vs_Multiple

Structural Similar Compounds

Name	Glimepiride	Labetalol	Lansoprazole
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.599	0.808	0.820
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

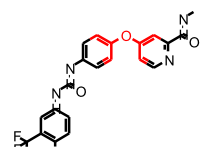
Model Applicability

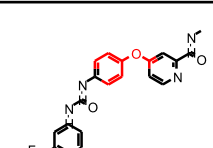
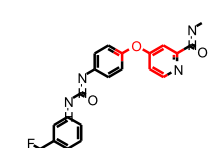
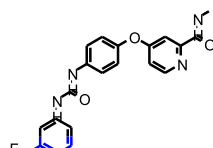
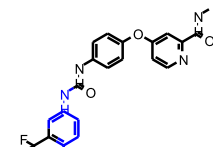
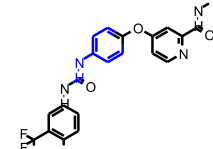
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

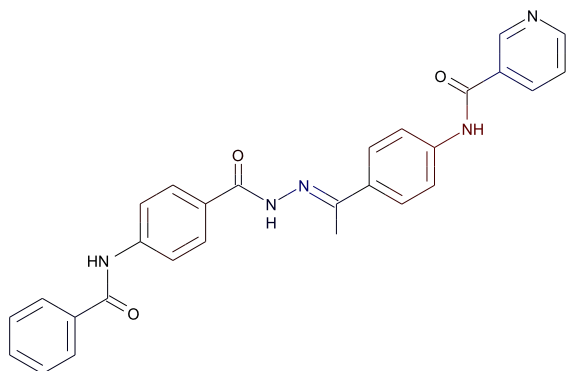
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1336678434: [*][c](:[*]):[c](C([*])([*])([*])):c:[*]
3. Unknown ECFP_2 feature: -1952889961: [*]:[c](:[*])C(F)(F)F

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	-834094296	 [*]:[cH]:[c](O[c](:[c]H):[*]):[c]H):[*]	0.351	1 out of 1

ECFP_4	1407472008	 [*]:[c](*)O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.351	1 out of 1
ECFP_4	143734695	 [*]:[c]1[*]:[cH]:[cH]:[c](O[c](:[*]):[*]):[cH]:1	0.351	1 out of 1
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.8	0 out of 3
ECFP_4	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.8	0 out of 3
ECFP_4	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.597	0 out of 2



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.266

Enrichment: 0.905

Bayesian Score: -1.56

Mahalanobis Distance: 13.9

Mahalanobis Distance p-value: 1.01e-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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Glimepiride	Glyburide	Bicalutamide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.538	0.581	0.612
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

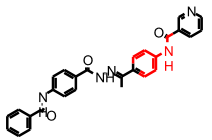
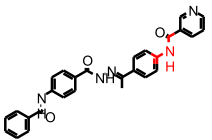
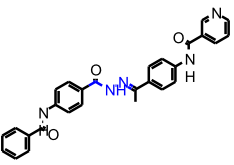
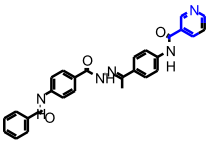
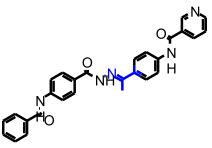
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

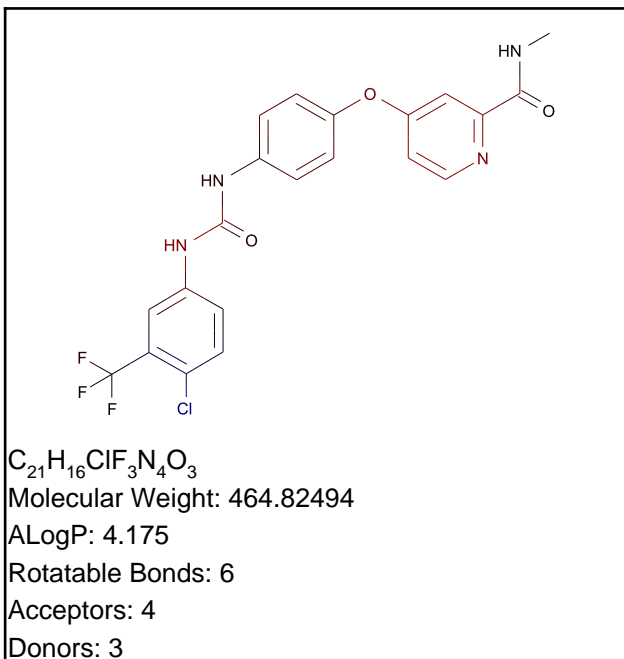
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.77	4 out of 5

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	1294255210	 [*]C(=[*])N[c](:[*]): [*]	0.441	12 out of 28
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885520711	 [*]C(=[*])NN=[*]	-0.839	0 out of 5
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]N=C(/C)[c](:[*]) :[*]	-0.489	3 out of 21

Sorafenib



Model Prediction

Prediction: **Carcinogen**

Probability: 0.444

Enrichment: 1.51

Bayesian Score: 4.21

Mahalanobis Distance: 20.3

Mahalanobis Distance p-value: 1.28e-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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

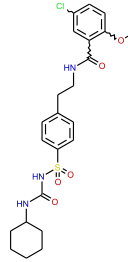
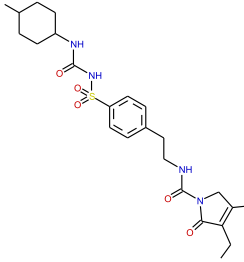
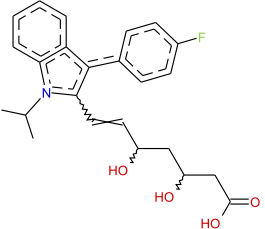
Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen

Structural Similar Compounds

Name	Glyburide	Glimepiride	Fluvastatin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.594	0.599	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

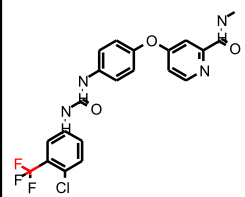
Model Applicability

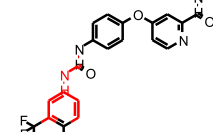
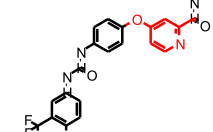
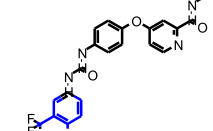
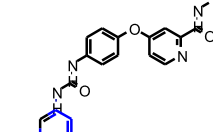
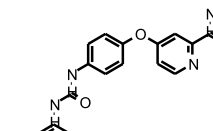
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

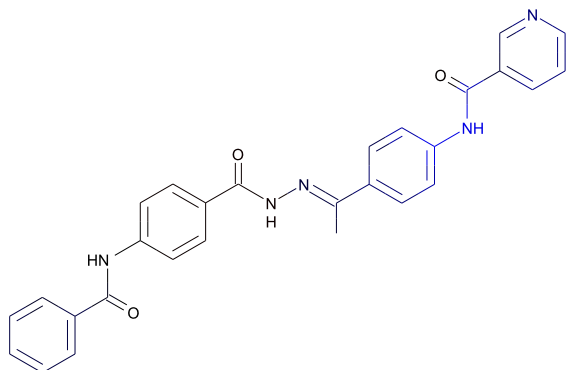
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

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

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	140656626	 [*]O[c]1:[cH]:[cH]:n: [c](:[cH]:1)C(=[*])[*]	0.46	1 out of 1
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	2104062943	 [*]C(=[*])([*])[c]1:[c H]:[*]:[cH]:[cH]:[c] :1Cl	-1.01	1 out of 17
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]): [c](Cl):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59


 $C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.151

Enrichment: 0.503

Bayesian Score: -11.8

Mahalanobis Distance: 15.6

Mahalanobis Distance p-value: 9.06e-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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Glimepiride	Bicalutamide	Primidolol
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.551	0.682	0.795
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

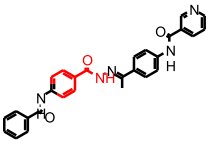
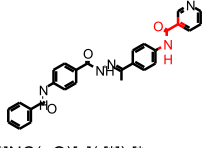
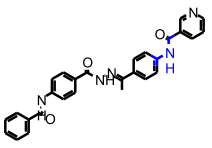
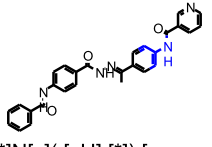
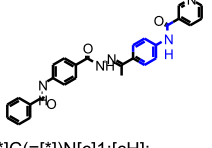
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]
3. Unknown FCFP_2 feature: -885520711: [*]C(=[*])NN=[*]

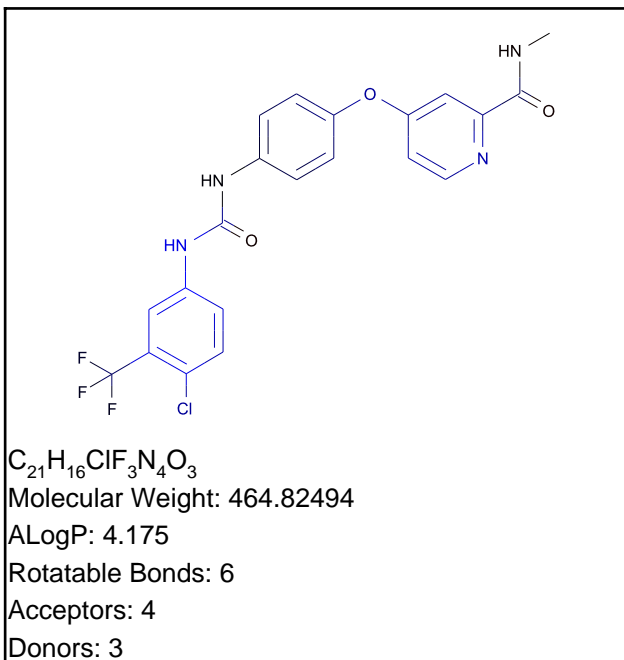
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	547884906	 [*][c]1:[*]:[cH]:[cH] :n:[cH]:1	0.4	1 out of 1

FCFP_12	-581879738	 <chem>[*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1</chem>	0.239	2 out of 4
FCFP_12	-1549103449	 <chem>[*]NC(=O)[c](:[*]):[*]]</chem>	0.168	3 out of 7
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 <chem>[*]C(=[*])N[c](:[*]):[*]</chem>	-1.63	0 out of 12
FCFP_12	590925877	 <chem>[*]N[c](:[cH]:[*]):[cH]:[*]</chem>	-0.998	1 out of 13
FCFP_12	-1838187238	 <chem>[*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1</chem>	-0.859	0 out of 4

Sorafenib



Model Prediction

Prediction: Single-Carcinogen

Probability: 0.139

Enrichment: 0.461

Bayesian Score: -14.7

Mahalanobis Distance: 21.3

Mahalanobis Distance p-value: 4.93e-011

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

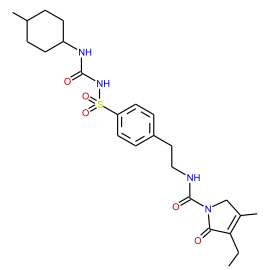
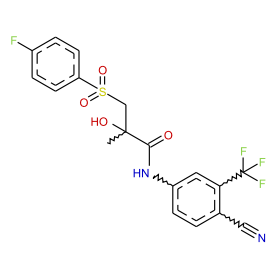
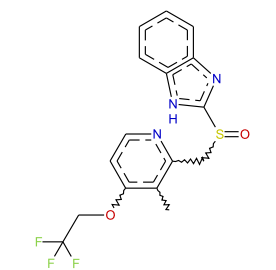
Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

