Design, synthesis, docking, MD simulations, and anti-proliferative evaluation of thieno[2,3-d]pyrimidine derivatives as new EGFR inhibitors

Eman A. Sobh^{a*}, Mohammed A. Dahab^{b*}, Eslam B. Elkaeed^c, Aisha A. Alsfouk^d, Ibrahim M. Ibrahim ^e, Ahmed M. Metwaly ^{f, g}, Ibrahim H. Eissa^{b*}.

- ^a Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Menoufia University, Menoufia, Shibin-El-Kom, Gamal Abd Al-Nasir Street, Egypt.
- ^b Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo, 11884, Egypt.
- ^c Department of Pharmaceutical Sciences, College of Pharmacy, AlMaarefa University, Riyadh 13713, Saudi Arabia.
- ^e Department of Pharmaceutical Sciences, College of Pharmacy, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia.
- ^f Biophysics Department, Faculty of Science, Cairo University. Cairo 12613, Egypt.
- ^f Pharmacognosy and Medicinal Plants Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo, Egypt
- ^g Biopharmaceutical Products Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications (SRTA-City), Alexandria 21934, Egypt

* Corresponding authors:

Eman A. Sobh

Pharmaceutical Chemistry Department, Faculty of Pharmacy, Menoufia University, Menoufia, Egypt.

E-mail: Eman_abd-elfattah@phrm.menofia.edu.eg

Mohammed A. Dahab

Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo, 11884, Egypt

E-mail: mohammeddahab@azhar.edu.eg

Ibrahim H. Eissa

Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo, 11884, Egypt

Email: <u>Ibrahimeissa@azhar.edu.eg</u>

	C	Content	
1. Chemistry			
2. Biology			
3. In silico stud	lies		
4. Spectral dat	a		

1. Chemistry

All commercially available materials were purchased from commercial sources and used without further purification. All reactions were performed in a temperature-controlled oil/wax bath. Reactions were monitored by analytical thin layer chromatography (TLC), using aluminum-backed plates, cut to size. TLC visualization was achieved by UV. ¹H NMR and ¹³C NMR spectra were recorded on Bruker AVANCE 400 MHz spectrometers at ambient temperature. Spectral data is reported in ppm, coupling constants (*J*) in Hz. IR spectra were recorded as KBr pellets on a PerkinElmer Spectrum one FT-IR spectrometer. Mass spectroscopy (MS) was performed at NAWAH scientific Mass Spectroscopy Centre on Advion compact mass spectrometer (CMS) NY | USA for APCI. All melting points were taken on a Gallenkamp melting point apparatus and are uncorrected. Nomenclature was determined using ChemBioDraw Ultra 14.0.

Ethyl 2-amino-6-methyl-4,5,6,7-tetrahydro[1]benzothiophene-3-carboxylate (2)

A mixture of 4-methylcyclohexanone (1) (4.48 g, 40 mmol), ethyl cyanoacetate (4.52 g, 40 mmol), sulphur (1.28 g, 40 mmol) and morpholine (5 mL) in absolute ethanol (25 mL) was heated on water-bath at 40°C for 4 h. The reaction mixture was allowed to cool to room temperature, the formed solid was filtered, dried and crystallized from ethanol to give 2.

$$\sim$$
 COOEt \sim NH₂

6-Methyl-5,6,7,8-tetrahydro[1]benzothieno[2,3-d]pyrimidin-4(3H)-one (3)

A mixture of ethyl 2-amino-6-methyl-4,5,6,7-tetrahydro[1]benzothiophene-3-carboxylate (2) (0.24 g, 1 mmol) and formamide (99.5%) (15 mL) was heated under reflux for 5 h. The solvent was then evaporated under reduced pressure and the residue was crystallized isopropanol.

$$H_3C$$
 NH
 NH

2. Biological testing

2.1.In vitro antiproliferative activity

The *in vitro* antiproliferative activities of all the synthesized compounds against a panel of four human tumor cell lines namely, A549 and MCF-7cell were evaluated quantitatively as described in the literature, using MTT assay protocol. two commercially available drugs (erlotinib) were used in this test as positive controls. The anti-proliferative activity was assessed quantitatively as follows.

Human cancer cell lines were dropped in 96-well plates at a density of $3\text{-}8 \times 10^3$ cells/well. Next, the wells were incubated for 12 h in a 5% CO₂ incubator at 37 °C. Then, for each well, the growth medium was exchanged with 0.1 ml of fresh medium containing graded concentrations of the test compounds to be or equal DMSO and incubated for two days. Then $10~\mu l$ MTT solution (5 $\mu g/m l$) was added to each well, and the cells were incubated for additional 4 h. The crystals of MTT-formazan were dissolved in $100~\mu l$ of DMSO; the absorbance of each well was measured at 490 nm using an automatic ELISA reader system (TECAN, CHE). The IC₅₀ values were calculated using the nonlinear regression fitting models (Graph Pad, Prism Version 5). The data represented the mean of three independent experiments in triplicate and were expressed as means \pm SD. The IC₅₀ value was defined as the concentration at which 50% of the cells could survive.

2.2.Safety assay

The normal cell lines, W138, were utilized. MTT assay was applied as described before.

2.3.In vitro EGFR inhibition

The tested compounds were examined for their inhibitory activities against both EGFR^{WT} and EGFR^{T790M}. Homogeneous time resolved fluorescence (HTRF) assay was applied in this test. EGFR^{WT}, EGFR^{T790M} (Sigma). firstly, the EGFR^{WT} and/or EGFR^{T790M} and their substrates were incubated with the tested compounds in enzymatic buffer for 5 min. ATP (1.65 μM) was added into the reaction mixture to allow starting the enzymatic reaction. The assay was conducted for 30 min at room temperature. The reaction was stopped by addition of detection reagents which contain EDTA. The detection step continued for 1 h, and then the IC₅₀ values were determined by GraphPad Prism 5.0. Three independent experiments were performed for each concentration.

2.4.Cell cycle analysis

A549 cells were exposed to compound 5b at a concentration of 17.79 μ M μ M for 72h. Then, the tested cells were collected by trypsinization and washed in PBS. Ice-cold absolute ethanol was used for fixation of the collected cells. The cells were stained with Cycle TESTTM PLUS DNA Reagent Kit (BD Biosciences, San Jose, CA) according to the manufacturer's instructions. Cell-cycle distribution was evaluated using a flow cytometer.

2.5.Apoptosis analysis

To detect the apoptosis induced by compound **5b**, A549 cells were seeded and incubated overnight and then treated with compound **5b** at concentrations of 17.79 µM for 72 h. DMSO was chosen as the negative control. After that, the cells were collected and washed with PBS two successive times. The cells were exposed to centrifugation. Apoptosis detection kit (BD Biosciences, San Jose, CA) was used in this test. According to the manufacturer's protocol the cells were stained by Annexin V-FITC and propidium iodide (PI) in the binding buffer for 20min at room temperature in the dark. Using a flow cytometer, Annexin V-FITC and PI binding were analyzed. flowjo software was used to analyze the frequencies in all quadrants.

2.6.Quantitative Real Time Reverse-Transcriptase PCR technique (determination of BAX and Bcl-2)

The quantity of BAX and Bcl-2 mRNA in control and compound 5b (at the IC₅₀ concentration)-treated A549 cells was assessed by qRT-PCR (reference). Total RNA from vehicle-treated control (0.01% DMSO) and 10k-treated HepG2 cells were extracted as-per the manufacturer instructions (RNeasy mini kit, Qiagen, Germany). After RNA extraction, cDNA was prepared

using the Revert Aid First Strand cDNA Synthesis kit (Thermo Scientific, USA). Amplification of target cDNA for apoptosis markers and GAPDH [as a normalization (housekeeping) gene] was done using one-step RT-PCR SYBR® Green kit Master Mix (Bio-Rad Laboratories, USA) on Rotor-Gene Q real-time PCR thermal cycler instrument. cDNA (2 µl aliquots) was mixed with 1 µl of forward primer, 1 µl reverse primer, 10 µl master mixture, and the reaction volume was completed to 20 µl with nuclease-free water. All experiments were performed in triplicates.

The sequences for primers used in quantitative Real Time Reverse-Transcriptase PCR (qRT-PCR)

Gene	Primer sequence		
BAX	F: 5'- CCCGAGAGGTCTTTTTCCGAG -3'		
	R: 5'- CCAGCCCATGATGGTTCTGAT -3'		
Bcl-2	F: 5'- TTGTGGCCTTCTTTGAGTTCGGTG -3'		
	R: 5'- GGTGCCGGTTCAGGTACTCAGTCA -3'		
GABDH	F: 5'- GACCCCTTCAT GACCTCAAC -3'		
	R: 5'- CTTCTCCATGGTGGT GAAGA -3'		

3. In silico studies

3.1.Docking studies

The crystal structures of the target enzymes EGFR^{WT} (PDB ID: 4HJO, resolution 2.75 Å and EGFR^{T790M} (PDB ID: 3W2O, resolution 2.35 Å) were downloaded from Protein Data Bank (http://www.pdb.org). Molecular Operating Environment (MOE) was used for the docking analysis [91]. In these studies, the free energies and binding modes of the designed molecules against EGFR^{WT} and EGFR^{T790M} were determined. At first, the water molecules were removed from the crystal structures of EGFR^{WT} and EGFR^{T790M}, retaining only one chain in each enzyme. Erlotinib and TAK-285 (The co-crystallized ligands) were utilized as references in the docking processes against both EGFR^{WT} and EGFR^{T790M}, respectively. After that, in order to prepare the target molecules for binding with the designed compounds, the target proteins were subjected to protonation step. Then, the hydrogen atoms were hidden to make the areas

of interaction clearer. Next, the energy of all systems was minimized followed by identification the binding pockets of the target proteins.

The structures of the designed compounds and the co-crystallized ligands, erlotinib and TAK-285 were drawn using ChemBioDraw Ultra 14.0 and saved as SDF format. Then, the saved files were opened using MOE and 3D structures were protonated. Next, the energy of the molecules was minimized. Validation process was performed for each target by running the docking process for only the co-crystallized ligand. Low RMSD values between docked and crystal conformations indicate valid performance. The docking procedures were carried out utilizing a default protocol. In each case, 10 docked structures were generated using genetic algorithm searches. The output from of MOE was further analyzed and visualized using Discovery Studio 4.0 software.

3.2.M D simulations

Molecular dynamics simulation of the protein-ligand complexes was performed using GROMACS 2021 and Linux 5.4 package. The GROMOS96 54a7 forcefield was selected as the force field for proteins and the ligand topologies were generated from the PRODRG server. All the complexes were solvated using simple point charge (SPC) water molecules in a rectangular box. To make the simulation system electrically neutral, required number of Na⁺ and Cl⁻ions were added while 0.15 mol/L salt concentrations were set in all the systems. Using the steepest descent method, all the solvated systems were subjected to energy minimization for 5000 steps. Afterwards, NVT (constant number of particles, volume, and temperature) series, NPT (constant number of particles, pressure, and temperature) series, and the production run were conducted in the MD simulation. The NVT and the NPT series were conducted at a 300 K temperature and 1 atm pressure for the duration of 300 ps. V-rescale thermostat and Parrinello-Rahman barostat were selected of the performed simulation. Finally, the production run was performed at 300 K for a duration of 100 ns (nanoseconds). Thereafter, a comparative analysis was performed measuring root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA) and hydrogen bonds to analyze their stability. The Xmgrace program was used to represent the analyses in the form of plots.