Structural Similar Compounds

Name	Glimepiride	Bicalutamide	Lansoprazole
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.626	0.700	0.866
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

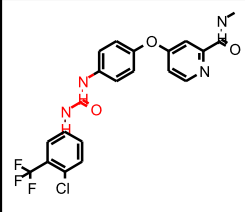
Model Applicability

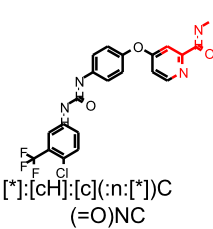
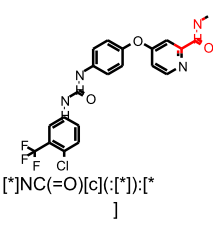
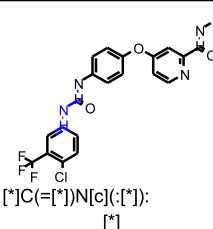
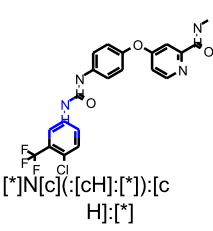
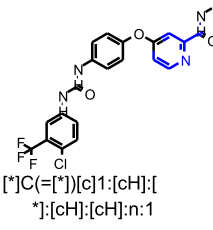
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

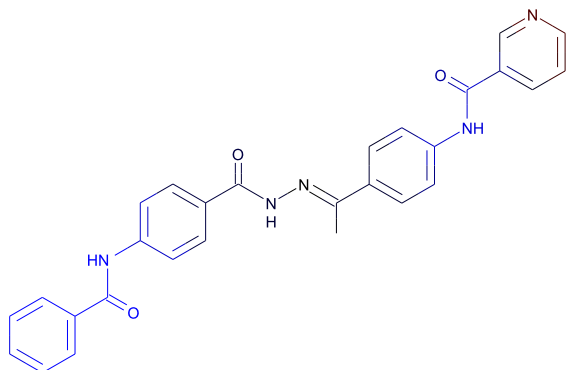
- All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1499521844	 [*]NC(=O)N[*]	0.39	5 out of 9

FCFP_12	-904785030	 [*]:[cH]:[c](:n:[*])C (=O)NC	0.174	1 out of 2
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]]	0.168	3 out of 7
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-1.63	0 out of 12
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	-1462709112	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:n:1	-0.994	0 out of 5



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.132

Enrichment: 0.191

Bayesian Score: -10.3

Mahalanobis Distance: 7.9

Mahalanobis Distance p-value: 0.942

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	ANTHRAQUINONE; 1-((2-HYDROXYETHYL)AMINO)-4-(METHYLAMINO)-
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.724	0.783	0.796
Reference	28ZPAK-;125;72	28ZPAK-;124;72	28ZPAK 245;72

Model Applicability

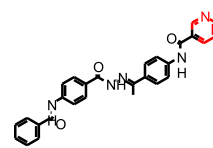
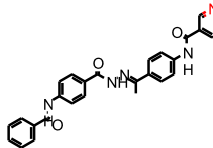
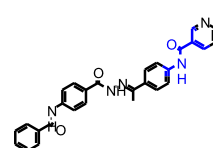
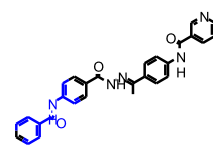
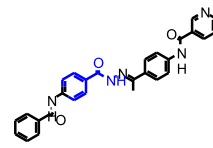
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- All properties and OPS components are within expected ranges.
- Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]

Feature Contribution

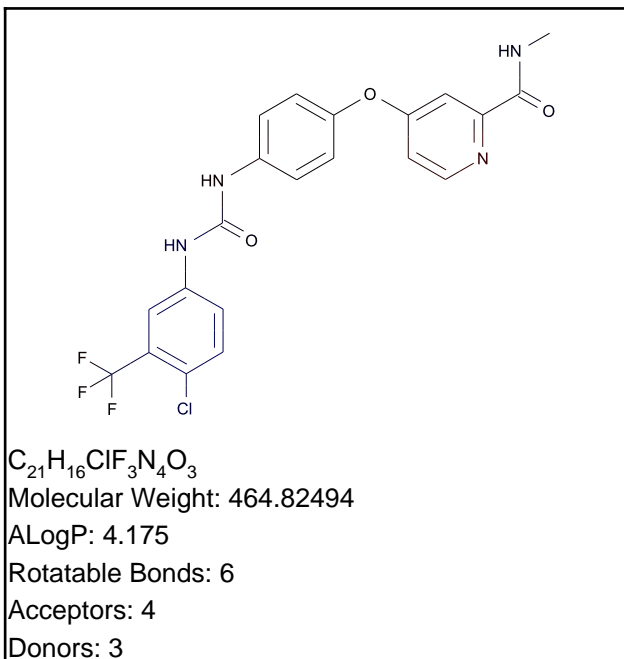
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	547884906	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	0.317	4 out of 4

FCFP_10	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.285	10 out of 11
FCFP_10	-124655670	 [*]:[cH]:[cH]:n:[*]	0.259	14 out of 16
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]):[*]])C(=O)N[c](:[*]):[*]]	-1.29	0 out of 4
FCFP_10	241406177	 [*]:[cH]:[c](NC(=O)[c] 1:[cH]:[cH]:[*]:[cH]):[cH]:1):[cH]:[*]	-1.29	0 out of 4
FCFP_10	-581879738	 [*]NC(=O)[c]1:[cH]:[c] H]:[*]:[cH]:[cH]:1	-1.29	0 out of 4

Sorafenib

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



Model Prediction

Prediction: Mild

Probability: 0.776

Enrichment: 1.13

Bayesian Score: -1.8

Mahalanobis Distance: 8.95

Mahalanobis Distance p-value: 0.537

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

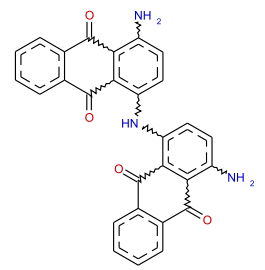
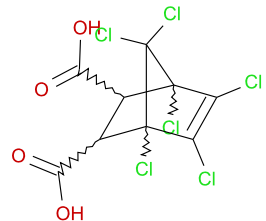
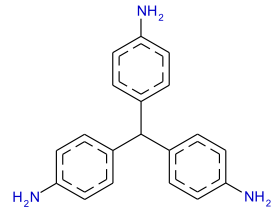
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	METHANE;TRIS(4-AMINOPHENYL)-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe
Distance	0.799	0.816	0.827
Reference	28ZPAK-;125;72	28ZPAK-;92;72	28ZPAK-;73;72

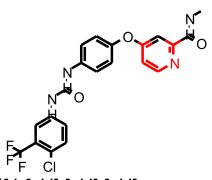
Model Applicability

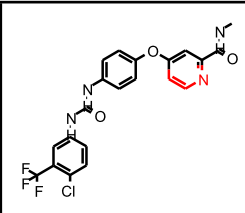
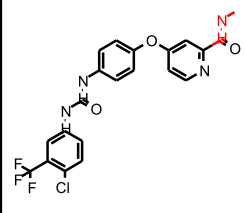
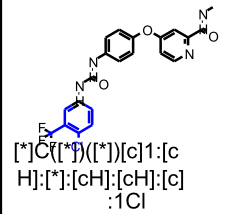
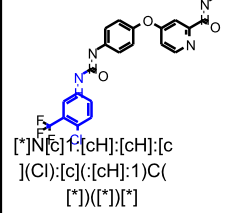
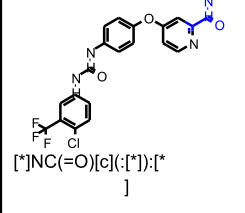
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

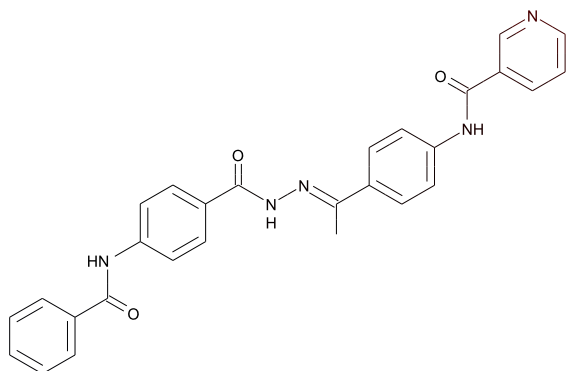
- All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.285	10 out of 11

FCFP_10	-124655670	 <chem>[*]:[cH]:[cH]:n:[*]</chem>	0.259	14 out of 16
FCFP_10	-885550502	 <chem>[*]C(=[*])NC</chem>	0.239	54 out of 64
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	2104062943	 <chem>[*]C([*])([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl</chem>	-0.745	7 out of 24
FCFP_10	-174293376	 <chem>[*]N([c]([cH]:[cH]:[c](Cl):[c](-[cH]:1)C([*])([*])[*])</chem>	-0.507	0 out of 1
FCFP_10	-1549103449	 <chem>[*]NC(=O)[c]([*]):[*]</chem>	-0.504	2 out of 6



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.72

Mahalanobis Distance: 7.1

Mahalanobis Distance p-value: 0.997

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	ANTHRAQUINONE; 1-((2-HYDROXYETHYL)AMINO)-4-(METHYLAMINO)-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.716	0.780	0.781
Reference	28ZPAK-;125;72	28ZPAK 245;72	28ZPAK-;124;72

Model Applicability

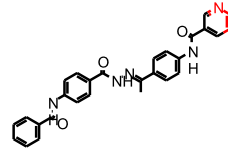
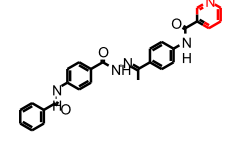
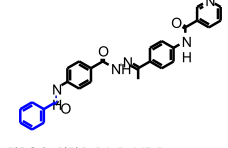
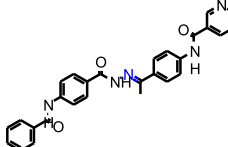
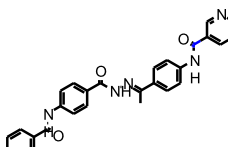
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- All properties and OPS components are within expected ranges.
- Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]

Feature Contribution

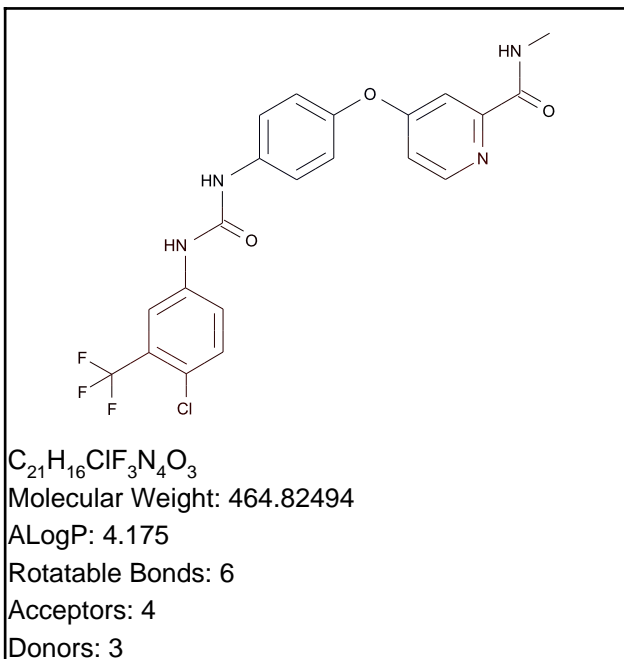
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*]:[cH]:n:[cH]:[*]	0.208	44 out of 44

FCFP_12	-124655670	 [*]:[cH]:[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.194	11 out of 11
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	1	 [*]N=[*]	0	872 out of 1051
FCFP_12	0	 [*]C(=[*])[*]	0	1184 out of 1397

Sorafenib

TOPKAT_Ocular_Irritancy_None_vs_Irritant



Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 3.04

Mahalanobis Distance: 6.28

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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

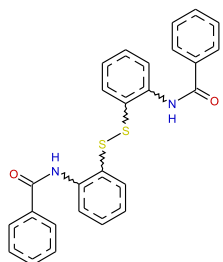
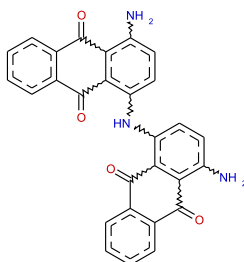
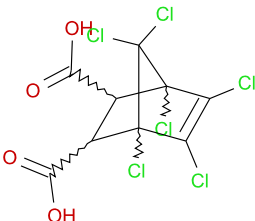
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	BENZANILIDE;2';2'''-DITHIOBIS-	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	5-NORBORNENE-2;3-DICARBOXYLIC ACID;1;4;5;6;7;7'-HEXACHLORO-
Structure			
Actual Endpoint	Non-Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.743	0.791	0.801
Reference	28ZPAK-;173;72	28ZPAK-;125;72	28ZPAK-;92;72

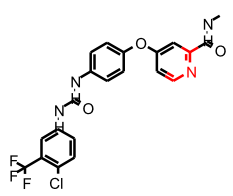
Model Applicability

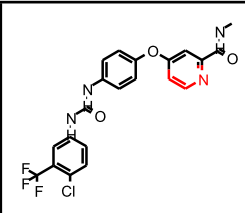
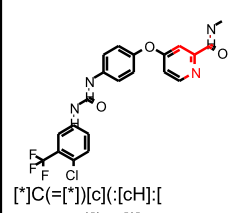
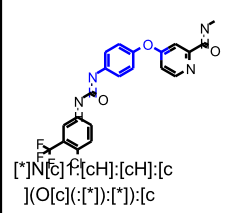
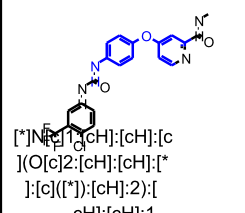
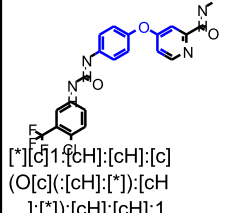
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

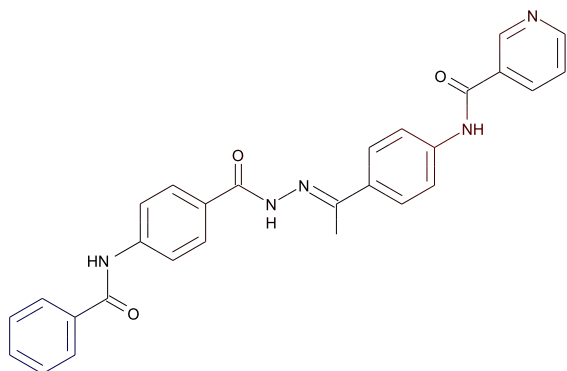
- All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*]:[cH]:n:[cH]:[*]	0.208	44 out of 44

FCFP_12	-124655670	 <chem>[*]:[cH]:[cH]:n:[*]</chem>	0.2	16 out of 16
FCFP_12	-1539132615	 <chem>[*]C(=[*])[c](-[cH]:[*]):n:[*]</chem>	0.197	13 out of 13
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-747629521	 <chem>[*]N([c]([cH]:[cH]:[c](O[c](-[*]):[*]):[cH]:[cH]:1</chem>	-0.268	1 out of 2
FCFP_12	702861189	 <chem>[*]N([c]([cH]:[cH]:[c](O[c]2:[cH]:[cH]:[*]):[c]([*]):[cH]:2):[cH]:[cH]:1</chem>	-0.268	1 out of 2
FCFP_12	859018953	 <chem>[*][c]1:[cH]:[cH]:[c](O[c](-[cH]:[*]):[cH]:[*]):[cH]:[cH]:1</chem>	0	7 out of 9



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: **Carcinogen**

Probability: 0.296

Enrichment: 0.92

Bayesian Score: -0.121

Mahalanobis Distance: 12

Mahalanobis Distance p-value: 0.00462

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Glimepiride	Glyburide	Glipizide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.562	0.603	0.656
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

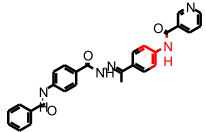
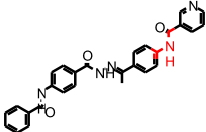
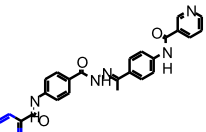
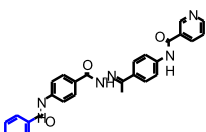
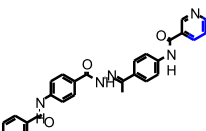
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]N=C(/C)\[c](:[*]):[*]

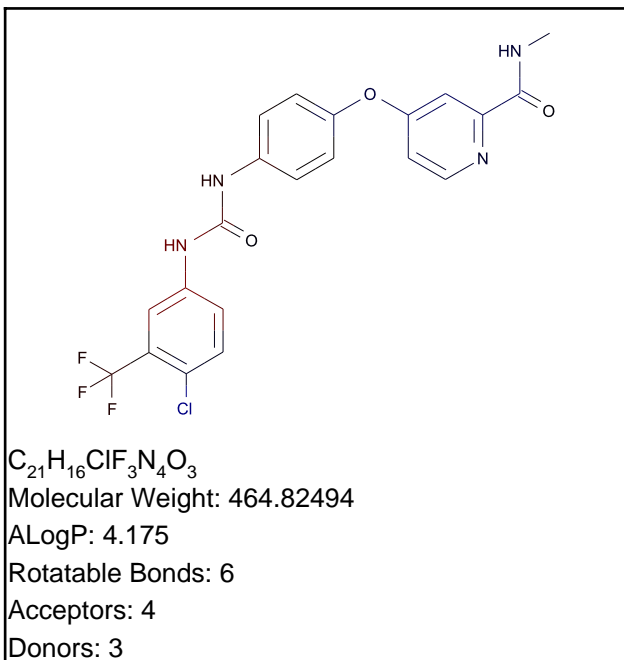
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.613	2 out of 2

ECFP_12	-177077903	 <chem>[*]N[c](:[cH]:[*]):[cH]:[*]</chem>	0.529	6 out of 10
ECFP_12	-1236483485	 <chem>[*]C(=[*])N[c](:[*]):[*]</chem>	0.46	9 out of 17
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1571214559	 <chem>[*]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1</chem>	-0.56	11 out of 64
ECFP_12	-281505363	 <chem>[*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1</chem>	-0.56	11 out of 64
ECFP_12	1997021792	 <chem>[*]:[cH]:[cH]:[cH]:[*]</chem>	-0.296	36 out of 156

Sorafenib



Model Prediction

Prediction: Non-Carcinogen

Probability: 0.236

Enrichment: 0.734

Bayesian Score: -3.76

Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.00229

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

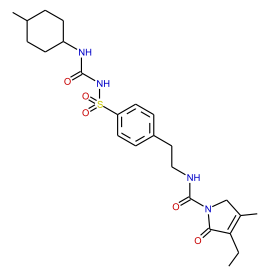
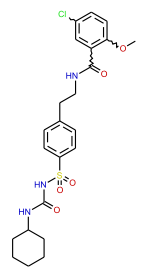
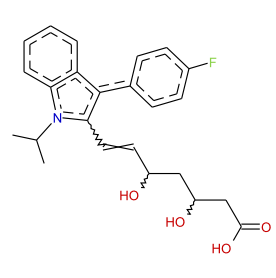
Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

Structural Similar Compounds

Name	Glimepiride	Glyburide	Fluvastatin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.620	0.635	0.635
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

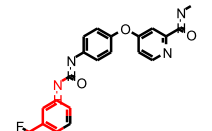
Model Applicability

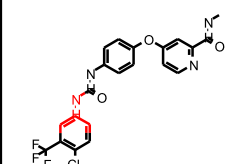
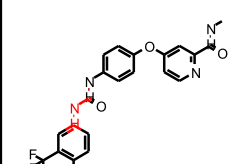
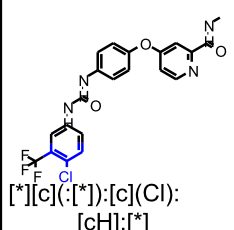
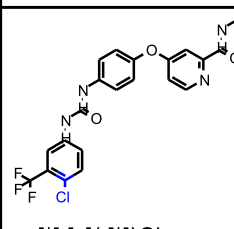
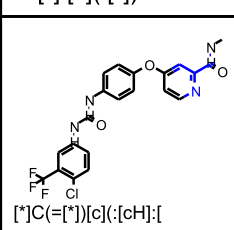
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

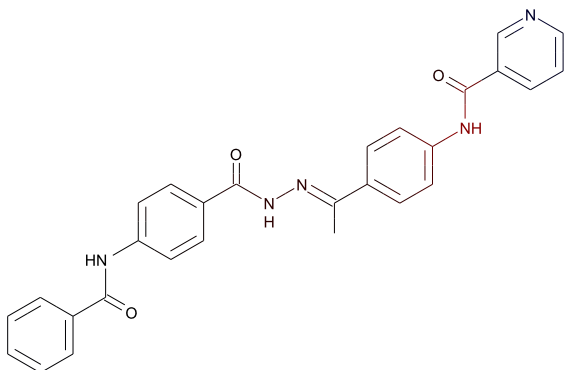
- All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-970385855	 <chem>[*]N(c1c([cH]:[*]:[c]([*]:[c]([*]:[c]([cH]:1)C([*]:[c]([cH]:1)C([*])([*])[*])])])</chem>	0.613	2 out of 2

ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]): [*]	0.46	9 out of 17
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	 [*][c](:[*]):[c](Cl): [cH]:[*]	-1.11	2 out of 26
ECFP_12	99947387	 [*]:[c](:[*])Cl	-0.817	8 out of 62
ECFP_12	1413420509	 [*]C(=[*])[c](:[cH]:[*]):n:[*]	-0.661	0 out of 3


 $C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.605

Enrichment: 1.62

Bayesian Score: 3.33

Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 0.000192

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Bicalutamide	Torsemide	Ursodiol
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.638	0.714	0.738
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

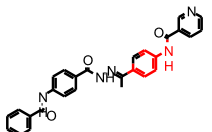
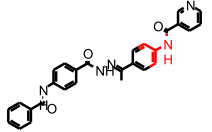
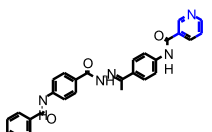
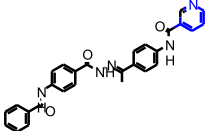
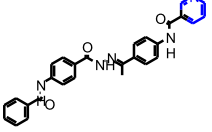
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