3.3.MM-GBSA

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

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

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

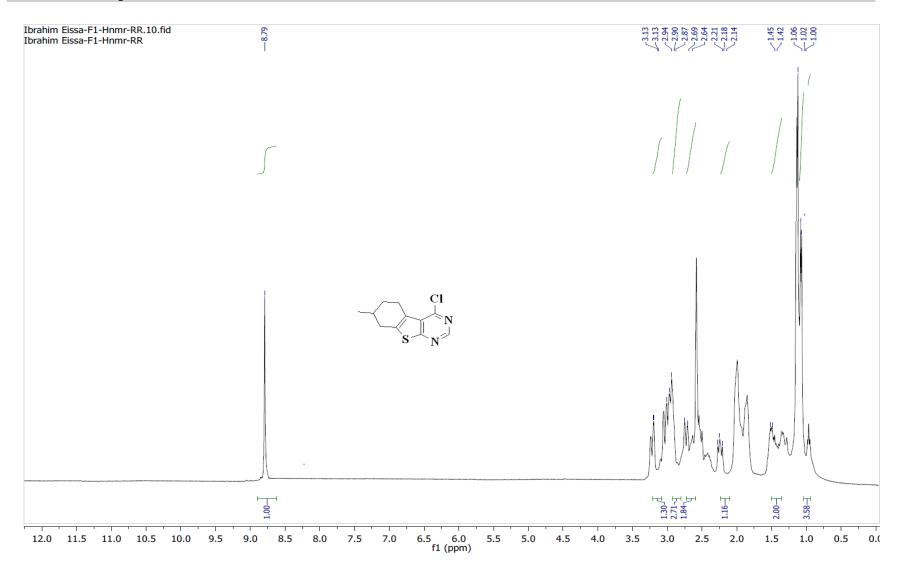
3.4.ADMET studies

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

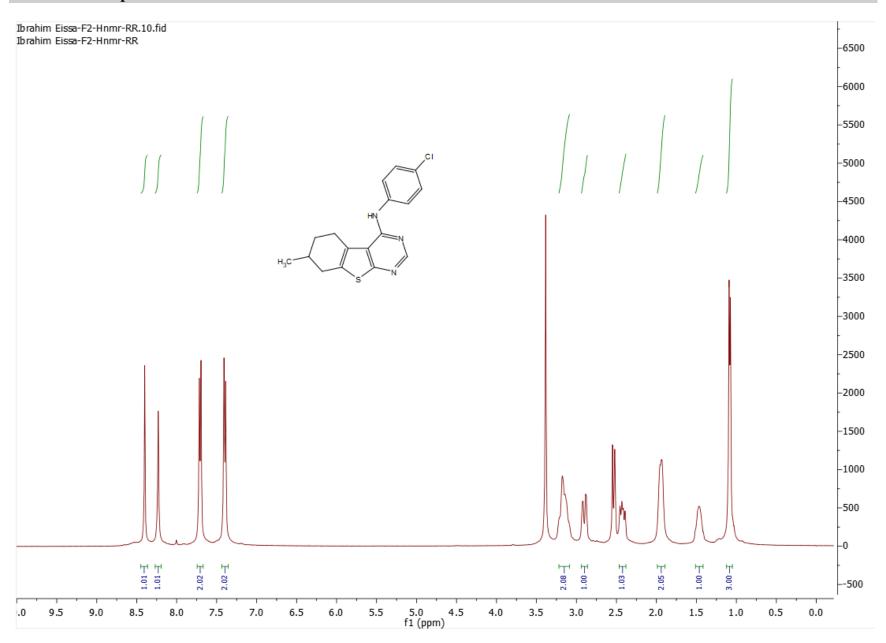
3.5. Toxicity studies

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

H¹ NMR of compound 4

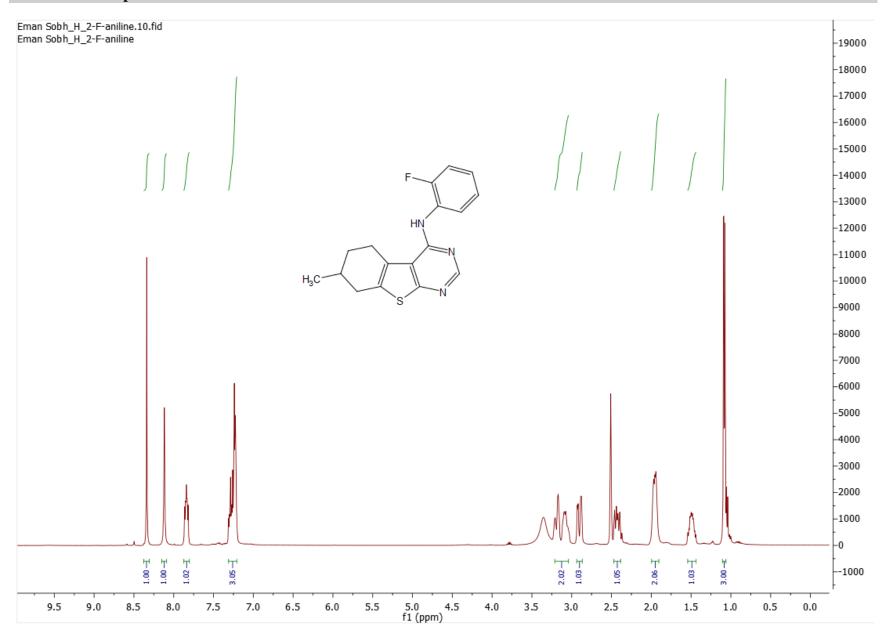


H¹ NMR of compound 5a



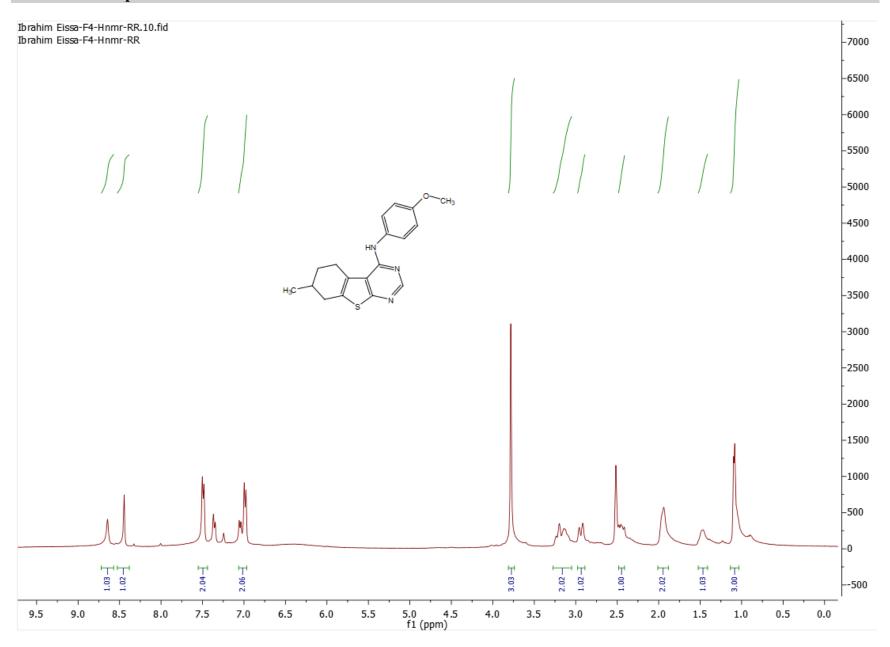
Page **9** of **42**

H1 NMR of compound 5b

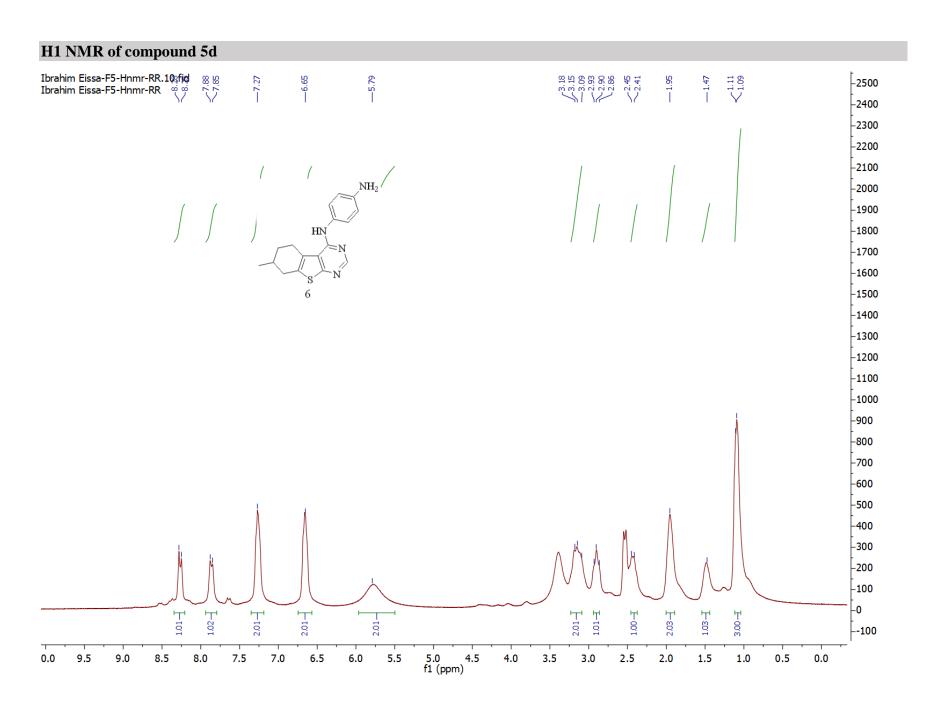


Page **10** of **42**

H1 NMR of compound 5c

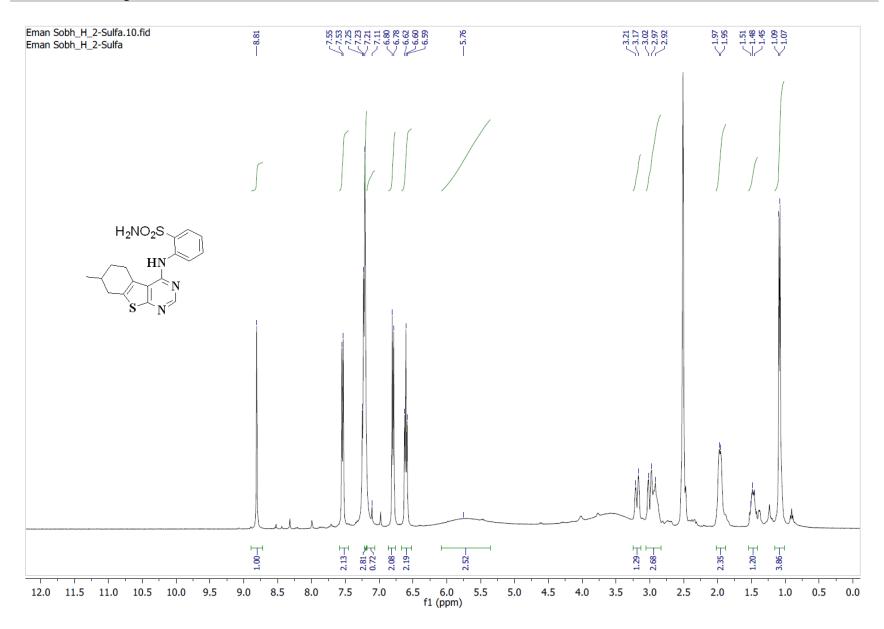


Page **11** of **42**



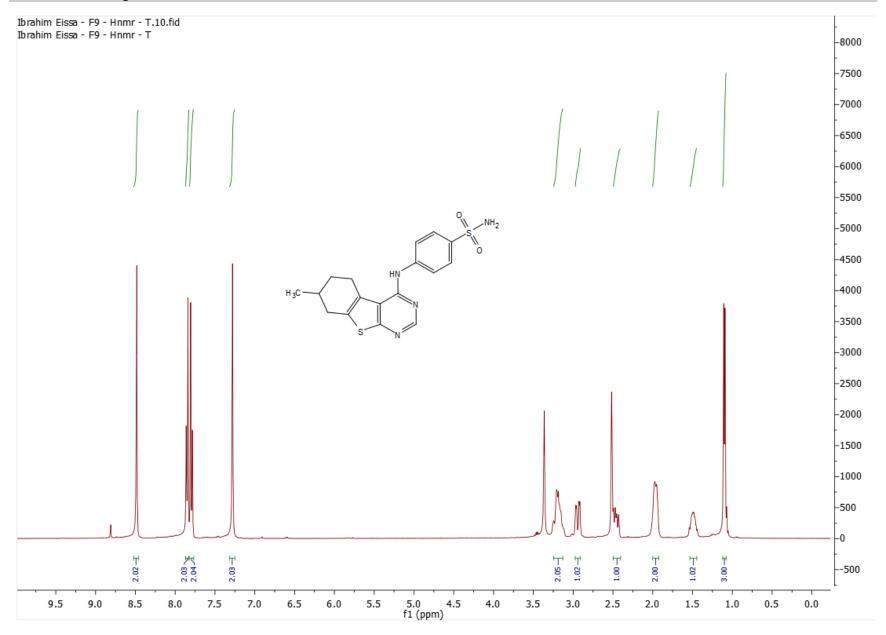
Page **12** of **42**

H1 NMR of compound 5e



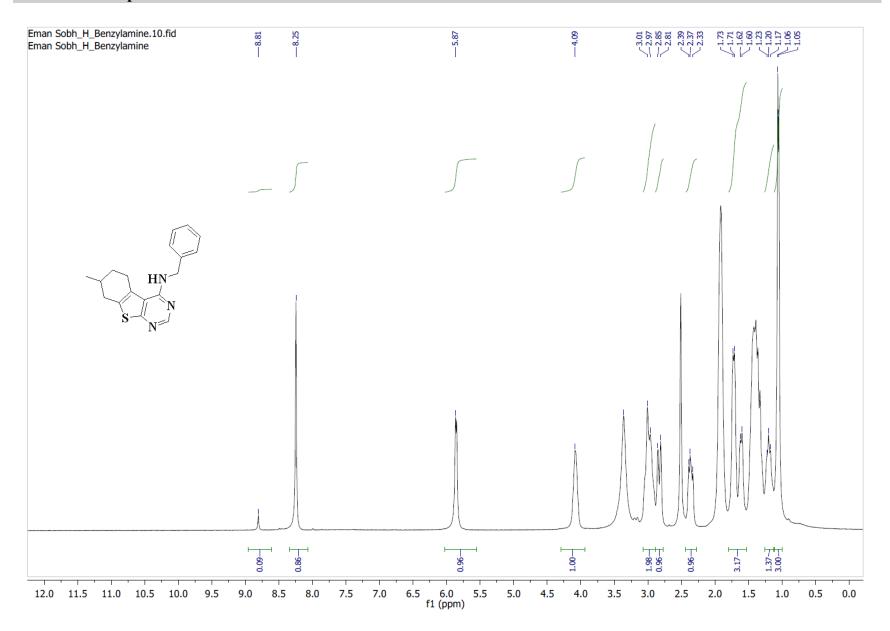
Page **13** of **42**

H1 NMR of compound 5f



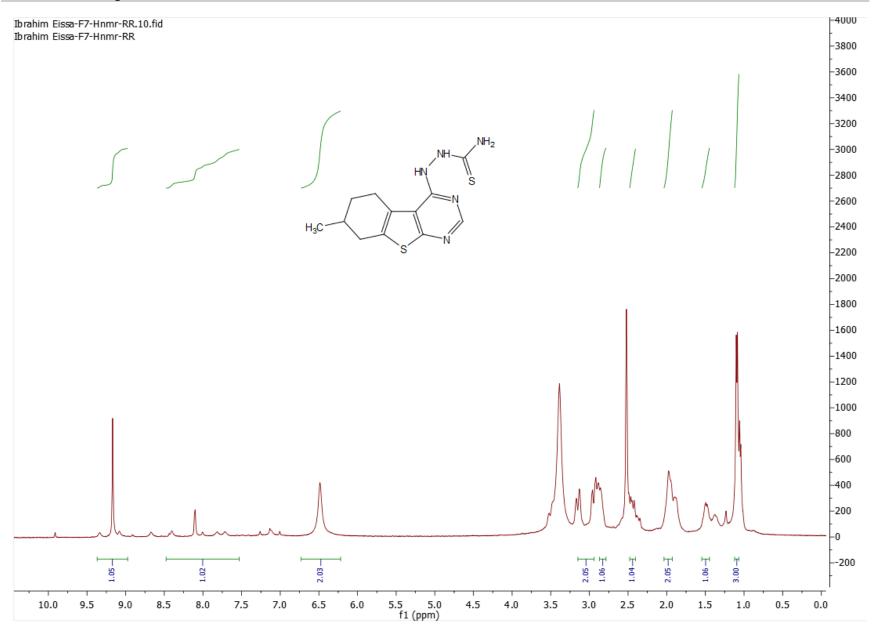