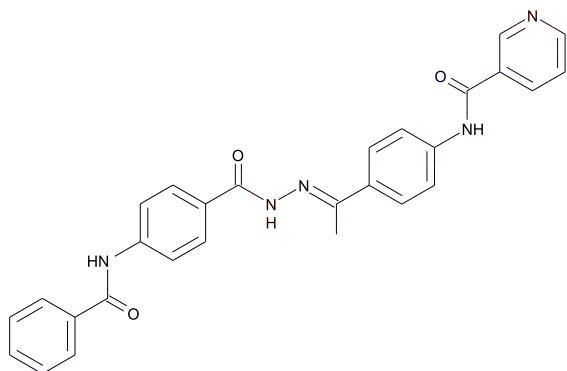
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	1631845520	 [*]C(=[*])N[c]([*]): [*]	0.601	6 out of 9

SCFP_4	-1375926917	 <chem>[*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1</chem>	0.522	6 out of 10
SCFP_4	1205586762	 <chem>[*]N[c](:[cH]:[*]):[cH]:[*]</chem>	0.451	7 out of 13
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	1188429584	 <chem>[*][c]1:[*]:[cH]:[cH]:n:[cH]:1</chem>	-0.666	0 out of 3
SCFP_4	-758850909	 <chem>[*][c]1:[*]:n:[cH]:[cH]:[cH]:1</chem>	-0.489	0 out of 2
SCFP_4	-937094999	 <chem>[*]1:[cH]:[cH]:[cH]:n:[cH]:1</chem>	-0.368	1 out of 6



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.342

Enrichment: 1.02

Bayesian Score: -0.494

Mahalanobis Distance: 14.5

Mahalanobis Distance p-value: 5.99e-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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Glimepiride	Glyburide	Glipizide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.546	0.585	0.632
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

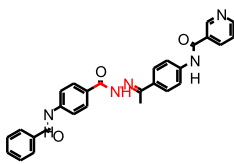
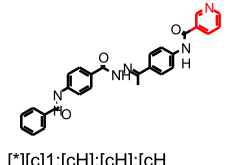
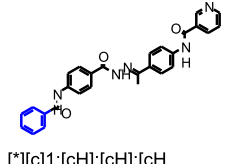
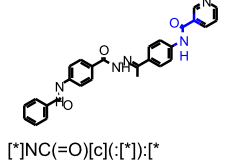
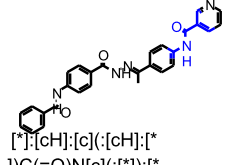
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

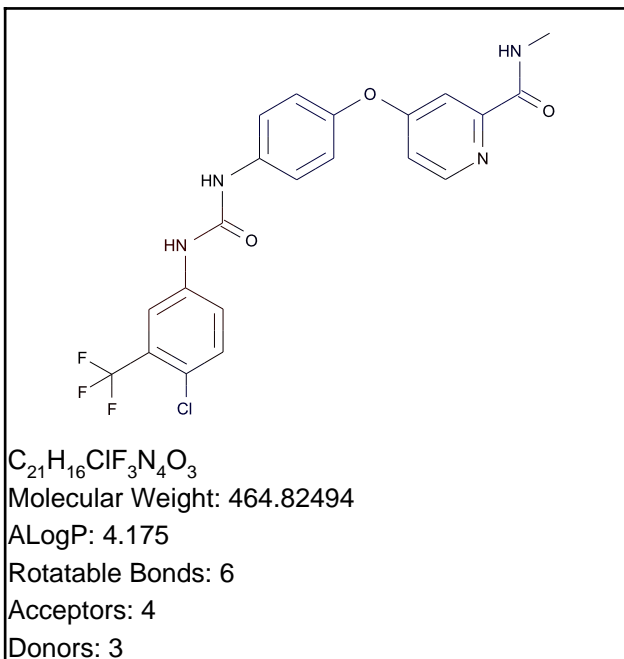
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	347048986	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1	0.615	5 out of 7

SCFP_6	698322229	 <chem>[*]C(=[*])NN=[*]</chem>	0.415	1 out of 1
SCFP_6	-105808146	 <chem>[*][c]1:[cH]:[cH]:[cH]:[cH]:n:[cH]:1</chem>	0.415	1 out of 1
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1653911926	 <chem>[*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1</chem>	-0.504	12 out of 64
SCFP_6	1257084377	 <chem>[*]NC(=O)[c](:[*]):[*]</chem>	-0.436	4 out of 21
SCFP_6	124026986	 <chem>[*]:[cH]:[c](:[cH]):[*])C(=O)N[c](:[*]):[*]</chem>	-0.278	0 out of 1

Sorafenib

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



Model Prediction

Prediction: Non-Carcinogen

Probability: 0.293

Enrichment: 0.878

Bayesian Score: -2.4

Mahalanobis Distance: 17.6

Mahalanobis Distance p-value: 1.1e-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 estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

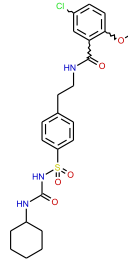
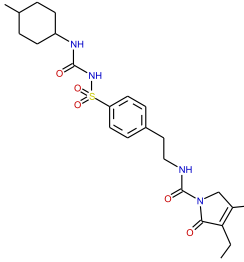
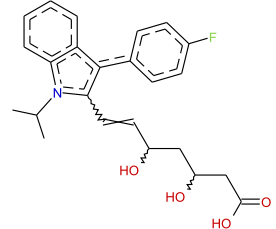
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Glyburide	Glimepiride	Fluvastatin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.593	0.600	0.615
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

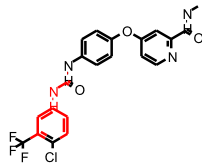
Model Applicability

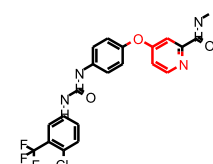
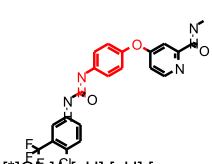
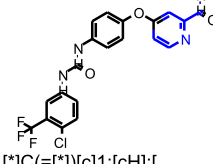
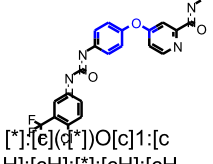
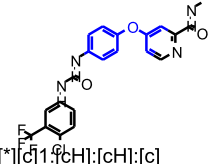
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

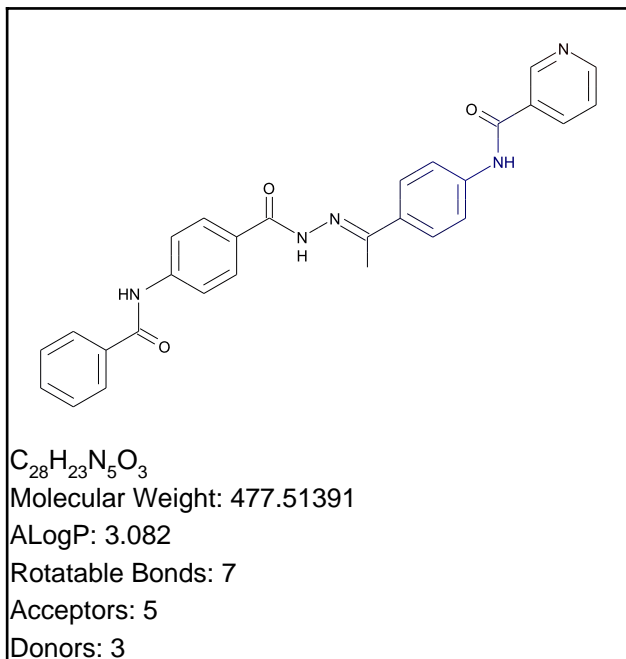
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	347048986	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1	0.615	5 out of 7

SCFP_6	-754059116	 [*]O[c]1:[cH]:[*]:n:[cH]:[cH]:1	0.415	1 out of 1
SCFP_6	-531283893	 [*]O[c]1:[cH]:[cH]:[c](NC(=[*])([*]):[cH]:[cH]:1	0.273	2 out of 4
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-827073191	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:n:1	-0.674	0 out of 3
SCFP_6	-488587948	 [*]:[e](q*)O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.496	0 out of 2
SCFP_6	-975241316	 [*][c]1:[cH]:[cH]:[c](O[c](:[cH]:[*]):[cH]:[*]):[cH]:[cH]:1	-0.496	0 out of 2



Model Prediction

Prediction: Non-Irritant

Probability: 0.911

Enrichment: 0.989

Bayesian Score: -2.62

Mahalanobis Distance: 7.39

Mahalanobis Distance p-value: 0.981

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

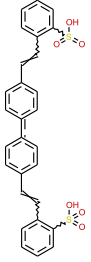
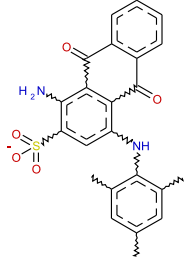
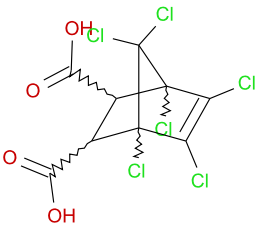
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzenesulfonic acid, 2,2'-(4,4'-biphenyldiylidene)di-, disodium salt	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6-trimethylanilino)-, monosodium salt	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.833	0.865	0.910
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volume(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organické Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	85JCAE "Prehled Prumyslove Toxikologie; Organické Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986

Model Applicability

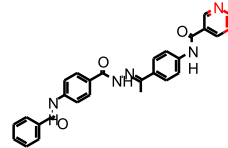
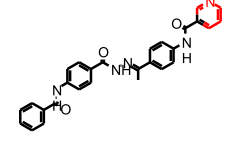
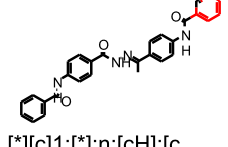
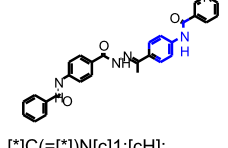
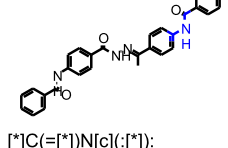
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

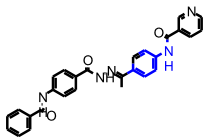
- All properties and OPS components are within expected ranges.

Feature Contribution

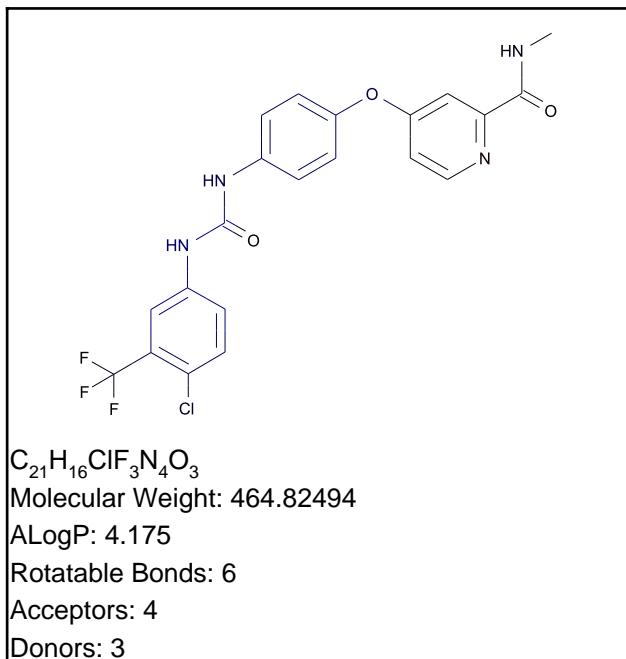
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-124655670	 [*]:[cH]:[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.0772	7 out of 7
FCFP_12	730557100	 [*][c]1:[*]:n:[cH]:[c H]:[cH]:1	0.0756	6 out of 6
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-0.486	12 out of 22

FCFP_12	-773983804	 <chem>O=C1C=NC2=C(C=C1)N(C2)C(=O)Nc3ccncc3</chem>	-0.444	46 out of 79
---------	------------	--	--------	--------------

Sorafenib



Model Prediction

Prediction: Non-Irritant

Probability: 0.264

Enrichment: 0.287

Bayesian Score: -5.23

Mahalanobis Distance: 8.27

Mahalanobis Distance p-value: 0.791

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.