Page **14** of **42**

H1 NMR of compound 6



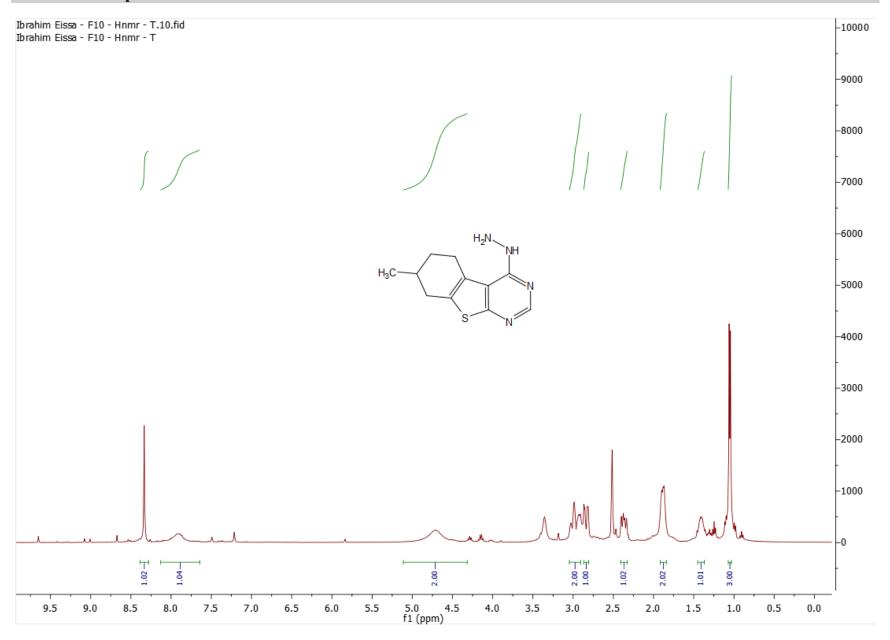
Page **15** of **42**

H1 NMR of compound 7a



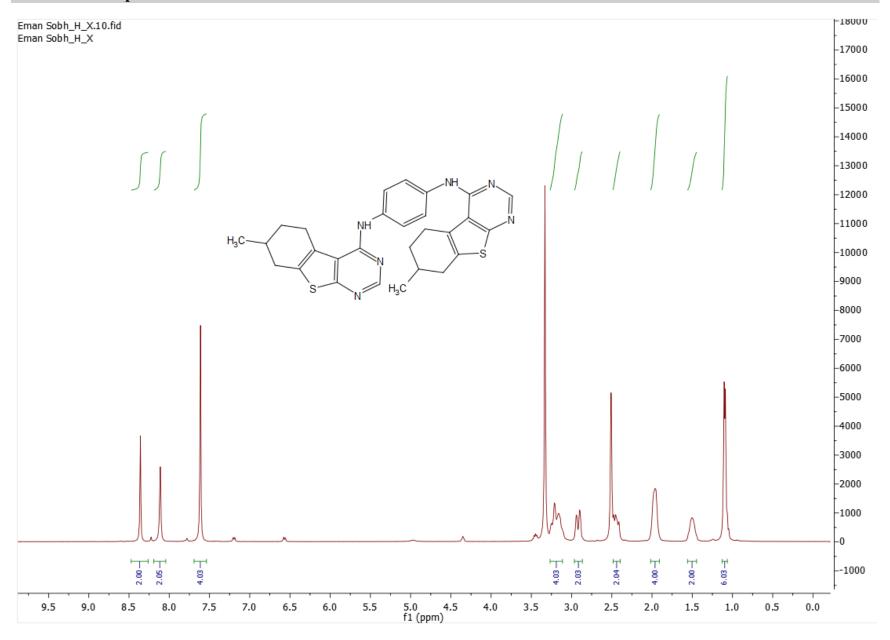
Page **16** of **42**

H1 NMR of compound 7b



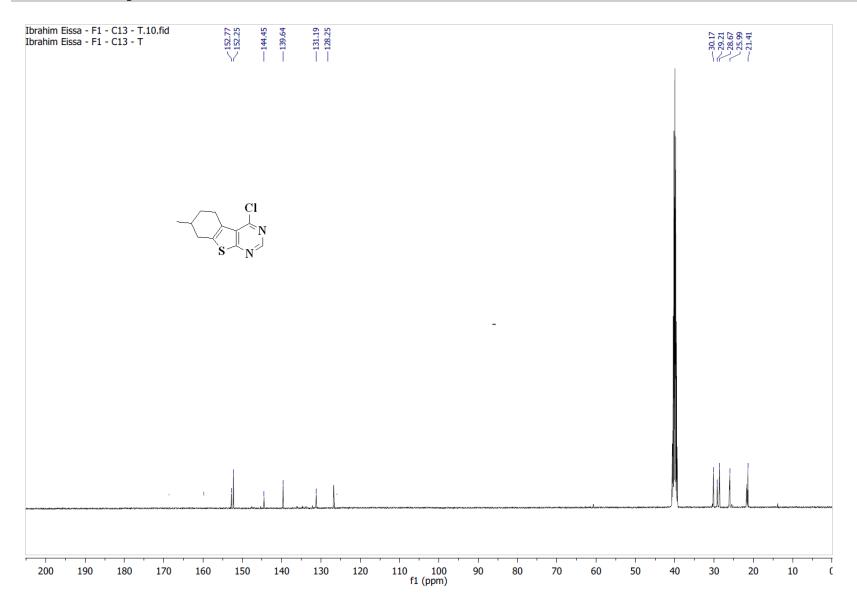
Page **17** of **42**

H1 NMR of compound 8

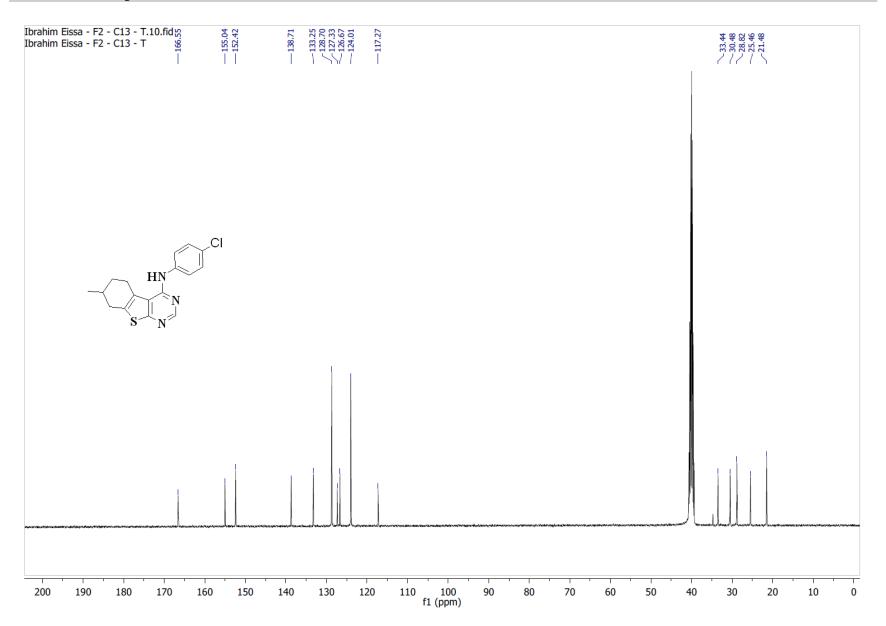


Page **18** of **42**

¹³C NMR of compound 4

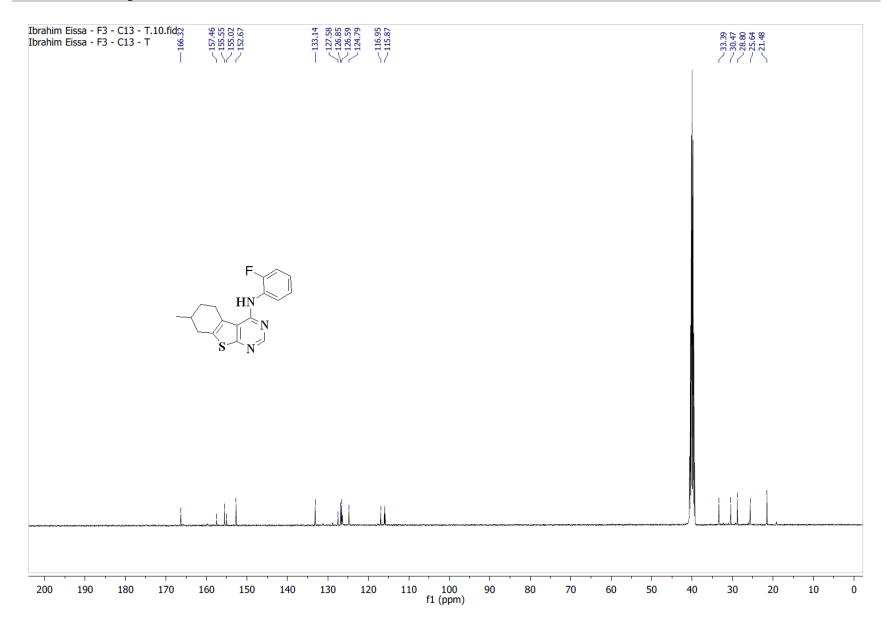


¹³C NMR of compound 5a



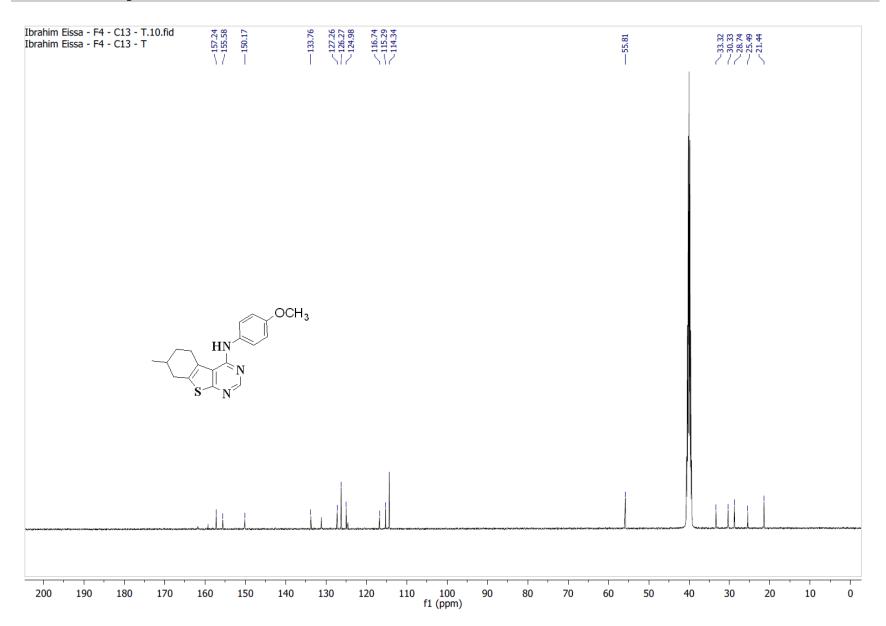
Page **20** of **42**

¹³C NMR of compound 5b

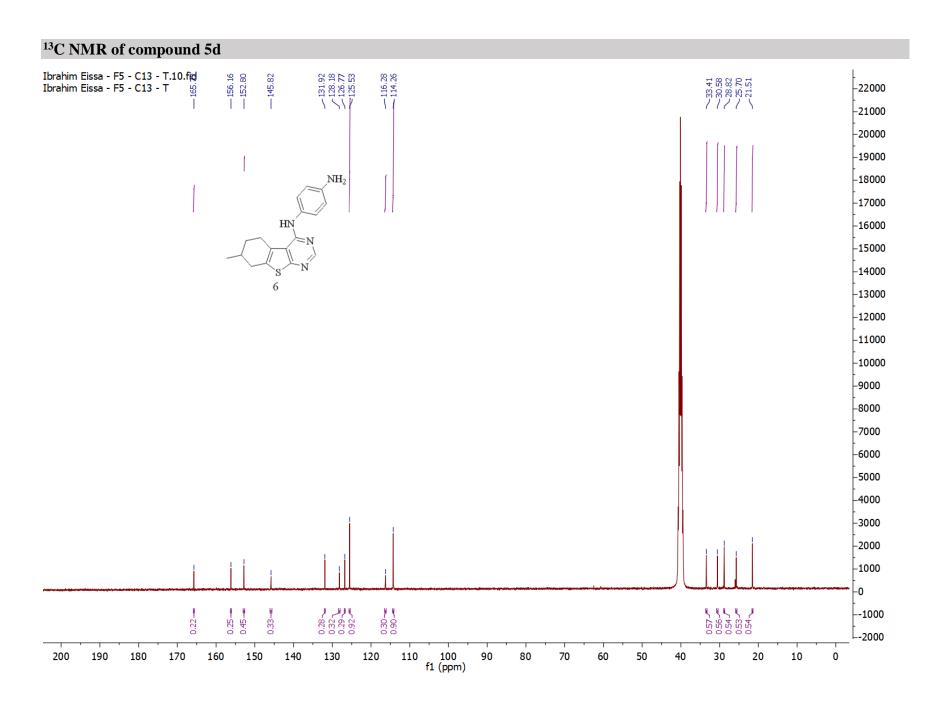


Page **21** of **42**

¹³C NMR of compound 5c

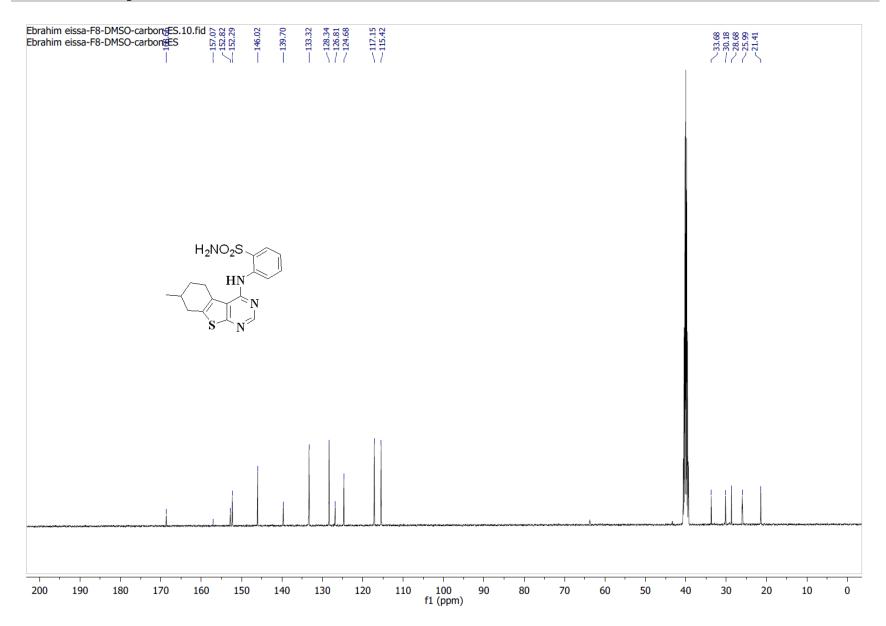


Page **22** of **42**



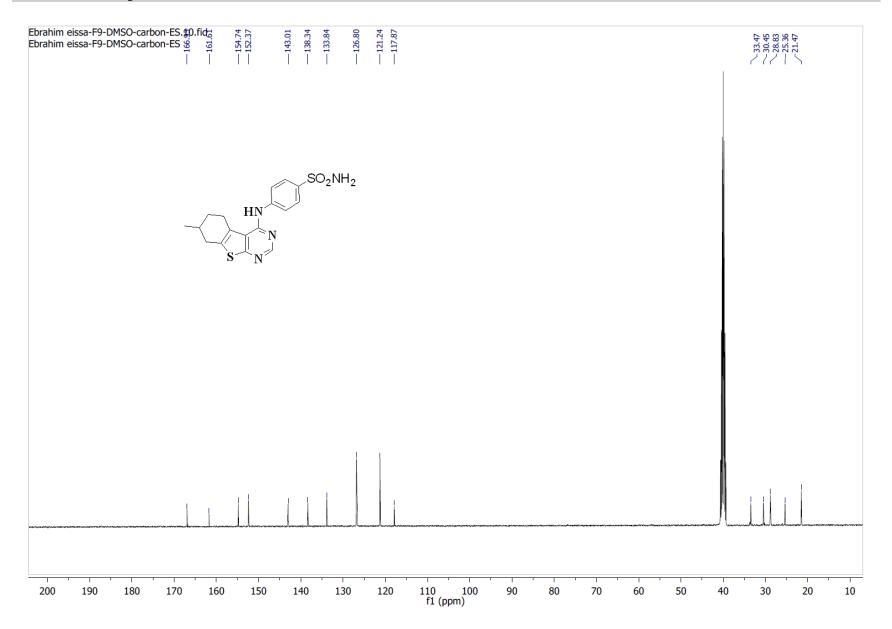
Page **23** of **42**

¹³C NMR of compound 5e



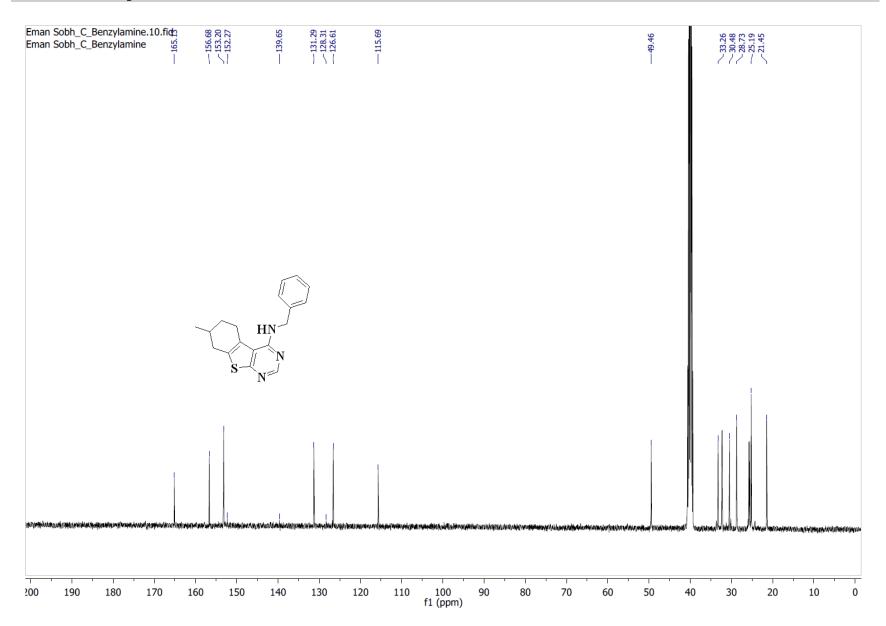
Page **24** of **42**

¹³C NMR of compound 5f



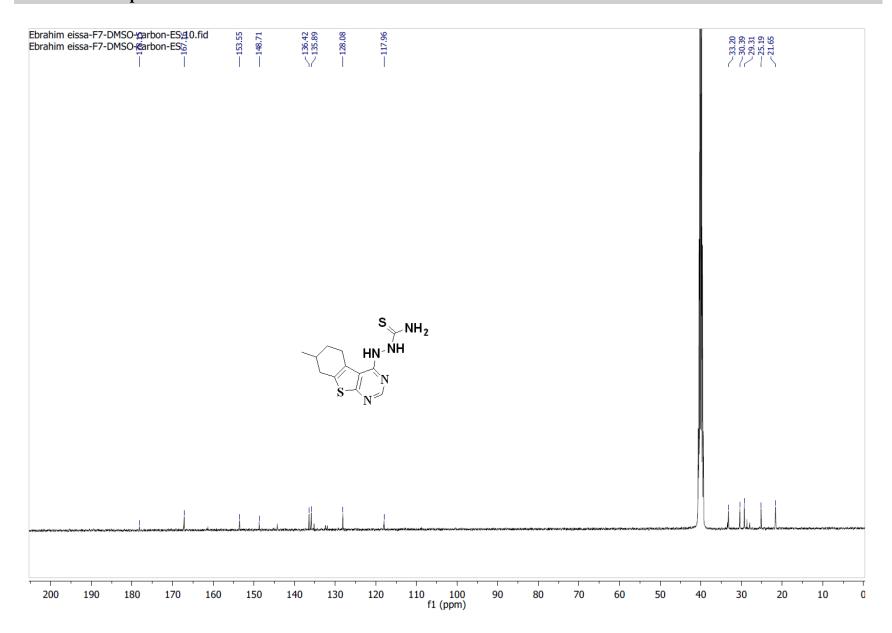