Probability: The estimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.

Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.

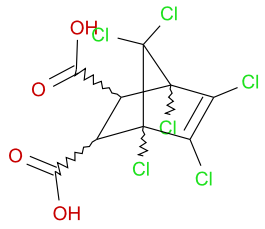
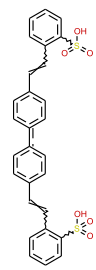
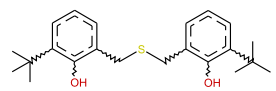
Bayesian Score: The standard Laplacian-modified Bayesian score.

Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD 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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

TOPKAT_Skin_Irritancy_None_vs_Irritant

Structural Similar Compounds

Name	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylene)di-, disodium salt	Sulfide, bis(4-t-butyl-m-cresyl)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.844	0.871	0.884
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organické Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volume(issue)/page/year: 2,193,1973	AMIHBC AMA Archives of Industrial Hygiene and Occupational Medicine. (Chicago, IL) V.2-10, 1950-54. For publisher information, see AEHLAU. Volume(issue)/page/year: 5,311,1952

Model Applicability

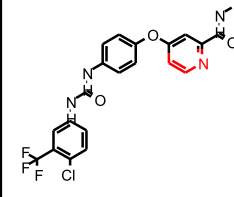
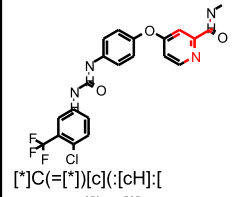
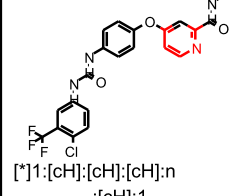
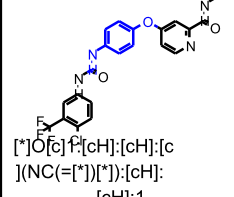
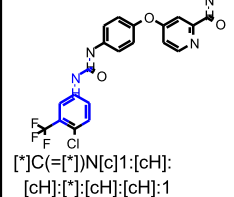
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

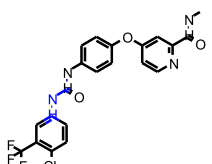
1. All properties and OPS components are within expected ranges.

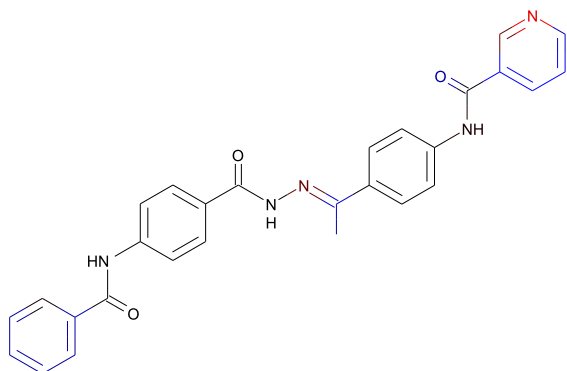
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-124655670	 [*]:[cH]:[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1539132615	 [*]C(=[*])[c](-[cH]:[*]):n:[*]	0.0795	9 out of 9
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.0772	7 out of 7
Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-789307649	 [*]O[c]([*])[cH]:[cH]:[c](NC(=[*])[*]):[cH]:[cH]:1	-1.54	0 out of 4
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12

FCFP_12	1294255210	 <chem>[*]C(=[*])N(c(:[*])):</chem> <chem>[*]</chem>	-0.486	12 out of 22
---------	------------	---	--------	--------------


 $C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: 13.5

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.6

Mahalanobis Distance p-value: 0.000154

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

Structural Similar Compounds

Name	Ochratoxin A	542	Tamoxifen citrate
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	5.05965
Predicted Endpoint (-log C)	3.6353	3.6353	4.24168
Distance	0.816	0.816	0.866
Reference	CPDB	CPDB	CPDB

Model Applicability

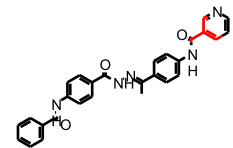
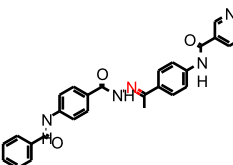
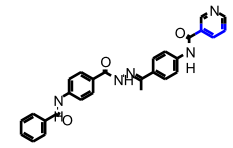
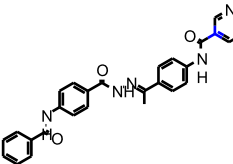
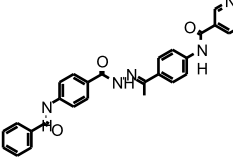
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC23 out of range. Value: -2.85. Training min, max, SD, explained variance: -2.6901, 3.3252, 1.05, 0.0138.
2. Unknown ECFP_2 feature: 128986386: [*]N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

Feature Contribution

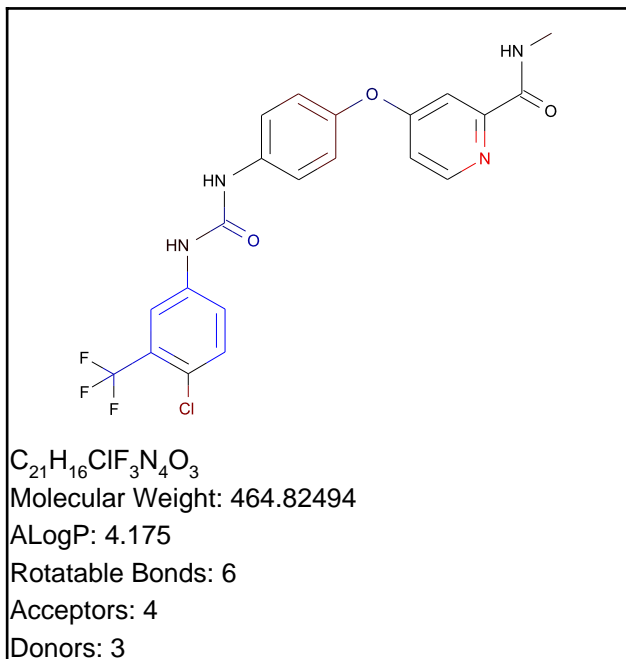
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	-175146122	 <chem>[*]C(=[*])[c](:[cH]:[*]):[cH]:[*]</chem>	0.107
ECFP_6	-1087070950	 <chem>[*]N=[*]</chem>	0.104
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 <chem>[*][c](:[*]):[cH]:[cH]:[*]</chem>	-0.251
ECFP_6	642810091	 <chem>[*][c](:[*]):[*]</chem>	-0.247
ECFP_6	-182236392	 <chem>[*]:[cH]:[*]</chem>	-0.232

Sorafenib

TOPKAT_Carcinogenic_Potency_TD50_Mouse



Model Prediction

Prediction: 19.2

Unit: mg/kg_body_weight/day

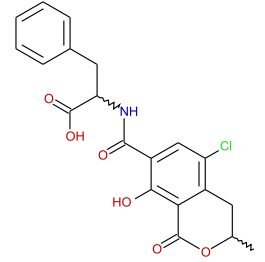
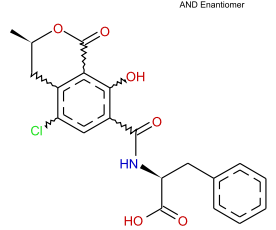
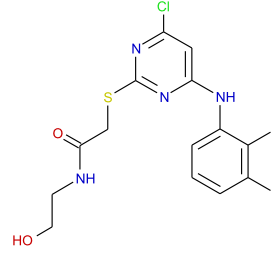
Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 2.94e-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 prediction. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Ochratoxin A	542	4-Chloro-6-(2,3-xylylidino)-2-pyridinylthio(N-b-hydroxy-ethyl) acetamide
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	3.91517
Predicted Endpoint (-log C)	3.6353	3.6353	3.92186
Distance	0.718	0.718	0.738
Reference	CPDB	CPDB	CPDB

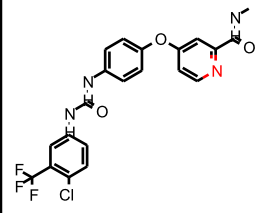
Model Applicability

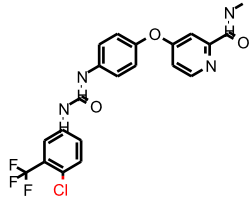
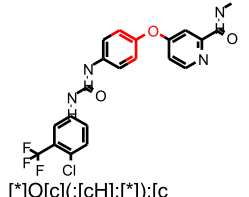
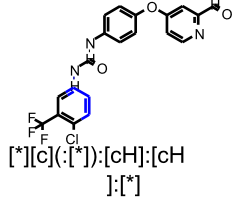
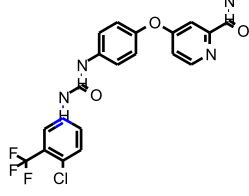
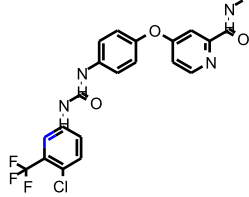
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- All properties and OPS components are within expected ranges.
- Unknown ECFP_2 feature: 1338334141: [*]C(=[*])NC
- Unknown ECFP_2 feature: 1413420509: [*]C(=[*])[c](:n:[*]):c:[*]

Feature Contribution

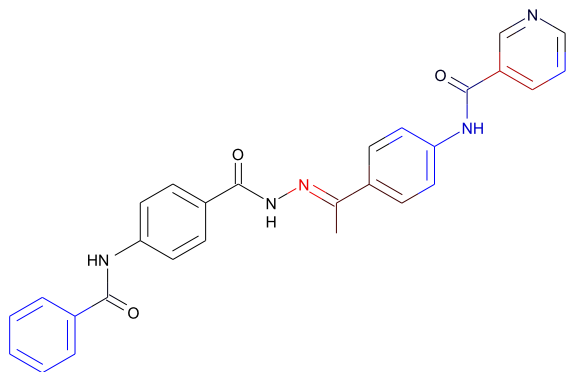
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.0818
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232

Q4

TOPKAT_Carcinogenic_Potency_TD50_Rat


 $C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: 45.4

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13

Mahalanobis Distance p-value: 5.88e-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 prediction. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	913	Fluvastatin	Ochratoxin A
Structure			
Actual Endpoint (-log C)	3.51742	3.51742	6.47264
Predicted Endpoint (-log C)	5.41573	5.41573	5.06501
Distance	0.730	0.730	0.767
Reference	CPDB	CPDB	CPDB

Model Applicability

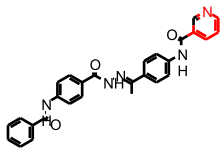
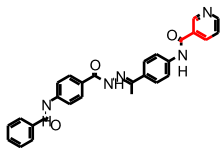
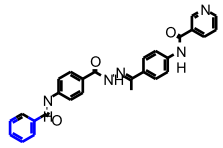
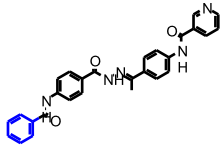
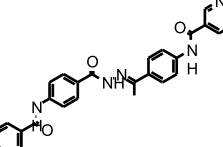
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

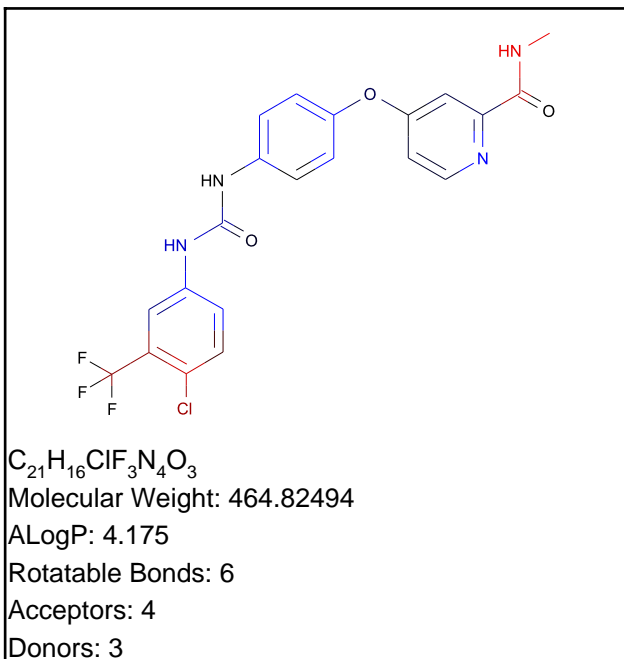
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	<p>[*]N=[*]</p>	0.234

FCFP_6	730557100	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	0.141
FCFP_6	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.137
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	-2093839777	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.378
FCFP_6	16	 [*]:[cH]:[*]	-0.354

Sorafenib

TOPKAT_Carcinogenic_Potency_TD50_Rat



Model Prediction

Prediction: 14.2

Unit: mg/kg_body_weight/day

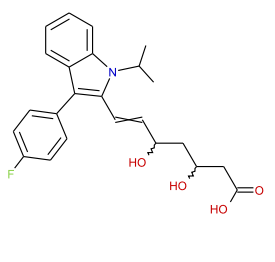
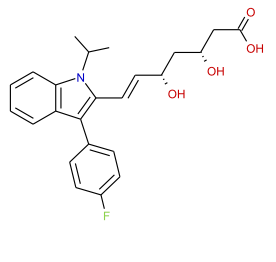
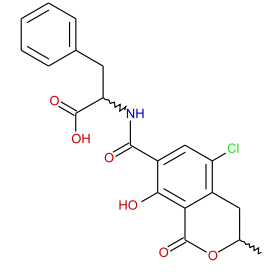
Mahalanobis Distance: 20.4

Mahalanobis Distance p-value: 9.56e-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 prediction. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluvastatin	913	Ochratoxin A
Structure			
Actual Endpoint (-log C)	3.51742	3.51742	6.47264
Predicted Endpoint (-log C)	5.41573	5.41573	5.06501
Distance	0.597	0.597	0.666
Reference	CPDB	CPDB	CPDB

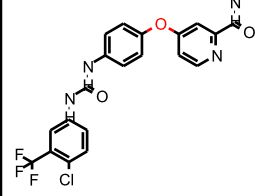
Model Applicability

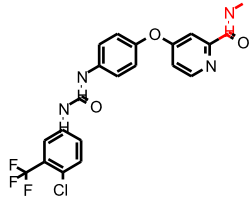
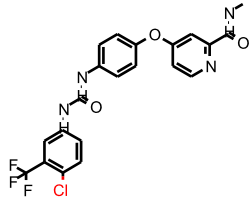
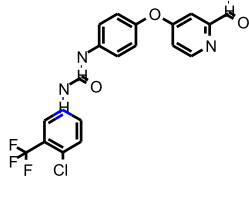
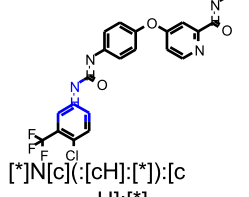
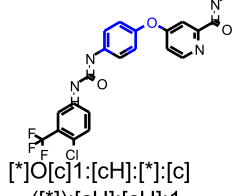
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

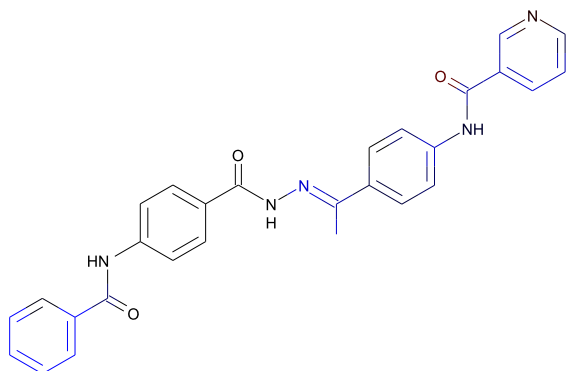
1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1029533685: [*]:[c](:[*])C(F)(F)F

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]N=[*]	0.234

FCFP_6	-885550502	 [*]C(=[*])NC	0.229
FCFP_6	32	 [*]Cl	0.154
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.323
FCFP_6	1674451008	 [*]O[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.233



$C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: 0.336

Unit: g/kg_body_weight

Mahalanobis Distance: 31.9

Mahalanobis Distance p-value: 3.77e-028

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

Structural Similar Compounds

Name	GLYBURIDE	GLIPIZIDE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	4.21661	3.94991	2.28997
Predicted Endpoint (-log C)	4.21035	3.95594	3.52921
Distance	0.787	0.790	0.837
Reference	UPJ-26452	NDA-17583	NTP 411 146

Model Applicability

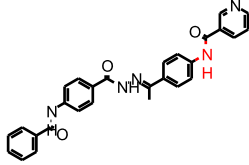
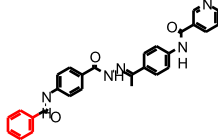
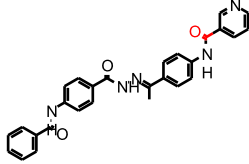
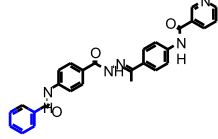
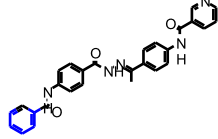
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

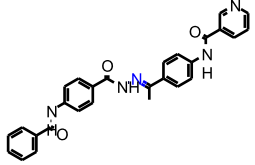
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
3. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]
4. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]
5. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
6. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
7. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
8. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
9. Unknown ECFP_6 feature: 128986386: [*]N=C(/C)[c](:[*]):[*]
10. Unknown ECFP_6 feature: 560380707: [*]NN=C([*])[*]
11. Unknown ECFP_6 feature: 544048674: [*]C(=[*])NN=[*]

Feature Contribution

Top features for positive contribution

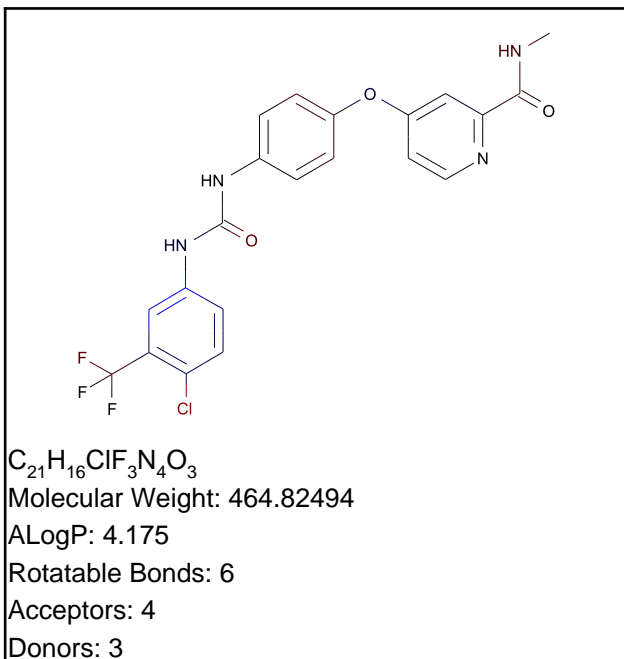
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	3	 [*]N[*]	0.0924
FCFP_6	-2093839777	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.078
ECFP_6	2099970318	 [*]C(=O)[*]	0.0766
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133

FCFP_6	1	 [*]N=[*]	-0.102
--------	---	---	--------

Sorafenib

TOPKAT_Chronic_LOAEL



Model Prediction

Prediction: 0.00483

Unit: g/kg_body_weight

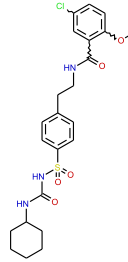
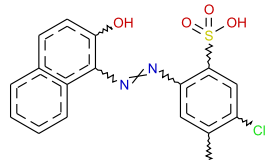
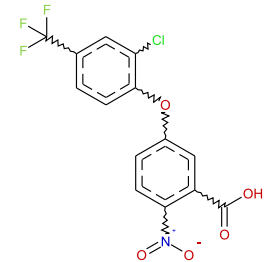
Mahalanobis Distance: 30

Mahalanobis Distance p-value: 1.21e-024

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

Structural Similar Compounds

Name	GLYBURIDE	D & C RED 9	SODIUM ACIFLUORFEN
Structure			
Actual Endpoint (-log C)	4.21661	3.87715	4.16036
Predicted Endpoint (-log C)	4.21035	3.6546	4.65915
Distance	0.636	0.722	0.736
Reference	UPJ-26452	NTP REPORT # 225	EPA COVER SHEET 0192;891101;(1)

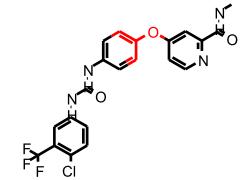
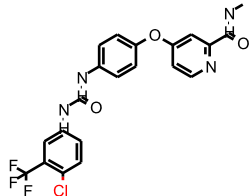
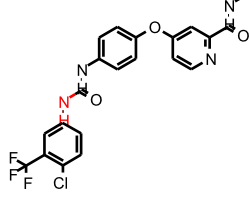
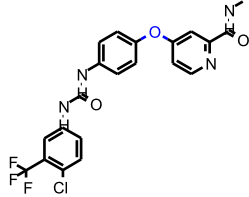
Model Applicability

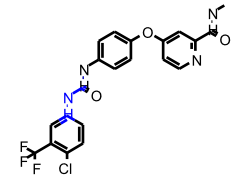
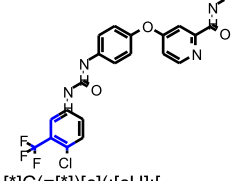
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -1046436026: [*]F
3. Unknown ECFP_6 feature: 99947387: [*]:c(:[*])Cl
4. Unknown ECFP_6 feature: 226796801: [*]C([*])([*])F
5. Unknown ECFP_6 feature: 1305253718: [*]:c(:[*])O[c(:[*]):[*]]
6. Unknown ECFP_6 feature: -677309799: [*]c(:[*]):n:[cH]:[*]
7. Unknown ECFP_6 feature: 1338334141: [*]C(=[*])NC
8. Unknown ECFP_6 feature: -177077903: [*]N[c(:[cH]:[*]):[cH]:[*]]
9. Unknown ECFP_6 feature: 1336678434: [*][c(:[*]):c(:[cH]:[*])C([*])([*])[*]
10. Unknown ECFP_6 feature: -649580166: [*]NC(=O)N[*]
11. Unknown ECFP_6 feature: -1952889961: [*]:c(:[*])C(F)(F)F
12. Unknown ECFP_6 feature: 1413420509: [*]C(=[*])[c(:[cH]:[*]):n:[*]
13. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]
14. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c(:[*]):[*]
15. Unknown ECFP_6 feature: 864287155: [*]NC