Page **25** of **42**

¹³C NMR of compound 6



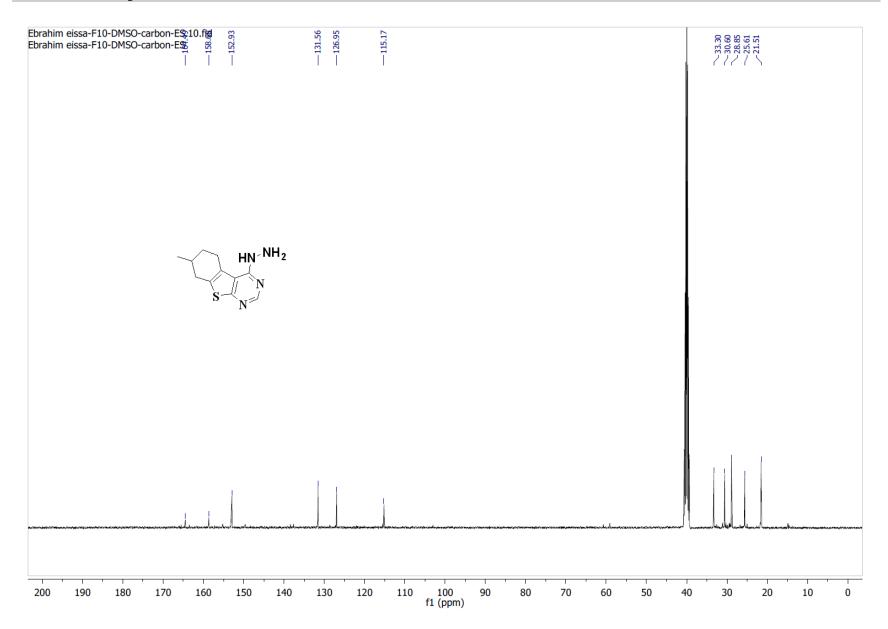
Page **26** of **42**

¹³C NMR of compound 7a

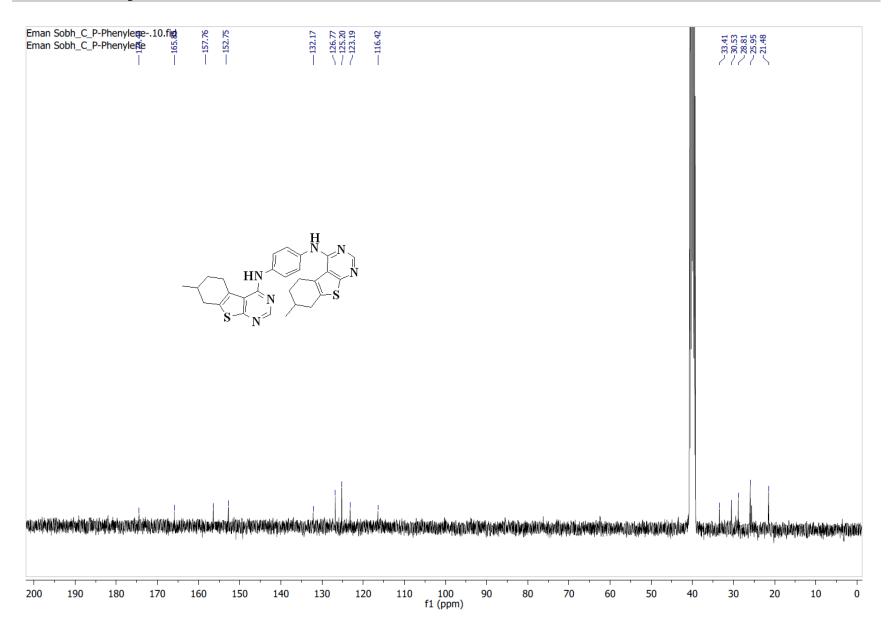


Page **27** of **42**

¹³C NMR of compound 7b

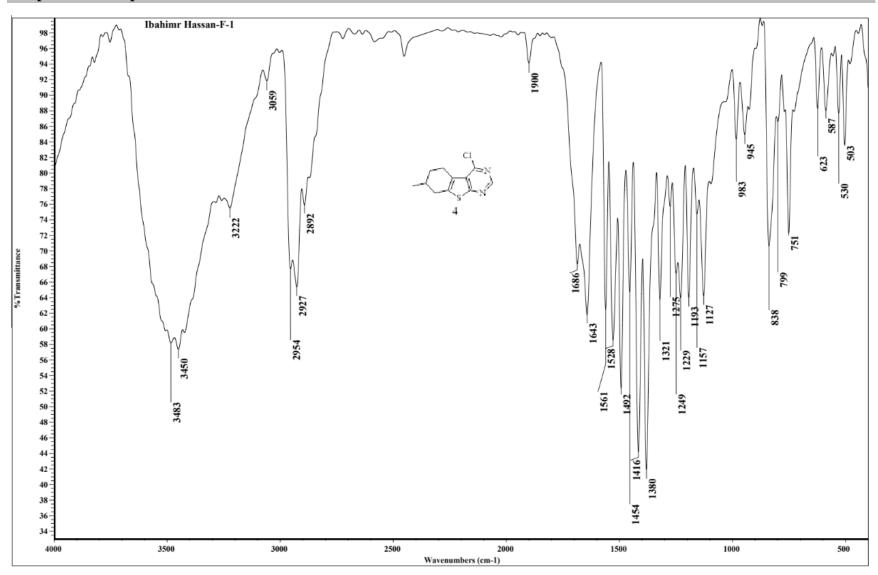


¹³C NMR of compound 8

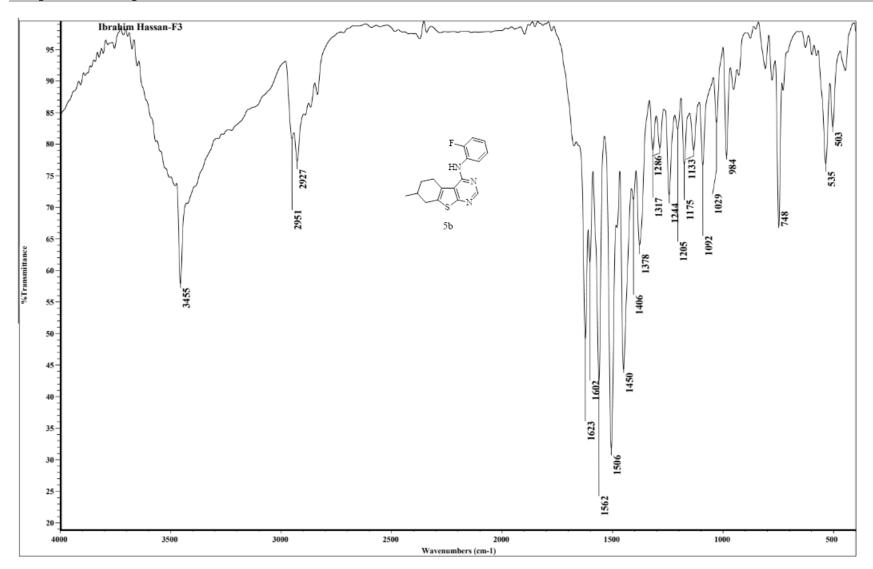


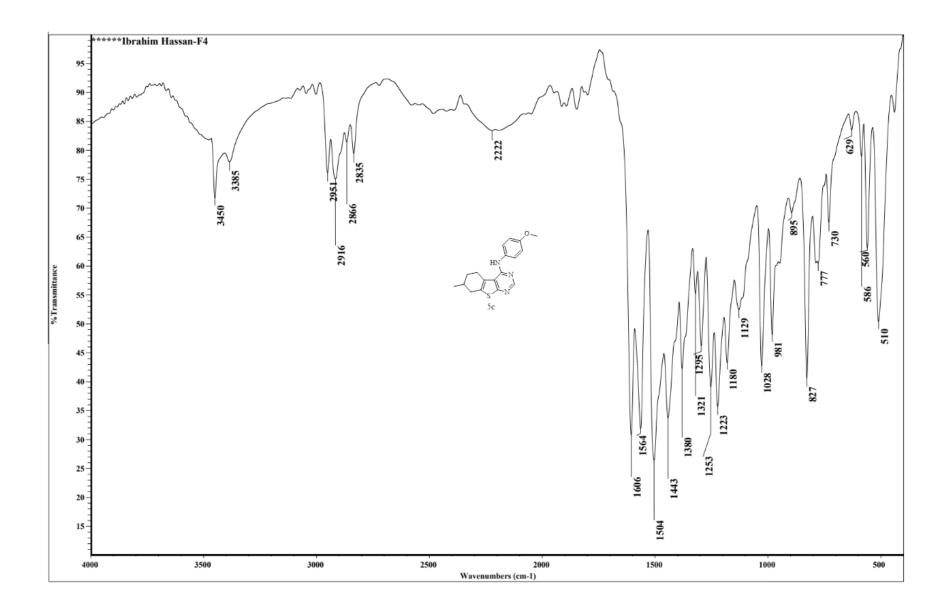
Page **29** of **42**

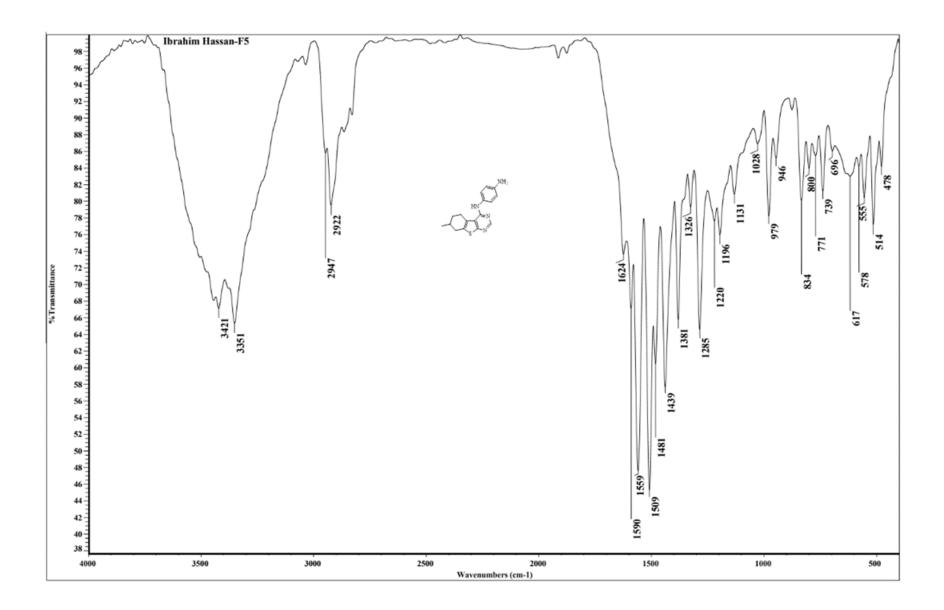
IR spectra of compound 4



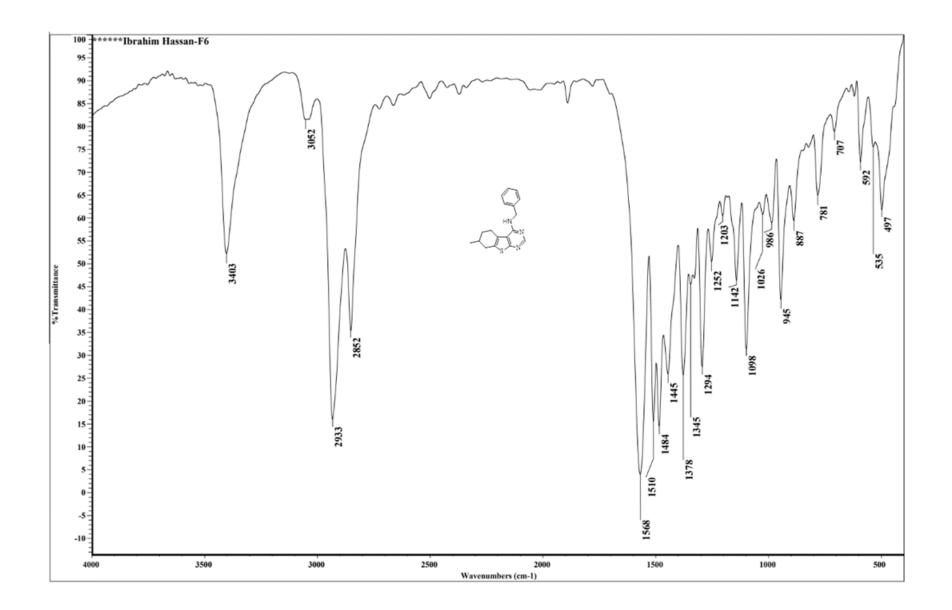
IR spectra of compound 5b

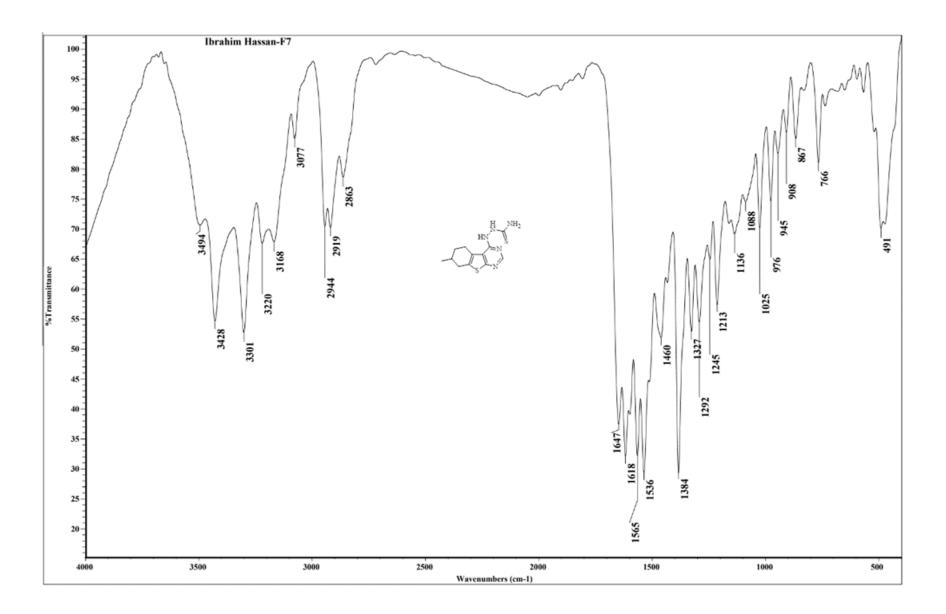




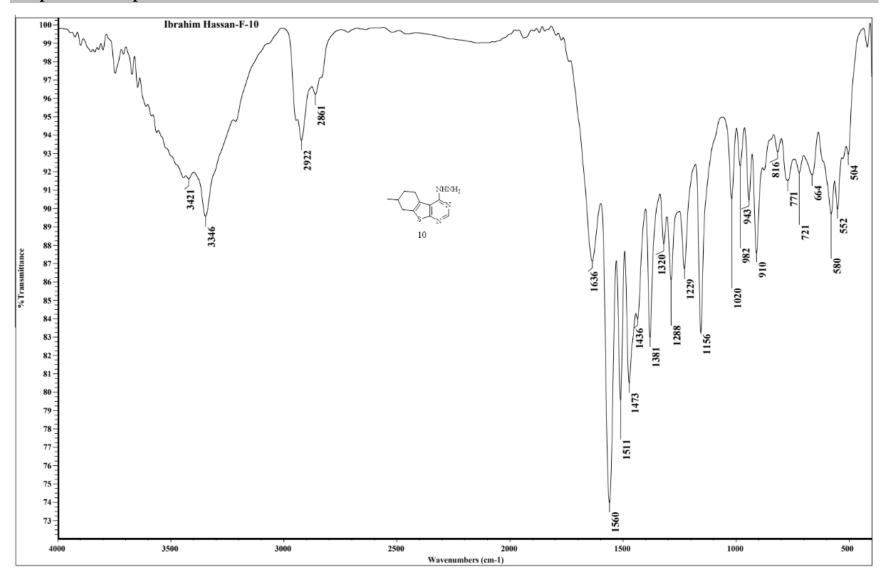


IR spectra of compound 6

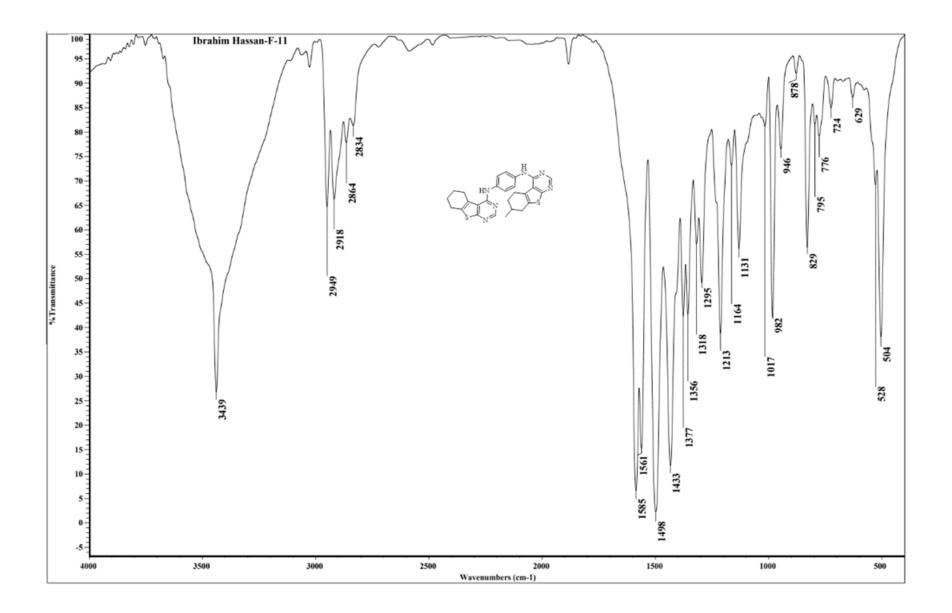




IR spectra of compound 7b

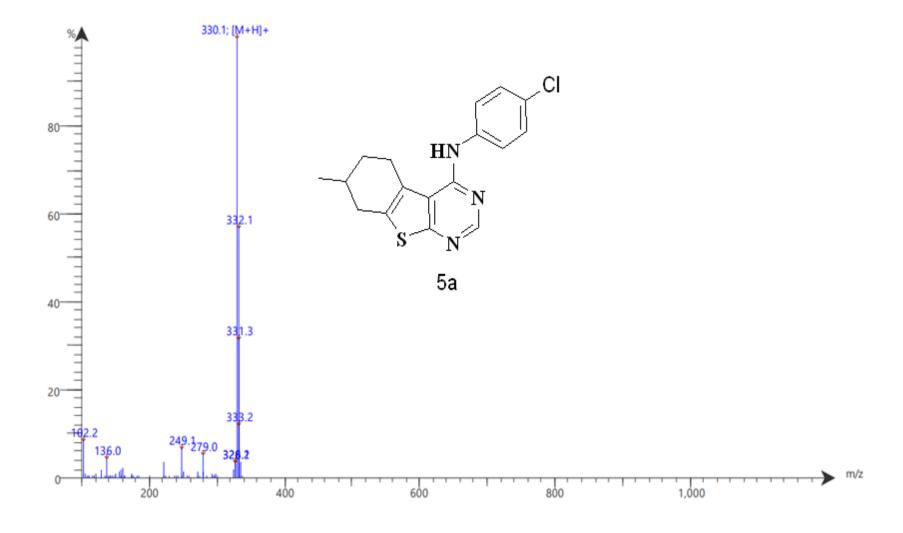


IR spectra of compound 8



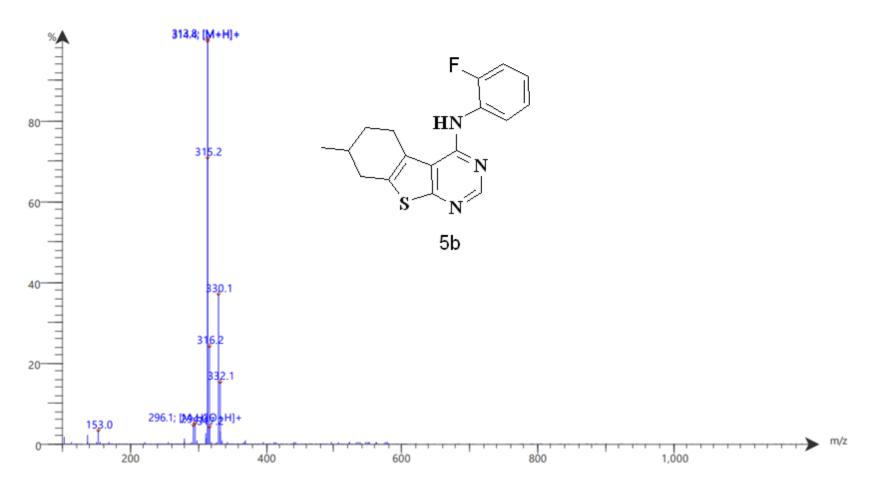
Mass spectra of compound 5a

Spectrum RT 0.30 - 0.48 (11 scans) 5a_Scan1_is1 2022.07.07 13:31:40 ; APCI + Max: 1.9E9