Feature Contribution

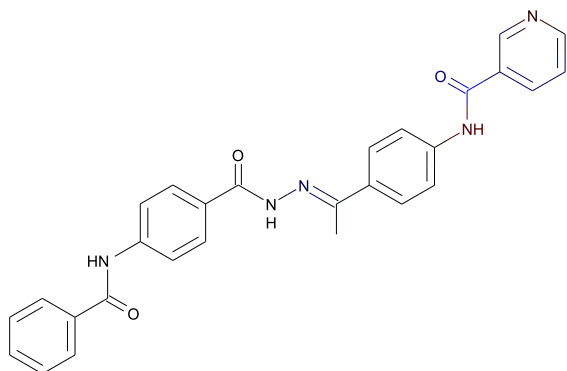
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-176455838	 <chem>[*]O[c](:[cH]:[*]):[cH]:[*]</chem>	0.106
FCFP_6	32	 <chem>[*]Cl</chem>	0.101
FCFP_6	3	 <chem>[*]N[*]</chem>	0.0924
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 <chem>[*]N=[*]</chem>	-0.102

<p>ECFP_6</p>	<p>1236483485</p>	 <p><chem>[*]C(=O)N(c1ccc(Oc2cc(O)nc(=O)n2)cc1)c3ccc(Oc4cc(O)nc(=O)n4)cc3</chem></p>	<p>-0.0747</p>
<p>FCFP_6</p>	<p>203677720</p>	 <p><chem>[*]C(=O)[c]([cH]:[*]):[*]</chem></p>	<p>-0.0713</p>

Q4

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed

C₂₈H₂₃N₅O₃

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: 0.117

Unit: g/kg_body_weight

Mahalanobis Distance: 9.42

Mahalanobis Distance p-value: 0.000352

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

Structural Similar Compounds

Name	SALICYLAZOSULFAPYRIDINE	FUROSEMIDE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	3.375	4.04236	2.30052
Predicted Endpoint (-log C)	2.80292	2.8614	3.55333
Distance	0.711	0.831	0.868
Reference	NCI/NTP TR-457	NCI/NTP TR-356	NCI/NTP TR-411

Model Applicability

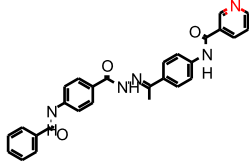
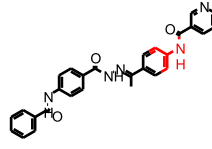
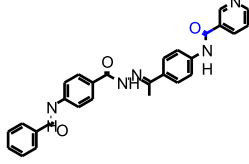
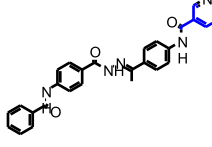
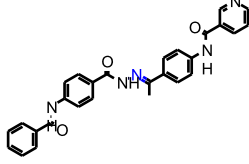
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

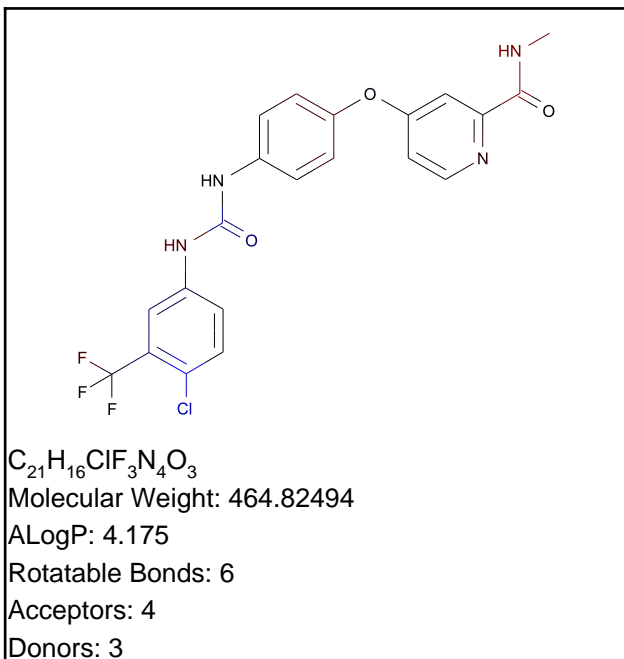
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3	 [*]N[*]	0.0737

FCFP_2	17	 [*]:n:[*]	0.0441
FCFP_2	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	0.00762
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.0829
FCFP_2	1	 [*]N=[*]	-0.0796

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



Model Prediction

Prediction: 0.0885

Unit: g/kg_body_weight

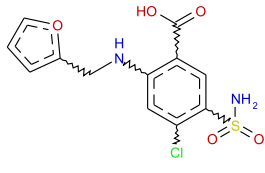
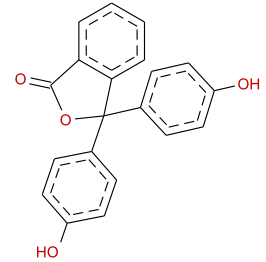
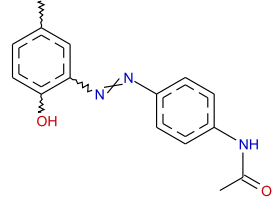
Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 1.76e-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 prediction. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	FUROSEMIDE	PHENOLPHTHALEIN	DISPERSE YELLOW 3
Structure			
Actual Endpoint (-log C)	4.04236	2.20184	2.77703
Predicted Endpoint (-log C)	2.8614	2.8857	2.80195
Distance	0.741	0.780	0.799
Reference	NCI/NTP TR-356	NCI/NTP TR-465	NCI/NTP TR-222

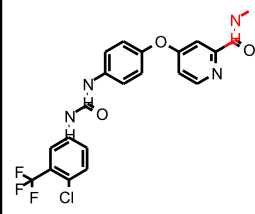
Model Applicability

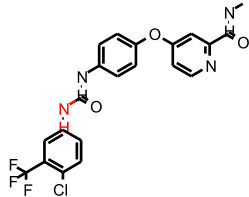
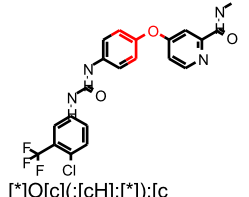
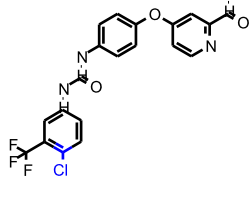
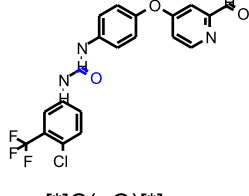
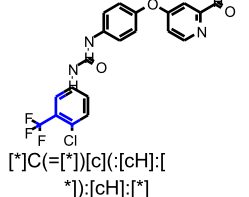
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

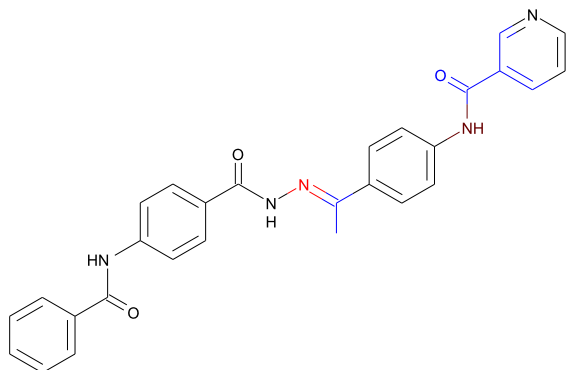
- All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-885550502	 <chem>[*]C(=[*])NC</chem>	0.115

FCFP_2	3	 <chem>[*]N[*]</chem>	0.0737
FCFP_2	332760439	 <chem>[*]O[c](:[cH]:[*]):[cH]:[*]</chem>	0.0611
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	 <chem>[*]:[c](:[*])Cl</chem>	-0.134
FCFP_2	1872154524	 <chem>[*]C(=O)[*]</chem>	-0.105
FCFP_2	203677720	 <chem>[*]C(=[*])[c](:[cH]:[*]):[cH]:[*]</chem>	-0.0829


 $C_{28}H_{23}N_5O_3$

Molecular Weight: 477.51391

ALogP: 3.082

Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: 0.148

Unit: g/kg_body_weight

Mahalanobis Distance: 12.3

Mahalanobis Distance p-value: 3.08e-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 prediction. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK
Structure			
Actual Endpoint (-log C)	6.28396	2.82494	2.54455
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702
Distance	0.921	1.081	1.282
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

Model Applicability

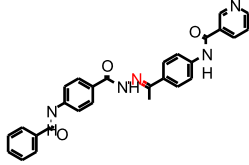
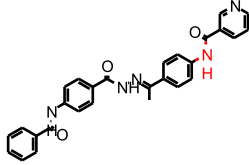
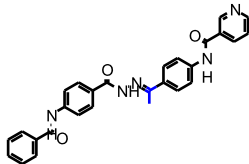
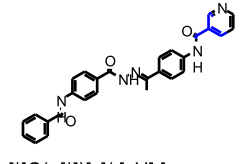
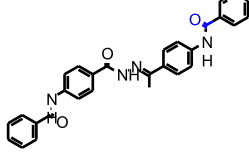
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 477.51. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_AromaticRings out of range. Value: 4. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
3. OPS_PC6 out of range. Value: -2.6475. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
4. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
5. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]
6. Unknown FCFP_2 feature: -885520711: [*]C(=[*])NN=[*]

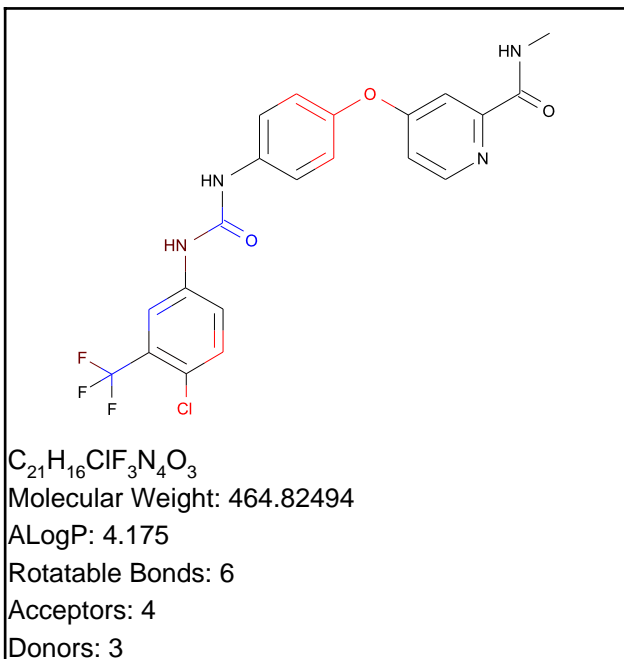
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	1	 [*]N=[*]	0.511
FCFP_2	3	 [*]N[*]	0.104
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