Mass spectra of compound 5b

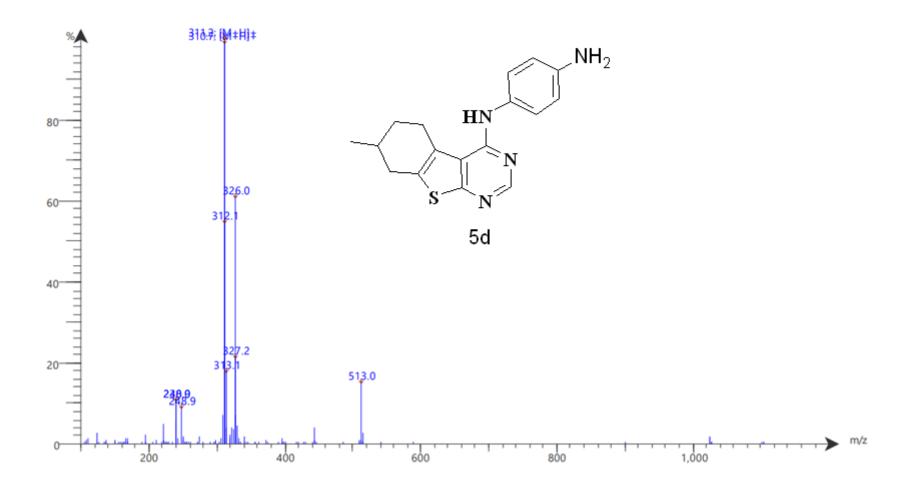
Spectrum RT 0.26 - 0.41 {9 scans} 5b_Scan1_is1 2022.07.07 13:34:32 ; Intensity APCI + Max: 2.2E9



Page **39** of **42**

Mass spectra of compound 5d

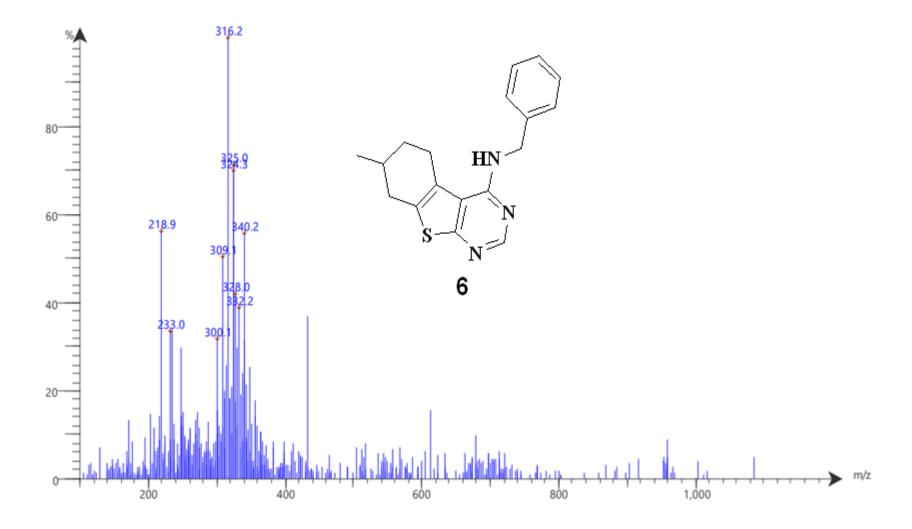
Spectrum RT 0.26 - 0.32 {4 scans} 6_Scan1_is1 2022.07.07 13:41:56 ; Intensity APCI + Max: 1.6E9



Mass spectra of compound 6

Spectrum RT 0.24 - 0.27 {3 scans} 7-_Scan2_is2 2022.07.07 13:48:54 ; APCI - Max: 7.5E6

Intensity



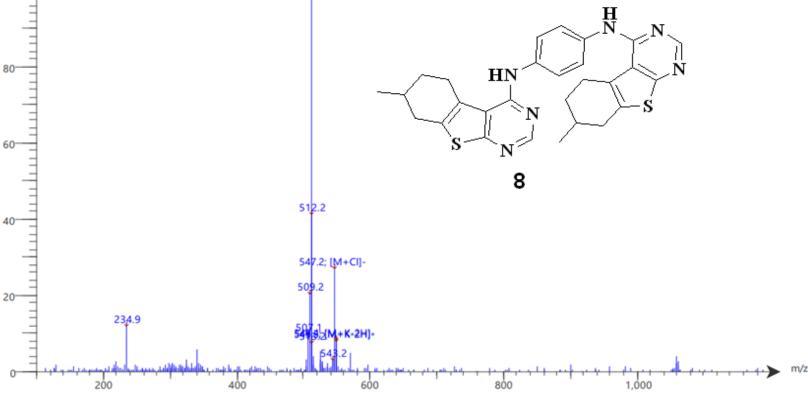
Page **41** of **42**

Mass spectra of compound 8

Intensity

Spectrum RT 0.54 - 0.61 (5 scans) 11_Scan2_is2 2022.07.07 13:57:10 ; APCI - Max: 2.2E7





Page **42** of **42**

AND Enantiomer S N N NH CI

 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.612 Enrichment: 1.1

Bayesian Score: -4.69

Mahalanobis Distance: 9.56

Mahalanobis Distance p-value: 0.558

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

	•		
Name	50264-69-2	89873-24-5	4947-27-7
Structure	HO NO CI	N N N N N N N N N N N N N N N N N N N	NH O
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen
Distance	0.555	0.557	0.559
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

Model Applicability

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

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

	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set		
SCFP_12	112346096	AND Enantiomer N	0.36	1035 out of 1263		

SCFP_12	1310748454	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.312	126 out of 161
SCFP_12	10	AND Enantiomer N N N N H CI [*]N[*]	0.306	1774 out of 2287
		ures for negative		1
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-55982834	AND Enantiomer N	-1.82	0 out of 9
SCFP_12	-1043339860	AND Enantiomer S N N N H CI [*]CC(C)C[*]	-1.09	76 out of 399
SCFP_12	1905487031	AND Enantomer N N N N N H CI [*][c]1:[cH]:[cH]:[c] (CI):[cH]:[cH]:1	-0.59	23 out of 74

AND Enantiomer N NH F

C₁₇H₁₆FN₃S

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.678 Enrichment: 1.21 Bayesian Score: -2.6

Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.0403

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural	Structural Sillillar Compounds						
Name	4947-27-7	50264-69-2	115664-51-2				
Structure		но	0				

	NH O	CI	O NH
Actual Endpoint	Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Mutagen	Non-Mutagen	Mutagen
Distance	0.539	0.546	0.546
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320
	·	·	·

Model Applicability

Structural Similar Compounds

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

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

	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set		
SCFP_12	112346096	AND Enantomer N N N N H F [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.36	1035 out of 1263		

SCFP_12	-1551011249	AND Enantioner S N H F [*]N[c]1:[cH]:[*]:[cH]:[cH]:[c]:1F	0.322	13 out of 16
SCFP_12	1310748454	AND Enamtiomer S N H F [*][c]1:[*]:[*]:[c](: [*]):s:1	0.312	126 out of 161
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-55982834	AND Examinmer N N N N H F [*]:[c]1:[*]C[C@@H](C)CC1	-1.82	0 out of 9
SCFP_12	-1043339860	AND Enantiomer S N NH F [*]CC(C)C[*]	-1.09	76 out of 399
SCFP_12	1328855840	AND Enantiomer S N N H F [*]:[cH]:[c]([N[c](:n: [*]):[c](:[*]):[*]): [cH]:[*]	-0.496	1 out of 4

N NH NH

 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Mutagen
Probability: 0.76

Enrichment: 1.36
Bayesian Score: 1.2

Mahalanobis Distance: 7.63

Mahalanobis Distance p-value: 0.999

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	1229-55-6	3-(4'- Chlorobenzylidenamino)- 5H-1;2;3-triazin- [5;4b]indol-4-one	135086-93-0
Structure	OH N N N N N N N N N N N N N N N N N N N	N N N N N N N N N N N N N N N N N N N	CI N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.539	0.540	0.540
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Mutagenesis 7(1):37-39; 1992	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

Model Applicability

Feature Contribution

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

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

i catalo co	iid ibadioii				
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set	
SCFP_12	-1825024803	[*]:[c]1:[*]:[c]2	0.472	18 out of 19	

CCCC[c]:1:2

SCFP_12	112346096	[*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.36	1035 out of 1263
SCFP_12	1310748454	[*][c]1:[*]:[c](: [*]):s:1	0.312	126 out of 161
		ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1328855840	[*]:[cH]:[c](N[c](:n: [*]):[c](:[*]):[*]): [cH]:[*]	-0.496	1 out of 4
SCFP_12	-1272798659	[*]CCC([*])[*]	-0.466	439 out of 1225
SCFP_12	815479470		-0.452	0 out of 1

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4 Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.725 Enrichment: 1.3

Bayesian Score: -0.687 Mahalanobis Distance: 9.01

Mahalanobis Distance p-value: 0.823

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	2425-85-6	50-65-7	139953-76-7
Structure	OH NAME OF THE PARTY OF THE PAR	OH OHO	HO N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.557	0.557	0.562
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

Model Applicability

Unknown features are fingerprint features in 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 fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1728149235	AND Enantiomer N N N N N N N N N N N N N	0.442	21 out of 23

SCFP_12	112346096	AND Enantiomer S	0.36	1035 out of 1263
SCFP_12	-192131231	N [c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.318	192 out of 244
		ures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-55982834	AND Enantiomer S N N N N N N (*):[c]1:[*]C[C@@H](C)CC1	-1.82	0 out of 9
SCFP_12	-1043339860	AND Enantiomer S N N N H ₂ N [*]CC(C)C[*]	-1.09	76 out of 399
SCFP_12	1328855840	AND Examinmer N N N N N N N N N N N N N	-0.496	1 out of 4

0.557

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678
Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.473
Enrichment: 0.848
Bayesian Score: -8.11
Mahalanobis Distance: 8.95

Mahalanobis Distance p-value: 0.843

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	50264-69-2	3771-19-5	4947-27-7		
Structure	HO O O O O O O O O O O O O O O O O O O	OH O	NH O		
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen		
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen		

Model Applicability

Distance

Reference

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

0.554

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

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

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

0.501

Feature Co	Feature Contribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set		
SCFP_12	-1231050790	AND Enantomer S (*]CN[c]1:n:[cH]:[*]: [c]2:[*]:[*]:[c]([*]):[c]:1:2	0.388	3 out of 3		

SCFP_12	112346096	AND Enantiomer N	0.36	1035 out of 1263
SCFP_12	1310748454	AND Enantiomer N N N H [*][c]1:[*]:[*]:[c](: [*]):S:1	0.312	126 out of 161
		ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-55982834	AND Enantiomer N	-1.82	0 out of 9
SCFP_12	-1043339860	AND Enantiomer S N NH NH [*]CC(C)C[*]	-1.09	76 out of 399
SCFP_12	227993601	AND Enantiomer N	-0.998	0 out of 3

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Mutagen

Probability: 0.608 Enrichment: 1.09 Bayesian Score: -4.8

Mahalanobis Distance: 13.4

Mahalanobis Distance p-value: 2.93e-007

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	64808-48-6	2;3-DIHYDRO-2-(5-NITRO- 2-THIENYL)-4- QUINAZOLINE	33389-33-2		
Structure	Na Na Na Na	O=N _t NH	NH NH NH NH NH NH NH NH NH NH NH NH NH N		
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen		
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen		
Distance	0.611	0.615	0.616		
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	EMIC	Kazius et. al., J. Med. Chem. (2005) 48, 312-320		

Model Applicability

Unknown features are fingerprint features in 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 fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1848869096	AND Enantiomer N NH HN NH S NH ₂ S S S S S S S S S S S S S S S S S S S	0.38	18 out of 21
		[*]NNC(=[*])[*]		

SCFP_12	112346096	AND Enantiomer S N N N H NH2 S [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[6]:1: [*]	0.36	1035 out of 1263
SCFP_12	1310748454	AND Enantiomer N N N N N N H N NH H N S [*][c]1:[*]:[c](: [*]):s:1	0.312	126 out of 161
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-55982834	AND Enantomer N N N N N N N H H N N S [*]:[c]1:[*]C[C@@H](C)CC1	-1.82	0 out of 9
SCFP_12	-1043339860	AND Enantiomer S NH HN NH S NH [*]CC(C)C[*]	-1.09	76 out of 399
SCFP_12	382734644	AND Enantlomer S NH NH HN NH ₂ [*]NC(=S)N	-0.811	3 out of 14

NH NH O NH 2

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.423
Enrichment: 0.757
Bayesian Score: -9.21
Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 0.00178

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

98644-23-6	130-17-6	139953-77-8
	μ,	
HO _II_O		
		0
* ***********************************	N, // OH	но
Na Na) s	N N H
	//	N N N N N N N N N N N N N N N N N N N
	HO 11 0	HO SO OH OH

Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.534	0.565	0.567
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

Model Applicability

Structural Similar Compounds

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

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

	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set		
SCFP_12	112346096	AND Enantomer N N NH ₂ [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.36	1035 out of 1263		

SCFP_12	896331226	AND Enantiomer N= NH ₂ [*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]:[c]:1N[c](:[*]):[*]	0.337	6 out of 7
SCFP_12	1310748454	AND Enantiomer N	0.312	126 out of 161
		ures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-55982834	AND Enantiomer S N N N N N N N N N N N N	-1.82	0 out of 9
SCFP_12	-1358544872	AND Enantiomer NH NH S NH O NH O (*]S(=[*])(=[*])N	-1.57	1 out of 15
SCFP_12	-1463646519	AND Enantiomer S N N N N N N N N N N N N	-1.29	1 out of 11

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

AND Enantiomer N N NH NH O=S H₂N O

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.308
Enrichment: 0.552
Bayesian Score: -11.6
Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 0.00178

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	98644-23-6	130-17-6	139953-77-8	
Structure	HO II O	H ₂ N	HO HO H	
Actual Endpoint	Mutagen	Mutagen	Mutagen	
Predicted Endpoint	Mutagen	Mutagen	Mutagen	
Distance	0.539	0.567	0.571	

Model Applicability

Reference

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

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

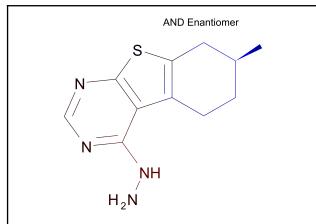
1. All properties and OPS components are within expected ranges.

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

Feature Co	Feature Contribution				
	Top fe	atures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set	
SCFP_12	112346096	AND Enantiomer N	0.36	1035 out of 1263	

SCFP_12	1310748454	AND Enantiomer N N N N N N H H N (*) [*][c]1:[*]:[c](: [*]):s:1	0.312	126 out of 161
SCFP_12	10	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.306	1774 out of 2287
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1892918731	[*]N[c]1:[cH]:[cH]:[c](:[cH]:[cH]:0) (=0)N	-2.2	0 out of 14
SCFP_12	-55982834	AND Enantiomer S N N N H H N (*]:[e]1:[*]C[C@@H](C)CC1	-1.82	0 out of 9
SCFP_12	-1358544872	AND Enantiomer N N N H H N S N H N H N H N N H N N H N N H N N H N N H N N H N N N N H N	-1.57	1 out of 15



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.712 Enrichment: 1.28 Bayesian Score: -1.25 Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.117

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Picloram	158321-20-1	96014-36-7
Structure			
	ÇI		
	}		H N
	N. CI		
		Hann	H ₂ N ^N

Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen
Distance	0.522	0.540	0.540
Reference	Mut. Res. 204: 17-115; 1988	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

Model Applicability

Structural Similar Compounds

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

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

	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set	
SCFP_12	260535663	AND Enantiomer S NH H ₂ N [*]NN	0.442	13 out of 14	

SCFP_12	-915345805	AND Enantiomer S NH H ₂ N	0.442	13 out of 14
SCFP_12	112346096	[*]:[c](:[*])NN AND Enartioner S N H ₂ NH [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.36	1035 out of 1263
	Top Fear	tures for negative	contribution)
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-55982834	AND Enantiomer S NH H ₂ N [*]:[c]1:[*]C[C@@H](C)CC1	-1.82	0 out of 9
SCFP_12	-1043339860	AND Enantiomer S NH H ₂ N [*]CC(C)C[*]	-1.09	76 out of 399
SCFP_12	-1272798659	AND Enantiomer S N NH H ₂ N [*]CCC([*])[*]	-0.466	439 out of 1225

AND Enantlomer N N N N N H S

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.658 Enrichment: 1.18 Bayesian Score: -3.27

Mahalanobis Distance: 9.58

Mahalanobis Distance p-value: 0.546

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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

Name	71422-67-8	Pigment red 2	69657-89-2
Structure	F _{rt} CI	HN NO H	HO to the total of
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen
Distance	0.693	0.699	0.721
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Environ. Mol. Mut. 19(21):1992	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

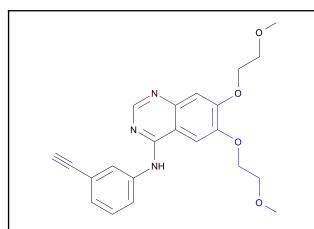
Model Applicability

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

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

	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set	
SCFP_12	112346096	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.36	1035 out of 1263	

SCFP_12	1310748454	AND Enantioner N N N N N N N N N N N N N N N N N N N	0.312	126 out of 161
SCFP_12	10	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.306	1774 out of 2287
	Top Fear	tures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-55982834	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-1.82	0 out of 9
SCFP_12	-1043339860	AND Enantiomer S N N N N N N N N N N N N N N N N N N	-1.09	76 out of 399
SCFP_12	1328855840	AND Enantiomer N S N N N N N N (*):[c](TN[c](:n: [*]):[c](:[*]):[*]): [cH]:[*]	-0.496	1 out of 4



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model	Prediction
IMOGEI	Frediction

Prediction: Non-Mutagen

Probability: 0.6
Enrichment: 1.07
Bayesian Score: -5.04
Mahalanobis Distance: 14.6

Mahalanobis Distance p-value: 6.74e-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 esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly linaccurate.

Structural Similar Compounds					
Name	Carvedilol	99522-79-9	HYCANTHONE FUROATE		
Structure		HN HN	9		

	N OH H		THE REPORT OF THE PARTY OF THE
Actual Endpoint Non-Mutagen		Non-Mutagen	Mutagen
Predicted Endpoint Non-Mutagen		Non-Mutagen	Mutagen
Distance 0.594		0.598	0.606
Reference Contrera, J.F., Matthews, E.J., Kruhlak, N.L., and Benz, R.D., Regulatory Toxicology and Pharmacology 2005, 313-323.		Kazius et. al., J. Med. Chem. (2005) 48, 312-320	EMIC

Model Applicability

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

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

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set		
SCFP_12	112346096	[*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.36	1035 out of 1263		

SCFP_12	10	[*]N[*]	0.306	1774 out of 2287
SCFP_12	-1380909229	[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.304	957 out of 1235
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1099149596	[*]OCCO[c](:[*]):[*]	-1.12	1 out of 9
SCFP_12	-677502852	N	-0.863	4 out of 19
SCFP_12	-417738003		-0.782	12 out of 48

AND Enantiomer S N N NH CI

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.419

Enrichment: 0.797
Bayesian Score: -3.75

Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.000182

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Triclabendazole	Meclofenamate Sodium (Free acid form)	Nitrofen
Structure	CI AND THE STATE OF THE STATE O	OH HN CI	O Multiple of the CI
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.531	0.601	0.613
Reference	Toxicology 43(3):283-287; 1987	Fundam Appl Toxicol 5:665-671; 1985	Environ Health Perspect 70:137-47; 1986

Model Applicability

Unknown features are fingerprint features in 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	-1181430618	AND Enantiomer N N N H CI [*]:n:[cH]:n:[*]	0.298	6 out of 8			

SCFP_6	815479470	AND Enuntiomer N N N N N (CI [*][C@H]1[*]C[o]2:[c] (C1):s:[o].([*)):[c] :2:[*]	0.271	1 out of 1
SCFP_6	1905487031	AND Enantiomer S	0.153	2 out of 3
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-300914917	AND Enumitorner N S N N N N N N N N N N N N N N N N N	-0.718	0 out of 2
SCFP_6	-1380909229	AND Enantiomer S N N N H C C I [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.449	6 out of 19
SCFP_6	951581613	AND Enantiomer S N N H CI [*]:[c](:[*])N[c](:[*]):[*]	-0.438	1 out of 4

AND Enantiomer S N NH F

C₁₇H₁₆FN₃S

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.435
Enrichment: 0.827
Bayesian Score: -3.23
Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.000905

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Triclabendazole	Meclofenamate Sodium (Free acid form)	N-Cyclohexyl-2- benzothiazylsulfenamide
Structure	CI NO THE SECOND	OH HNV	HN HN
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic
Distance	0.588	0.602	0.611
Reference	Toxicology 43(3):283-287; 1987	Fundam Appl Toxicol 5:665-671; 1985	J Appl Toxicol 9(3):187-90 1989

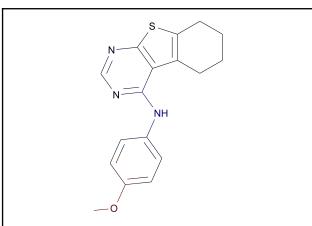
Model Applicability

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

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

Feature Contribution							
	Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set			
SCFP_6	-1181430618	AND Enantiomer S NH NH [*]:n:[cH]:n:[*]	0.298	6 out of 8			

SCFP_6	815479470	AND Enantomer N N N H F	0.271	1 out of 1
		[*][C@H]1[*]C[c]2:[c] (C1):s:[c](:[*]);[c] :2:[*]		
SCFP_6	-1551011249	AND Enantiomer N N N H H	0.271	1 out of 1
		[*]N[c]1:[cH]:[*]:[cH]:[cH]:[c]:1F		
	Top Feat	tures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1655894345	AND Enantomer N N H F [*][c]1:[cH]:[cH]:[cH]:[cH]:[c]:1F	-0.718	0 out of 2
SCFP_6	-1380909229	AND Enantomer N	-0.449	6 out of 19
SCFP_6	951581613	AND Enantiomer N= N H F [*]:[c](:[*])N[c](:[*]):[*]	-0.438	1 out of 4



C₁₇H₁₇N₃OS

Molecular Weight: 311.40138

ALogP: 4.568
Rotatable Bonds: 3
Acceptors: 4

Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.495

Enrichment: 0.941
Bayesian Score: -1.44
Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.00495

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	N-Cyclohexyl-2- benzothiazylsulfenamide	Nitrofen	Ronnel
Structure	N S = 1/3 S HN	O MO CI	Cl
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.580	0.607	0.608
Reference	J Appl Toxicol 9(3):187- 90; 1989	Environ Health Perspect 70:137-47; 1986	J Toxicol Environ Health 10:111-119; 1982

Model Applicability

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

1. OPS PC12 out of range. Value: 3.4923. Training min, max, SD, explained variance: -3.7514, 3.3159, 1.318, 0.0255.

Feature Co	Feature Contribution				
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	1237755852	CO[c]1:[cH]:[cH]:[*]: [cH]:[cH]:1	0.453	8 out of 9	

SCFP_6	591469355	[*]:[cH]:[c](OC):[cH] :[*]	0.411	10 out of 12	
SCFP_6	-1181430618	s NH NH NH NH (*]:n:[cH]:n:[*]	0.298	6 out of 8	
Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	-300914917	[*]:[c](:[*])N[c]1:[c H]:[cH]:[*]:[cH]:[cH	-0.718	0 out of 2	
SCFP_6	-1380909229	[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.449	6 out of 19	
SCFP_6	951581613	[*]:[c](:[*])N[c](:[*	-0.438	1 out of 4	

0.642

Oyo Yakuri 17(4):551-557;

 $C_{17}H_{18}N_{4}S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.336
Enrichment: 0.639

Bayesian Score: -6.97

Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.00165

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	D&C Yellow 8	N-Cyclohexyl-2- benzothiazylsulfenamide	Diflunisal	
Structure	OH OH	N S = J/J S HN	OH OH	
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic	
Predicted Endpoint	Non-Toxic	Non-Toxic	Toxic	

Model Applicability

Distance

Reference

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

0.625

90: 1989

J Appl Toxicol 9(3):187-

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

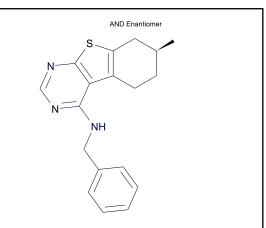
Food Chem Toxicol

24:819-823; 1986

0.614

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	-1181430618	AND Enantiomer N N N H H 2 N [*]:n:[cH]:n:[*]	0.298	6 out of 8		

SCFP_6	815479470	AND Enantiomer N N N N H ₂ N [*][C@H]1[*]C[c]2:[c] (C1):s:[c](:[*)]:[c] :2:[*]	0.271	1 out of 1	
SCFP_6	-1272798659	AND Enantiomer S N N N H ₂ N [*]CCC([*])[*]	0.0708	44 out of 78	
Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	-192131231	AND Enantiomer N N N N H H C]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.13	0 out of 4	
SCFP_6	1181470699	AND Enantiomer S N N N H 2N [*][c]1:[cH]:[c] (N):[cH]:[cH]:1	-0.718	0 out of 2	
SCFP_6	-1728149235	AND Enunitomer S N H H ₂ N [*]N[c]1:[cH]:[cH]:[c](N):[cH]:[cH]:1	-0.718	0 out of 2	



 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.427 Enrichment: 0.811 Bayesian Score: -3.51

Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.00391

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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

Name	N-Cyclohexyl-2- benzothiazylsulfenamide	N-Cyclohexyl-2- Triclabendazole benzothiazylsulfenamide		
Structure	N N HN	CI AND	OH HN CI	
Actual Endpoint	Non-Toxic	Toxic	Non-Toxic	
Predicted Endpoint	Non-Toxic	Toxic	Non-Toxic	
Distance	0.592	0.601	0.601	
Reference	J Appl Toxicol 9(3):187- 90; 1989	Toxicology 43(3):283-287; 1987	Fundam Appl Toxicol 5:665-671; 1985	

Model Applicability

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

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

Feature Contribution

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	-1181430618	AND Enantiomer S NH NH [*]:n:[cH]:n:[*]	0.298	6 out of 8		

SCFP_6	815479470	AND Enantiomer N N N N N N (*) C@H 1(*]C c 2:[c] (C1):s:[c](:[*)]:[c] :2:[*]	0.271	1 out of 1
SCFP_6	-1272798659	AND Enantiomer S N NH NH [*]CCC([*])[*]	0.0708	44 out of 78
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	AND Enantiomer N NH NH	-0.945	0 out of 3
SCFP_6	1155330592	AND Enantiomer S NH NH [*]NC[c](:[*]):[*]	-0.422	0 out of 1
SCFP_6	-55982834	AND Enardomer S N N H H (*):[c]1:[*]C[C@@H](C)CC1	-0.422	0 out of 1

0.741

Journal of Toxic Sciences

11:107-119; 1982

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Toxic Probability: 0.529 Enrichment: 1.01

Bayesian Score: -0.498 Mahalanobis Distance: 11.2

Mahalanobis Distance p-value: 0.000502

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Caffeic Acid	Azosemide	Guanabenz		
Structure	HO THE OF	N = N HN N N NH 2 O S O	H ₂ N NH		
Actual Endpoint	Toxic	Non-Toxic	Toxic		
Predicted Endpoint	Toxic	Non-Toxic	Toxic		

Model Applicability

Distance

Reference

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

0.732

5195; 1984

Kiso to Rinsho 18:5187-

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

Toxicol Appl Pharmacol 36(2):227-37; 1976

0.723

Feature Contribution						
	Top fe	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	382734644	AND Enantiomer S NH NH NH S NH NH	0.478	4 out of 4		

SCFP_6	1435188938	AND Enantlomer S NH HN NH S NH ₂ [*]C(=S)[*]	0.478	4 out of 4
SCFP_6	384861283	AND Enantiomer NH HN NH S NH S [*]C(=[*])N	0.441	3 out of 3
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-55982834	AND Enantiomer S N N N N N H H NH ₂ S [*]:[o]1:[*]C[C@@H](C)CC1	-0.422	0 out of 1
SCFP_6	-2147171373	AND Enantomer N N N N N N N N N N N N N N N N N N	-0.422	0 out of 1
SCFP_6	8	AND Enantiomer S NH NH NH S NH ₂ [*]:n:[*]	-0.278	24 out of 61

NH NH O NH 2 O

C₁₇H₁₈N₄O₂S₂

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.419
Enrichment: 0.797
Bayesian Score: -3.75
Mahalanobis Distance: 12

Mahalanobis Distance p-value: 3.47e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

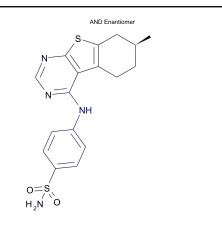
Structural Similar Compounds					
Name	D&C Yellow 8	D&C Yellow 8 Sulfonylurea Gliclazide			
Structure	OH OH	NH ONH	N N N N N N N N N N N N N N N N N N N		
Actual Endpoint	Non-Toxic	Toxic	Toxic		
Predicted Endpoint	Non-Toxic	Toxic	Toxic		
Distance	0.654	0.658	0.664		
Reference	Food Chem Toxicol 24:819-823; 1986	Yakuri to Chiryo 9:3551- 3571; 1981	Fundam Appl Toxicol 7(2):214-20; 1986		

Model Applicability

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

Feature Co	Feature Contribution						
	Top fea	tures for positive o	ontribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set			
SCFP_6	-1181430618	AND Enantiomer S NH S -NH ₂ [*]:n:[cH]:n:[*]	0.298	6 out of 8			

SCFP_6	1655803608	AND Enantiomer N= S N= N NH ₂ [*][c]1:[cH]:[cH]:[cH]]:[cH]:[c]:1S(=[*])(=[*])[*]	0.271	1 out of 1
SCFP_6	815479470	AND Enunthomer N N N N N N N N N N N N N N N N N N N	0.271	1 out of 1
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1463646519	AND Enantiomer S N N N N N N N N N N N N N N N N N N	-0.718	0 out of 2
SCFP_6	-1380909229	AND Enantomer N N N N N N N N N N N N N N N N N N	-0.449	6 out of 19
SCFP_6	951581613	AND Enantiomer N	-0.438	1 out of 4



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic

Probability: 0.411
Enrichment: 0.781
Bayesian Score: -4.03
Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 1.11e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

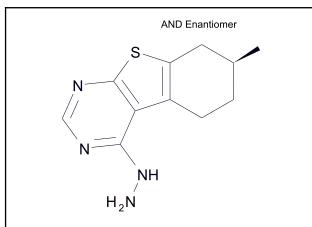
Name	Sulfonylurea Gliclazide	D&C Yellow 8	Amsacrine
Structure	NH ON NH	OH OH	N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.651	0.658	0.668
Reference	Yakuri to Chiryo 9:3551- 3571; 1981	Food Chem Toxicol 24:819-823; 1986	Fundam Appl Toxicol 7(2):214-20; 1986

Model Applicability

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

Feature Co	Feature Contribution							
	Top features for positive contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set				
SCFP_6	-1181430618	AND Enantiomer N N N H H N 1 (*) (*) (*) (*) (*) (*) (*) (*) (*)	0.298	6 out of 8				

SCFP_6	815479470	AND Enantiomer N N N N N N N N N (*)[C@H]1(*)[C]02:[c] (C1):s:[c](:[*)]:[c] :2:[*]	0.271	1 out of 1
SCFP_6	-1247518081	AND Enantiomer N= S N= N H ₂ N NS(=0)(=0)(0]1:[cH]:[cH]:[rH]:[cH]:1	0.271	1 out of 1
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-300914917	[*]:[c](*[*])N[c]1:[c H]:[cH]:[*]:[cH]:[cH	-0.718	0 out of 2
SCFP_6	-1463646519	AND Enantiomer N	-0.718	0 out of 2
SCFP_6	-1380909229	AND Enantiomer N N N N N N N N N N N N N	-0.449	6 out of 19



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.481

Enrichment: 0.915 Bayesian Score: -1.83

Mahalanobis Distance: 12.1

Mahalanobis Distance p-value: 2.12e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simil	ar Compounds		
Name	Bropirimine	3;6-Dichloropicolinic Acid	Caffeic Acid
Structure	Br N N N N N N N N N N N N N N N N N N N	CI OH OH	HO W OH
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Toxic
Distance	0.667	0.685	0.687
Reference	Teratology 38(1):7-14; 1988	Fundam Appl Toxicol 4:91-97; 1984	Toxicol Appl Pharmacol 36(2):227-37; 1976

Model Applicability

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

Feature Co	Feature Contribution					
	Top fea	tures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	-1181430618	AND Enantiomer S NH H ₂ N [*]:n:[cH]:n:[*]	0.298	6 out of 8		

SCFP_6	815479470	AND Enantiomer N N N H ₂ NH [*][C@H]1[*]C[c]2:[c]	0.271	1 out of 1
		(C1):s:[c](:[*]):[c] :2:[*]		
SCFP_6	5	AND Enantiomer NH H ₂ N	0.228	51 out of 77
		[*]N		
	Top Fea	tures for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-2147171373	AND Enantiomer S N NH H ₂ N [*INIGI4 to IGHI to I*]	-0.422	0 out of 1
		[*]N[c]1:n:[cH]:n:[*] :[c]:1:[*]		
SCFP_6	-55982834	AND Enantiomer N NH H ₂ N [*]:[c]1:[*]C[C@@H](C)CC1	-0.422	0 out of 1
SCFP_6	8	AND Enantiomer S NH H ₂ N [*]:n:[*]	-0.278	24 out of 61

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.435

Enrichment: 0.827 Bayesian Score: -3.24

Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 5.91e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar	Similar Compounds				
Name	Bromofenofos	Estramustine Phosphate Disodium (Free acid form)	Triclabendazole		
Structure	Br A.Br				

Name	Bromotenotos	Disodium (Free acid form)	Triciabendazoie
Structure	Br M OH OH	CI NO OH	CI NO THE STATE OF
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.761	0.793	0.838
Reference	Toxicol Letters 31:243- 247; 1986	Oyo Yakuri 20(6):1219- 1236; 1980	Toxicology 43(3):283-287; 1987