Sorafenib



Model Prediction

Prediction: 0.000918

Unit: g/kg_body_weight

Mahalanobis Distance: 12.2

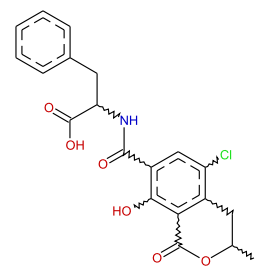
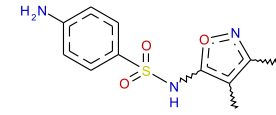
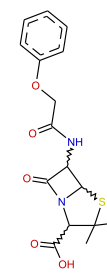
Mahalanobis Distance p-value: 4.69e-009

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

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage

Structural Similar Compounds

Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK
Structure			
Actual Endpoint (-log C)	6.28396	2.82494	2.54455
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702
Distance	0.758	0.997	1.159
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

Model Applicability

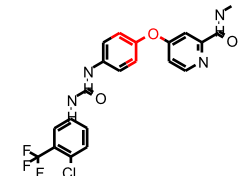
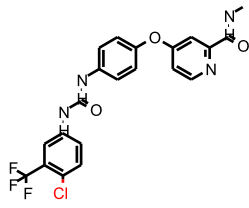
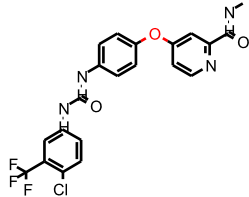
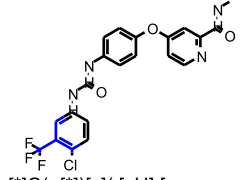
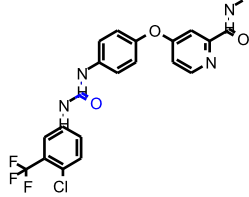
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

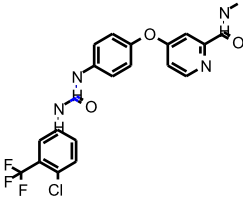
1. Molecular_Weight out of range. Value: 464.82. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
3. OPS_PC5 out of range. Value: -3.5737. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
4. OPS_PC7 out of range. Value: -3.8342. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
5. Unknown FCFP_2 feature: 1499521844: [*]NC(=O)N[*]
6. Unknown FCFP_2 feature: -1029533685: [*]:[c](:[*])C(F)(F)F
7. Unknown FCFP_2 feature: 136686699: [*]NC

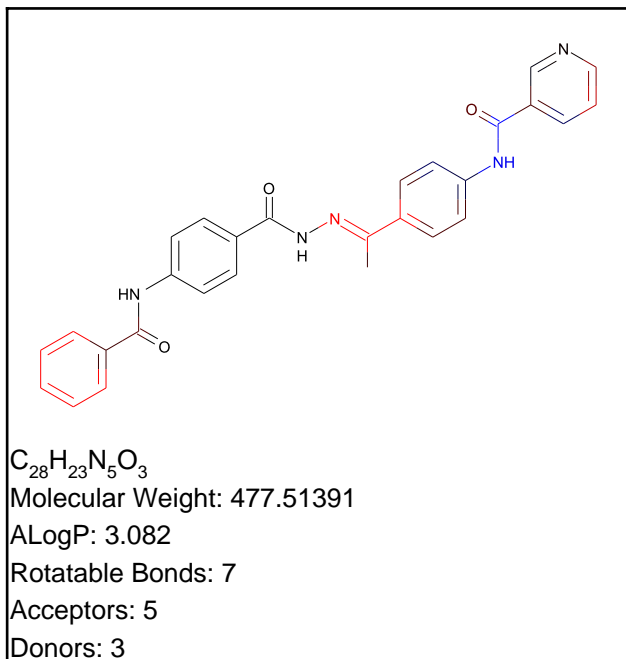
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 <chem>[*]O[c](:[cH]:[*]);[cH]:[*]</chem>	0.672
FCFP_2	32	 <chem>[*]Cl</chem>	0.526
FCFP_2	1	 <chem>[*]N=[*]</chem>	0.511
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 <chem>[*]C(=[*])[c](:[cH]:[*]);[cH]:[*]</chem>	-0.406
FCFP_2	1872154524	 <chem>[*]C(=O)[*]</chem>	-0.307

FCFP_2	0	 <chem>[*]C(=[*])[*]</chem>	-0.29
--------	---	---	-------



Model Prediction

Prediction: 2.81

Unit: g/kg_body_weight

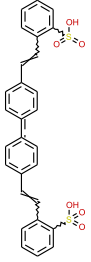
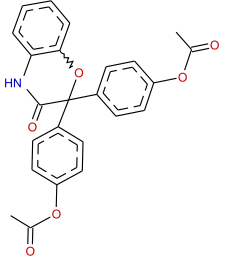
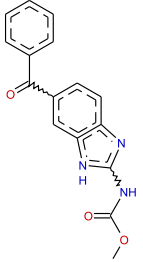
Mahalanobis Distance: 22.1

Mahalanobis Distance p-value: 1.54e-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 prediction. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	BENZENESULFONIC ACID; 2,2'-(4;4'-BIPHENYLYLENE)DI-; DISODIUM SALT (Na STRIPPED)	bis-OXATIN ACETATE	CARBAMIC ACID; N-(5-BENZOYL-BENZIMIDAZOL-2-YL)-; METHYL ESTER
Structure			
Actual Endpoint (-log C)	1.968	1.717	2.617
Predicted Endpoint (-log C)	1.72109	2.40947	2.2368
Distance	0.834	0.851	0.861
Reference	MVCRB3 2;193;73	NIIRDN 6;609;82	IYKEDH 19;735;88

Model Applicability

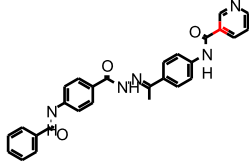
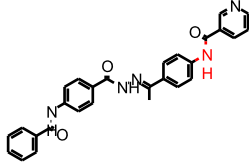
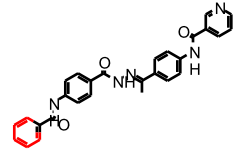
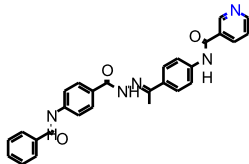
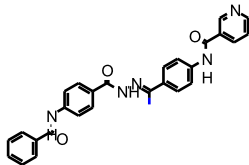
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

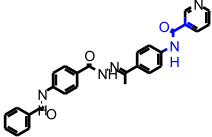
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]N=C(/C)\[c](:[*]):[*]
3. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
4. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
5. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]
6. Unknown FCFP_6 feature: 581019816: [*]NN=C(*)[*]
7. Unknown FCFP_6 feature: -885520711: [*]C(=[*])NN=[*]

Feature Contribution

Top features for positive contribution

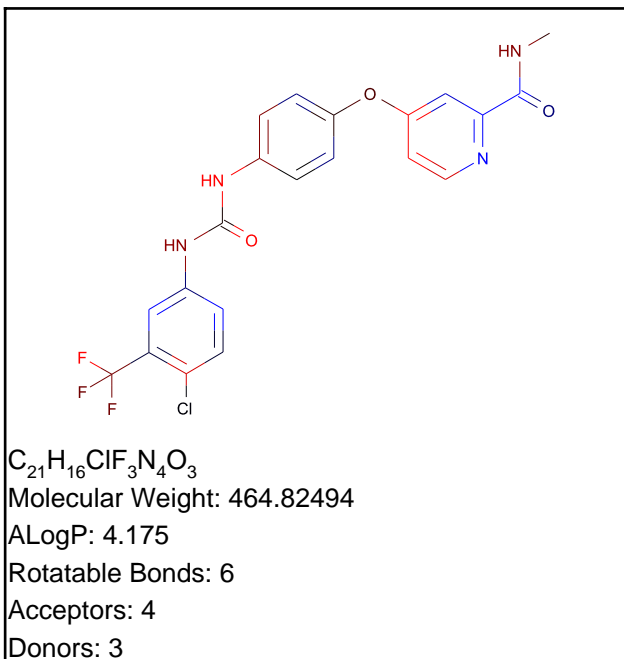
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	-0.239
ECFP_6	734603939	 [*]C	-0.201

FCFP_6	1549103449	 <chem>[*]NC(=O)c1ccc(cc1)Nc2ccc(cc2)Nc3ccncc3</chem>	0.117
--------	------------	---	-------

Sorafenib

TOPKAT_Rat_Oral_LD50



Model Prediction

Prediction: 0.823

Unit: g/kg_body_weight

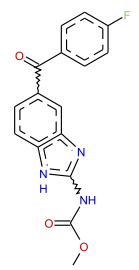
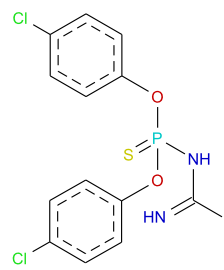
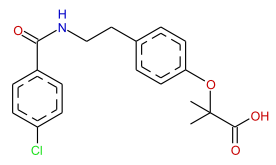
Mahalanobis Distance: 21

Mahalanobis Distance p-value: 1.93e-012

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

Structural Similar Compounds

Name	FLUBENDAZOLE	PHOSPHORAMIDOTHIOIC ACID; ACETIMIDOYL-; O; O-bis-(p-CHLOROPHENYL)ESTER	BEZAFIBRATE
Structure			
Actual Endpoint (-log C)	2.088	5.006	1.946
Predicted Endpoint (-log C)	2.69288	3.23989	2.54395
Distance	0.697	0.703	0.721
Reference	YRTMA6 9;11;78	FMCHA2 -;C149;89	ARZNAD 30;2023;80

Model Applicability

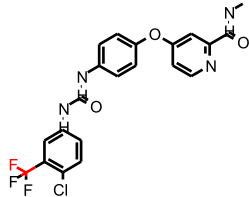
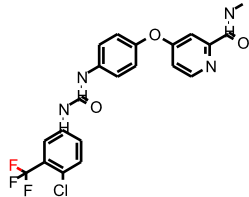
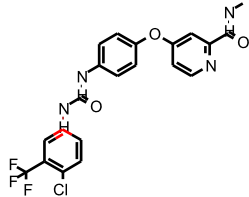
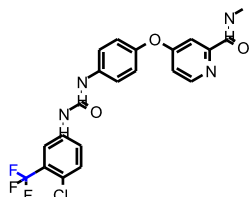
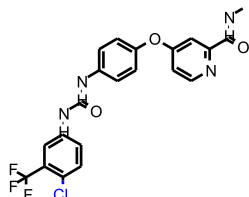
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

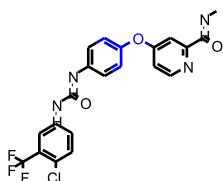
1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
3. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl
4. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]
5. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
6. Unknown FCFP_6 feature: 136686699: [*]NC

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	71953198	 [*]C*F	0.392
ECFP_6	-1046436026	 [*]F	0.349
ECFP_6	642810091	 [*]c(:[*]):[*]	0.281
Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	226796801	 [*]C*F	-0.32
ECFP_6	-817402818	 [*]Cl	-0.263

ECFP_6	-176455838	 <chem>O=C1N=CN=C1c2ccc(Oc3ccc(N4C=NC=C4)cc3)cc2</chem> [*]O[c](:[cH]:[*]):[c H]:[*]	-0.257
--------	------------	--	--------