Model Applicability

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

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

SCFP_6	815479470	AND Enuntioner N N N N N N N N N N N N N N N N N N	0.271	1 out of 1
SCFP_6	-1272798659	AND Enantioner N N N N N N N N N N N N N N N N N N N	0.0708	44 out of 78
		tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-300914917	AND Enuntioner N S	-0.718	0 out of 2
SCFP_6	-1380909229	AND Enantionmer N S N NH S NH [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.449	6 out of 19
SCFP_6	951581613	AND Enantionner N S . N N N N N N N N N N N N N N N N N N	-0.438	1 out of 4

TOPKAT_Developmental_Toxicity_Potential

N	
NH O	

 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.504

Enrichment: 0.959
Bayesian Score: -1.17

Mahalanobis Distance: 8.76

Mahalanobis Distance p-value: 0.258

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Nicardipine	Suxibuzone	Etofenamate	
Structure		HO	HO O O O O O O O O O O O O O O O O O O	
Actual Endpoint	Non-Toxic	Toxic	Non-Toxic	
Predicted Endpoint	Non-Toxic	Toxic	Non-Toxic	
Distance	0.608	0.620	0.635	
Reference	Kiso to Rinsho 13:1149- 1159; 1979	Oyo Yakuri 20:377-386; 1980	lyakuhin Kenkyu 13(4):896-909; 1982	

Model Applicability

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

Feature Co	Feature Contribution				
	Top fe	atures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	123285475	[*]O[c]1:[cH]:[c](:n: [*]):[c](:[*]):[*]:[c]:1[*]	0.478	4 out of 4	

SCFP_6	446954673	[*]CO[c]1:[cH]:[c]2:n :[cH]:[*]:[c]([*]):[c]:2:[cH]:[c]:10[*]	0.381	2 out of 2
SCFP_6	-1814968949	[*]CO[c]1:[cH]:[c]2:[c](N[*]):n:[*]:n:[c] :2:[cH]:[c]:10[*]	0.381	2 out of 2
		tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-2020651081	[*][c]1:[cH]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.718	0 out of 2
SCFP_6	2142015375	[*]:[cH]:[c]1:n:[cH]: n:[*]:[c]:1:[*]	-0.718	0 out of 2
SCFP_6	-300914917	[*]:[c](:[*])N[c]1:[c OH]:[cH]:[cH]:[*]:[cH]:[*]	-0.718	0 out of 2

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

AND Enantiomer S N N NH CI

 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.274
Enrichment: 0.855
Bayesian Score: 0.526
Mahalanobis Distance: 13.5

Mahalanobis Distance p-value: 8.88e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Levonorgestrel	Indomethacin
Structure	OH	OH OH	HO
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.629	0.647	0.652
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: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	ntribution			
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	51876938	AND Enantiomer N N N N N N CI [*]CC[c](:[*]):[*]	0.473	16 out of 31

ECFP_6	-1661653144	AND Enantiomer N N N N N N N (*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.442	2 out of 3
ECFP_6	553181281	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.424	1 out of 1
	Top Fea	tures for negative of	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1242906247	AND Enantiomer N N H CI [*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2
ECFP_6	-176494269	AND Enantiomer S N N N H C [*]:[cH]:[c](CI):[cH] :[*]	-0.476	5 out of 28
ECFP_6	99947387	AND Enandomer N N N N H CI [*]:[c](:[*])CI	-0.461	9 out of 48

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

AND Enantiomer S N NH F

C₁₇H₁₆FN₃S

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.305
Enrichment: 0.951
Bayesian Score: 1.75
Mahalanobis Distance: 11.3

Mahalanobis Distance p-value: 0.0631

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Nafenopin	Levonorgestrel	Diclofenac		
Structure	OH	OH OH	OH OH OH OH		
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen		
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen		
Distance	0.613	0.640	0.654		

Model Applicability

Reference

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[*]:[c]:1:[*]

Feature Co	ntribution			
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	51876938	AND Enantiomer S NH NH [*]CC[c](:[*]):[*]	0.473	16 out of 31

ECFP_6	-1661653144	AND Enantiomer N H F [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.442	2 out of 3
ECFP_6	-1448786963	AND Enantioner S N H F [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.424	1 out of 1
		ures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1242906247	AND Enantiomer S N H F [*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2
ECFP_6	-1311285389	AND Enantioner S N H F [*][c](:[*]):[c](F):[cH]:[*]	-0.459	1 out of 7
ECFP_6	-1672647522	AND Enantioner N N H F [*]C[c]1:s:[*]:[c]]:1[*]	-0.27	0 out of 1

N NH NH

C₁₇H₁₇N₃OS

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.244
Enrichment: 0.761
Bayesian Score: -1.06
Mahalanobis Distance: 12.6

Mahalanobis Distance p-value: 0.00166

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Indomethacin	Nafenopin	Phenolphthalein		
Structure	HO	OH OH	НО		
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen		
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen		
Distance	0.579	0.624	0.649		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &		

Model Applicability

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

Res.) Sept. 1997

Res.) Sept. 1997

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

Res.) Sept. 1997

2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	ntribution			
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	51876938	[*]CC[c](:[*]):[*]	0.473	16 out of 31

ECFP_6	-1661653144	[*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.442	2 out of 3
ECFP_6	-1448786963	[*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.424	1 out of 1
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1271104377	CO[c]1:[cH]:[cH]:[*]: [cH]:[cH]:1	-0.805	0 out of 4
ECFP_6	693720869	[*][c]1:[cH]:[cH]:[c] (OC):[cH]:[cH]:1	-0.805	0 out of 4
ECFP_6	1307307440	[*]:[c](:[*])OC	-0.558	4 out of 25

 $C_{17}H_{18}N_{4}S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.33
Enrichment: 1.03
Bayesian Score: 2.59
Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.035

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

	•		
Name	Niclosamide	Phenolphthalein	Pyrimethamine
Structure	CI AND H	НО	NH 2 N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.568	0.581	0.626
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: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Contribution

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	51876938	AND Enantiomer S N N N H ₂ N [*]CC[c](:[*]):[*]	0.473	16 out of 31	

ECFP_6	-1661653144	AND Enantiomer N N N H [*][c]([*]**]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.442	2 out of 3
ECFP_6	-1448786963	AND Enantiomer S N H H 12 [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.424	1 out of 1
	Top Featu	res for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1242906247	AND Enandomer N N N N H 2N H 2N [*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2
ECFP_6	-1672647522	AND Enantlomer N N N N N H ₂ N [*]C[c]1:s:[*]:[c]:1[*]	-0.27	0 out of 1
ECFP_6	-1660340418	AND Enantiomer N N N N N N H H 12,N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.232	2 out of 9

 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.35
Enrichment: 1.09
Bayesian Score: 3.17
Mahalanobis Distance: 12.9

Mahalanobis Distance p-value: 0.000738

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

	•		
Name	Nafenopin	Pergolide	Oxaprocin
Structure	OH OH	Z H	HO HO
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.566	0.613	0.617
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: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Contribution

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	769925792	AND Enantiomer NH NH [*]NC[c](:[*]):[*]	0.617	2 out of 2

ECFP_6	51876938	AND Enantiomer S NH NH [*]CC[c](:[*]):[*]	0.473	16 out of 31
ECFP_6	-1661653144	AND Enantiomer N N N N N N H H (*)[c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.442	2 out of 3
	Top Feat	tures for negative of	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1672647522	AND Enantiomer N N N N H	-0.27	0 out of 1
ECFP_6	-1660340418	AND Enantiomer S N N N H	-0.232	2 out of 9
ECFP_6	1559650422	AND Enantiomer N NH NH [*]C[*]	-0.164	50 out of 191

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4 Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.415 Enrichment: 1.29 Bavesian Score: 4.8

Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.00594

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Furosemide	Phenazopyridine	Niclosamide
Structure	HO O H	NH ₂ N _N	IH ₂
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen

Model Applicability

Predicted Endpoint

Distance

Reference

Churchinal Cimilar Campainda

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

Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.729

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.734

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown ECFP_2 feature: -571028867: [*]NC(=S)N

Carcinogen

Res.) Sept. 1997

0.715

ECFP_6	51876938	AND Enantiomer NH HN NH S NH S [*]CC[c](:[*]):[*]	0.473	16 out of 31
ECFP_6	-1661653144	AND Enantiomer N N N N N H H N NH H N NH (1):[a]:[a]:[c]:1:[c] ([a]):[a]:[a]:[a]:[a]: [b]:[a]:[a]:[a]: [c]:[a]:[a]:[a]:[a]: [c]:[a]:[a]:[a]:[a]:[a]:[a]:[a]:[a]:[a]:[a	0.442	2 out of 3
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1672647522	AND Enantioner N N N N N H N NH2 S [*]C[c]1:s:[*]:[c]:1[*]	-0.27	0 out of 1
ECFP_6	-1660340418	AND Enantioner N N N N N N N N N N N N N N N N N N	-0.232	2 out of 9
ECFP_6	85262808	AND Enantioner S N N N N N N N N N N N N N N N N N N	-0.112	1 out of 4

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.249
Enrichment: 0.776
Bayesian Score: -0.764
Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.0367

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Indapamide	Metolazone	Niclosamide	
Structure	HN ZZ H ₂ N O	H ₂ N S the NH	CIANOH	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Distance	0.579	0.616	0.616	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set ECFP_6 51876938 0.473 16 out of 31

ECFP_6	-1661653144	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.442	2 out of 3
ECFP_6	-1448786963	AND Enantiomer S N N NH2 [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.424	1 out of 1
	Top Featu	res for negative of	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1242906247	AND Enantiomer S N N N N N N N N N N N N	-0.482	0 out of 2
ECFP_6	1634465935	AND Enantomer N N N N N NH ₂ [*][c]1:[*]:[cH]:[cH] :[cH]:[c]:1S(=[*])(= [*])[*]	-0.27	0 out of 1
ECFP_6	-1114960102	AND Enantiomer N N N N N N N N N N 1 C NH ₂ [*][c]1:[cH]:[cH]:[cH]:[cH]:[c]:1S(=[*])(=[*])[*]	-0.27	0 out of 1

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.286
Enrichment: 0.893
Bayesian Score: 1.05
Mahalanobis Distance: 12.1

Mahalanobis Distance p-value: 0.00933

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	ilar Compounds Indapamide	Niclosamide	Metolazone
Structure	HN Z SO	CIANOH HNWO	H ₂ N NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen

Model Applicability

Distance

Reference

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

0.620

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

0.633

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

0.596

2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	Feature Contribution				
	Top fe	atures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	51876938	AND Enantiomer N N N H H 2N [*]CC[c](:[*]):[*]	0.473	16 out of 31	

ECFP_6	-1661653144	[*][c]('t**)*[c]1:[c] ([*]):[*]:[*]:[c]:1:	0.442	2 out of 3
ECFP_6	1049768340	AND Enantiomer S H, N H, N [*]N[c](:n:[*]):[c](: [*]):[*]	0.424	1 out of 1
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1242906247	AND Enantiomer S N N N N N N N N N N N N	-0.482	0 out of 2
ECFP_6	-1672647522	AND Enantiomer N N N N H H N S 1 (*)C[o]1:S:[*]:[*]:[c]:1[*]	-0.27	0 out of 1
ECFP_6	-797085356	AND Enantiomer N N N N H H 2N [*]S(=[*])(=[*])[*]	-0.263	5 out of 22

US FDA (Centre for Drug

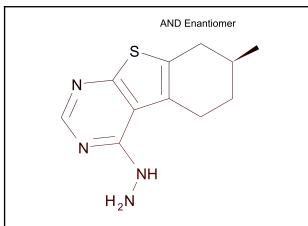
Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561
Rotatable Bonds: 1
Acceptors: 4

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.407 Enrichment: 1.27 Bayesian Score: 4.63 Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.0391

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Lamotrigine	Phenazopyridine	Dapsone		
Structure	CI N N N N N N N N N N N N N N N N N N N	N N N N N N N N N N N N N N N N N N N	H ₂ N OF S		
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen		
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen		
Distance	0.571	0.586	0.601		

Model Applicability

Reference

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	Feature Contribution				
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	-934039951	AND Enantiomer S N NH H ₂ N [*]NN	0.581	3 out of 4	

ECFP_6	51876938	AND Enantiomer S NH H ₂ N [*]CC[c](:[*]):[*]	0.473	16 out of 31
ECFP_6	-1661653144	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.442	2 out of 3
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1672647522	AND Enantiomer NH H ₂ N [*]C[c]1:s:[*]:[c]:1[*]	-0.27	0 out of 1
ECFP_6	-1660340418	AND Enantiomer S NH H ₂ N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.232	2 out of 9
ECFP_6	85262808	AND Enantiomer NH H ₂ N [*][c]1:[*]:[*]:[c](: [*]):S:1	-0.112	1 out of 4

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

AND Enantiomer S N N NH	
N H H	

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.318
Enrichment: 0.993
Bayesian Score: 2.21
Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 0.000434

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Fluticasone	Emetine	Bromocriptine		
Structure	HO state of the st		Br N N N N N N N N N N N N N N N N N N N		
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Distance	0.878	0.893	0.907		
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: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	ntribution			
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	51876938	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.473	16 out of 31

ECFP_6	-1661653144	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.442	2 out of 3
ECFP_6	553181281	AND Ensettioner N S N N N N N N N N N N N N N N N N N	0.424	1 out of 1
		tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1242906247	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.482	0 out of 2
ECFP_6	-1672647522	AND Enantomer N N N N N N N N N N N N N N N N N N N	-0.27	0 out of 1
ECFP_6	-1660340418	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.232	2 out of 9

	N O
N	O NH

 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.235
Enrichment: 0.733
Bayesian Score: -1.68
Mahalanobis Distance: 15.3

Mahalanobis Distance p-value: 7.99e-008

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Mycophenolate	Nicardipine	Nimodipine	
Structure	HOTHO	I N O O O O O O O O O O O O O O O O O O	H N N N N N N N N N N N N N N N N N N N	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Distance	0.632	0.664	0.675	
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: -182178874: [*]#C[c](:c:[*]):c:[*]
- 3. Unknown ECFP_2 feature: 1139738044: [*]:[c](:[*])C#C

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1939823063	\	0.866	8 out of 9

ECFP_6	-1545539812	N • 0	0.866	8 out of 9
		N H O		
		[*]C#C		
ECFP_6	-1114776580		0.755	11 out of 15
		N H O (
	Top Foatur	es for negative c	ontribution	
Fingerprint	Bit/Smiles		Score	Carcinogen in
				Carcinogen in training set
ECFP_6	2007300961	[*][c]1:[*]:[c]([*]): [cH]:[cH]:1	-0.652	5 out of 34
ECFP_6	-2063202154	[*]O[c]1:[cH]:[c](:n: [*]):[c](:[*]):[*]:[c]:1[*]	-0.482	0 out of 2
ECFP_6	-1242906247	[*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2

AND Enantiomer S N NH CI

 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.415
Enrichment: 1.01
Bayesian Score: 0.333
Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 3.16e-007

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Nafenopin	Nafenopin Levonorgestrel			
Structure	OH	OH CHARLES OH	₩ HO		
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen		
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen		
Distance	0.612	0.641	0.655		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &		

Model Applicability

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

Res.) Sept. 1997

Res.) Sept. 1997

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

Res.) Sept. 1997

- 2. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown ECFP_2 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
ECFP_4	553181281	AND Ensistement N N N N N N H CI [*][C@@H]1(*][c]2:[*] ::[*]:[c](-[*]):[c]:2 CC1	0.351	1 out of 1	

ECFP_4	-710237522	AND Enantiomer	0.351	1 out of 1
		N N N N N N N N N N N N N N N N N N N		
		[*]:n:[cH]:n:[*]		
ECFP_4	85262808	AND Enantiomer	0.351	1 out of 1
		[*][c]1:[*]:[c](:		
		[*]):s:1		
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	888054369	AND Enantiomer S N N N H	-0.8	0 out of 3
		ci [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1		
ECFP_4	-1661653144	AND Enantiomer N N N N H	-0.597	0 out of 2
		[*][c](:[^c]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]		
ECFP_4	-1897341097	AND Enantiomer	-0.356	8 out of 29
		CI [*11.1F*1		
		[*]N[*]		

AND Enantiomer S N NH NH

 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.472 Enrichment: 1.15 Bayesian Score: 1.6

Mahalanobis Distance: 13.5

Mahalanobis Distance p-value: 1.91e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Levonorgestrel	Pergolide
Structure	OH	OH OH	S T T T T T T T T T T T T T T T T T T T
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.593	0.636	0.655
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: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown ECFP_2 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
ECFP_4	220735655	AND Enantiomer S NH NH [*]:[c](:[*])F	0.497	4 out of 5	

ECFP_4	-710237522	AND Enantiomer S NH NH F [*]:n:[cH]:n:[*]	0.351	1 out of 1
ECFP_4	85262808	AND Enantiomer N N F [*][c]1:[*]:[*]:[c](: [*]):S:1	0.351	1 out of 1
	Top Feat	ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-1661653144	AND Enantiomer S N N N H F [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.597	0 out of 2
ECFP_4	-1897341097	AND Enantiomer S N N N F [*]N[*]	-0.356	8 out of 29
ECFP_4	292958156	AND Enantiomer S NH NH [*]CC(C)C[*]	-0.342	0 out of 1

 $C_{17}H_{18}N_{4}S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4 Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.305 Enrichment: 0.744 Bayesian Score: -2.95 Mahalanobis Distance: 15.9

Mahalanobis Distance p-value: 1.43e-007

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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

Name	Phenolphthalein	Oxazepam	Phenazopyridine
Structure		CI N N OH	H ₂ N ^N NH ₂
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.578	0.655	0.715
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. 1.
- 2. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*] 3.
- Unknown ECFP_2 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
ECFP_4	553181281	AND Enumitorner N N N N H ₂ N [*][C@@H]1[*][c]2:[*] :[*]:[c]:(*]):[c]:2 CC1	0.351	1 out of 1		

ECFP_4	-710237522	AND Enantiomer S N N H ₂ N [*]:n:[cH]:n:[*]	0.351	1 out of 1
ECFP_4	85262808	AND Enantiomer N N N N N H H 2N [*][c]1:[*]:[*]:[c](: [*]):s:1	0.351	1 out of 1
		res for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	888054369	AND Exantomer N N N H H 2N [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.8	0 out of 3
ECFP_4	924664308	AND Enantiomer N N N H H N [*][c]1:[*]:[cH]:[c](N):[cH]:[cH]:1	-0.597	0 out of 2
ECFP_4	-1661653144	AND Enantomer N N N N H [*][C]([**]):[C]1:[C] ([*]):[*]:[*]:[C]:1: [*]	-0.597	0 out of 2

 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678
Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.439
Enrichment: 1.07
Bayesian Score: 0.904
Mahalanobis Distance: 13.8

Mahalanobis Distance p-value: 1.14e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural	Similar	Compounds	

Name	Nafenopin	Pergolide	Levonorgestrel
Structure	OH	Z H	OH NATURAL OH
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.547	0.602	0.651
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: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
ECFP_4	110318898	AND Enantomer N N N	0.351	1 out of 1	

ECFP_4	-710237522	AND Enantiomer S NH NH (*):n:[cH]:n:[*]	0.351	1 out of 1
ECFP_4	85262808	[*][c]1:[*]:[c](: [*]):s:1	0.351	1 out of 1
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-1661653144	AND Enantiomer N	-0.597	0 out of 2
ECFP_4	-1897341097	AND Enantiomer N N NH NH [*]N[*]	-0.356	8 out of 29
ECFP_4	292958156	AND Enantiomer N N N N N [*]CC(C)C[*]	-0.342	0 out of 1

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.338
Enrichment: 0.825
Bayesian Score: -1.78
Mahalanobis Distance: 14.7

Mahalanobis Distance p-value: 1.82e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

	•		
Name	Furosemide	Phenazopyridine	Guanabenz
Structure	HO O NH2 NH2 NH2 O O O O O O O O O O O O O O O O O O O	N N N N N N N N N N N N N N N N N N N	CI NH H ₂ N NH
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.708	0.744	0.758
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: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown ECFP_2 feature: -571028867: [*]NC(=S)N

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
ECFP_4	1635339976	AND Enantomer S NH HN NH S [*]NNC(=[*])[*]	0.351	1 out of 1	

ECFP_4	553181281	AND Enantomer S N N N N H H NH ₂ [*][C@@H]1(*]:[o]2:[*] :[*]:[o]:[-1]:[o]:2 CC1	0.351	1 out of 1
ECFP_4	-1448786963	AND Enantiomer S N N N N H NH2 S [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.351	1 out of 1
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-1661653144	AND Enantiomer S N N N N H H NH ₂ [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.597	0 out of 2
ECFP_4	-932108170	AND Enantiomer S NH HN NH HN S NH ₂ [*]C(=[*])N	-0.545	1 out of 6
ECFP_4	-1897341097	AND Enantiomer S NH HN NH ₂ [*]N[*]	-0.356	8 out of 29

TOPKAT_Mouse_Female_FDA_Single_vs_Multiple

AND Enantiomer
NH O NH 2 O

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.265
Enrichment: 0.646
Bayesian Score: -4.84
Mahalanobis Distance: 14.3

Mahalanobis Distance p-value: 3.91e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Phenolphthalein	Sulfamethazine	Furosemide
Structure	НО	HN THE SOLUTION OF THE SOLUTIO	HO O NH2
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.719	0.719	0.719
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: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown ECFP_2 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	553181281	AND Enantionner N N N N N N N N N N N N N N 1 (NH ₂ (1) C@@H]1[1] c] 2:[1] :[1]: c]:(1] s c]:2 CC1	0.351	1 out of 1

ECFP_4	-1448786963	AND Enantioner S N N N N N N N N N N N N	0.351	1 out of 1
ECFP_4	1635992319	AND Enantiomer S N N N N N N N N N N N N	0.351	1 out of 1
	Top Feat	ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	2102150379	AND Enantiomer NH NH O S'-NH ₂ O [*]S(=[*])(=O)[*]	-1.11	0 out of 5
ECFP_4	-797085356	AND Enantiomer S NH O S NH ₂ O O [*]S(=[*])(=[*])[*]	-1.11	0 out of 5
ECFP_4	-1661653144	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.597	0 out of 2

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.229
Enrichment: 0.558
Bayesian Score: -7.71
Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 3.18e-007

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Sulfamethazine	Phenolphthalein	Furosemide
Structure	HN NH 2	НО	HO O NH2 O SO
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.700	0.732	0.744
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: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown ECFP_2 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-1448786963	AND Enantiomer S N N N H 2N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.351	1 out of 1

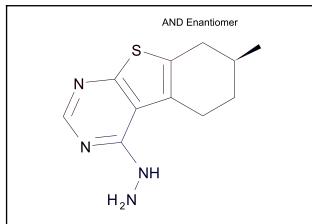
ECFP_4	-710237522	AND Enantomer N N H OLUMN H ₂ N [*]:n:[cH]:n:[*]	0.351	1 out of 1
ECFP_4	553181281	AND Enantiomer N N N H H N (*)[C@@H]1(*)[c]2:(*) :[*]:[o]:(*)]:[c]:2 CC1	0.351	1 out of 1
	Top Fea	tures for negative	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-797085356	AND Enantiomer N N N N H H N H (*) (*) (*) (*) (*) (*) (*) (*) (*) (*)	-1.11	0 out of 5
ECFP_4	2102150379	AND Enantomer N N N N H O H 2 N (1) (1) (2) (3) (4) (5) (7) (7) (8) (8) (9) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	-1.11	0 out of 5
ECFP_4	-177264675	AND Enantiomer N N N N N H N (*)S(=[*])(=[*])[c](: [cH]:[*]).[cH]:[*]	-0.968	0 out of 4

0.638

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.327 Enrichment: 0.798 Bayesian Score: -2.14 Mahalanobis Distance: 15.2

Mahalanobis Distance p-value: 6.92e-007

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simi	nar Compounds		
Name	Phenazopyridine	Hydralazine	Proflavine
Structure	N N N N N N N N N N N N N N N N N N N	N NH 2 H	H ₂ N ⁴ r ⁴ NH ₂
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen

Model Applicability

Distance

Reference

Structural Similar Compounds

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

0.608

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

0.584

- 2. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	ntribution			
	Top features for positive contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-1448786963	AND Enantiomer S N NH H ₂ N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.351	1 out of 1

ECFP_4	-710237522	AND Enantiomer S NH H ₂ N [*]:n:[cH]:n:[*]	0.351	1 out of 1
ECFP_4	85262808	AND Enantiomer NH H ₂ N [*][c]1:[*]:[*]:[c](: [*]):S:1	0.351	1 out of 1
	Top Feat	ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-1661653144	AND Enantlomer S N N H ₂ NH [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.597	0 out of 2
ECFP_4	-1897341097	AND Enantiomer S NH H ₂ N [*]N[*]	-0.356	8 out of 29
ECFP_4	1049768340	AND Enantiomer N NH H ₂ N [*]N[c](:n:[*]):[c](: [*]):[*]	-0.342	0 out of 1

 $|C_{28}H_{28}N_6S_2|$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.381
Enrichment: 0.929
Bayesian Score: -0.526
Mahalanobis Distance: 16.4

Mahalanobis Distance p-value: 6e-008

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Glimepride	Simvastatin	Lovastatin
Structure	NH NH		

	0"	ОН	ОН
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.899	0.911	0.940
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

Structural Similar Compounds

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown ECFP_2 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	-1448786963	AND Enantiomer N N NH S NH S (*)[c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.351	1 out of 1

ECFP_4	85262808	AND Enantiomer N S N N N N N N N N N N N N N N N N N	0.351	1 out of 1
ECFP_4	-710237522	AND Enantiomer N S N N N N N N N N N N N N N N N N N	0.351	1 out of 1
	Top Feat	tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
ECFP_4	888054369	**No Enantioner	-0.8	0 out of 3
ECFP_4	-1661653144	AND Enurtionmer N	-0.597	0 out of 2
ECFP_4	-1897341097	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.356	8 out of 29

AND Enantiomer S N NH CI

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.266
Enrichment: 0.904
Bayesian Score: -1.58
Mahalanobis Distance: 15.4

Mahalanobis Distance p-value: 1.47e-008

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Indomethacin	Indomethacin Nafenopin		
Structure	HO	OH	HO THO	
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen	
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen	
Distance	0.619	0.622	0.639	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	

Model Applicability

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

1. OPS PC19 out of range. Value: -3.237. Training min, max, SD, explained variance: -2.8152, 4.6113, 1.185, 0.0147.

Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set FCFP_6 -1441604640 0.77 4 out of 5

FCFP_6	-773983804	AND Enantiomer N N N N N H CI [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.409	10 out of 24
FCFP_6	-1043339860	AND Enantiomer N N N N N H CI [*]CC(C)C[*]	0.383	24 out of 61
	Top Feat	tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	AND Enantiomer S N H CI [*]:n:[cH]:n:[*]	-0.731	1 out of 12
FCFP_6	1293778554	AND Enantiomer N S N N N (C) (C) ([*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	AND Enantiomer N S N N H CI [*][c]1:[*]:[cH]:[c](CI):[cH]:[cH]:1	-0.433	8 out of 49

N NH NH

C₁₇H₁₆FN₃S

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.273 Enrichment: 0.926 Bayesian Score: -1.3

Mahalanobis Distance: 17.2

Mahalanobis Distance p-value: 2.45e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Diclofenac	Nafenopin	Indomethacin
Structure	OH O	ОН	HO
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.600	0.610	0.621
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

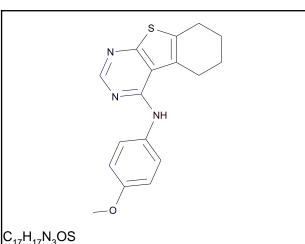
Model Applicability

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

1. OPS PC19 out of range. Value: -2.8919. Training min, max, SD, explained variance: -2.8152, 4.6113, 1.185, 0.0147.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1441604640	AND Enantiomer S N H F [*]:[c]1:[*]C[C@@H](C)CC1	0.77	4 out of 5

FCFP_6	-773983804	AND Enantiomer S H F [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.409	10 out of 24
FCFP_6	-1043339860	AND Enantiomer S N NH F [*]CC(C)C[*]	0.383	24 out of 61
	Top Feat	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-2095752315	AND Enantiomer N	-1.13	0 out of 8
FCFP_6	-124685461	AND Enantomer S NH NH [*]:n:[cH]:n:[*]	-0.731	1 out of 12
FCFP_6	1293778554	AND Enantiomer N H	-0.719	0 out of 4



Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.232 Enrichment: 0.789 Bayesian Score: -3.15 Mahalanobis Distance: 15.4

Mahalanobis Distance p-value: 1.79e-008

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Indomethacin	Indomethacin Nafenopin		
Structure	HO	OH	HO	
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen	
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen	
Distance	0.540	0.597	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. OPS PC19 out of range. Value: -3.0431. Training min, max, SD, explained variance: -2.8152, 4.6113, 1.185, 0.0147.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1060187936	[*]N[c]1:[cH]:[cH]:[c]](OC):[cH]:[cH]:1	0.46	1 out of 1

FCFP_6	-773983804	[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.409	10 out of 24
FCFP_6	590925877	[*]N[c](:[cH]:[*]):[c H]:[*]	0.369	13 out of 33
	Top Feat	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	**************************************	-0.731	1 out of 12
FCFP_6	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	-9847677	[*][c]1:[cH]:[c] (OC):[cH]:[cH]:1	-0.719	0 out of 4

 $C_{17}H_{18}N_{4}S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.339
Enrichment: 1.15
Bayesian Score: 1.16
Mahalanobis Distance: 13.3

Mahalanobis Distance p-value: 8.27e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Niclosamide	Phenolphthalein	Pyrimethamine
Structure	CIANDOH OH	НО	NH 2 N N N C
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.548	0.572	0.594
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

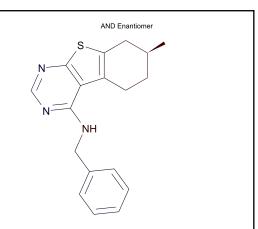
Model Applicability

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	-1441604640	AND Enantiomer S N N H H P 1':[c]1:[']C[C@@H](C)CC1	0.77	4 out of 5

FCFP_6	-773983804	AND Enantiomer N N H ₂ N [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.409	10 out of 24
FCFP_6	-1043339860	AND Enantiomer S N N H H CC(C)C[*]	0.383	24 out of 61
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	AND Enantiomer N N H H 2N [*]:n:[cH]:n:[*]	-0.731	1 out of 12
FCFP_6	1293778554	AND Enantiomer N= N+ H ₂ N [*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	-475316933	AND Enantiomer N N H ₂ N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	-0.423	0 out of 2



C₁₈H₁₉N₃S

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.325 Enrichment: 1.1

Bayesian Score: 0.677
Mahalanobis Distance: 14.1

Mahalanobis Distance p-value: 4.11e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Indomethacin	Pergolide
Structure		OH CI	S

		// °	H
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.548	0.591	0.601
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

Structural Similar Compounds

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

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

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1441604640	AND Enantiomer N N H H [*]:[c]1:[*]C[C@@H](C)CC1	0.77	4 out of 5

FCFP_6	1294255210	AND Enantiomer S N NH NH [*1CAL[c1(-[*1])-[*1]	0.441	12 out of 28
FCFP_6	-1043339860	[*]CN[c](:[*]):[*] AND Enantiomer N N N N N [*]CC(C)C[*]	0.383	24 out of 61
	Top Featu	res for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	AND Enantiomer NH NH [*]:n:[cH]:n:[*]	-0.731	1 out of 12
FCFP_6	-475316933	AND Enantiomer S N N N H I*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	-0.423	0 out of 2
FCFP_6	907007053	AND Enantiomer N N H H [*]C([*])C[c](:[*]):[*]	-0.366	11 out of 62

0.715

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4 Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.397 Enrichment: 1.35 Bayesian Score: 2.93 Mahalanobis Distance: 14

Mahalanobis Distance p-value: 7.25e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Furosemide	Phenazopyridine	Dapsone
Structure	HO O NH 2 NH 2 S O S O C I	N N N N N N N N N N N N N N N N N N N	H ₂ N O O O O O O O O O O O O O O O O O O O
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen

Model Applicability

Distance

Reference

0.683

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

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

0.712

US FDA (Centre for Drug

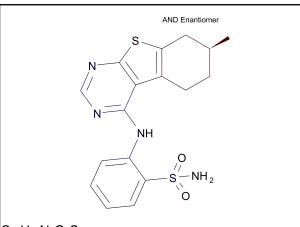
Res.) Sept. 1997

Eval.& Res./Off. Testing &

1. OPS PC30 out of range. Value: -2.6109. Training min, max, SD, explained variance: -2.5648, 3.3898, 0.9347, 0.0092.

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
FCFP_6	-1441604640	AND Enantiomer S N N N N N N H H N NH ₂ S ["]:[c]1:["]C[C@@H](C)CC1	0.77	4 out of 5		

FCFP_6	-885461129	AND Enantiomer S N NH HN NH ₂ S [*]NNC(=[*])[*]	0.547	3 out of 5
FCFP_6	1294344583	AND Enantiomer S NH HN NH S NH ₂ [*]NN[c](:[*]):[*]	0.517	2 out of 3
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	AND Enantiomer S NH HN NH S NH ₂ S [*]:n:[cH]:n:[*]	-0.731	1 out of 12
FCFP_6	-475316933	AND Enantomer N N N N N N N N N N N N N N N N N N	-0.423	0 out of 2
FCFP_6	907007053	AND Enantiomer S	-0.366	11 out of 62



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.286
Enrichment: 0.973
Bayesian Score: -0.74
Mahalanobis Distance: 15.1

Mahalanobis Distance p-value: 6.79e-008

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Indapamide	Niclosamide	Metolazone
Structure	HN H ₂ N O	CI MOH	H ₂ N ² S th
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.584	0.590	0.609
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing

Model Applicability

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

Res.) Sept. 1997

Res.) Sept. 1997

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

Res.) Sept. 1997

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	-1441604640	AND Enantiomer S N N N N N N N N N N N N N N N C N N N C N N C N C	0.77	4 out of 5	

FCFP_6	-773983804	AND Enartioner N N N N N N N N N N N N N N N N N N	0.409	10 out of 24
FCFP_6	-1043339860	AND Enantiomer S NH NH S NH S NH ₂ [*]CC(C)C[*]	0.383	24 out of 61
	Top Featur	es for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	AND Enantomer NH S NH S NH C S S NH C S S S NH C S S S S S S S S S S S S	-0.731	1 out of 12
FCFP_6	1293778554	AND Enantiomer S N N NH ₂ [*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	-103236997	AND Enuntioner S H NH ₂ [*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]:[cH]: [c]:1N[c](:[*]):[*]	-0.582	0 out of 3

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.618

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541
Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.321
Enrichment: 1.09
Bayesian Score: 0.528
Mahalanobis Distance: 14.5

Mahalanobis Distance p-value: 7.14e-007

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Niclosamide	Indapamide	Metolazone	
Structure	CIANOH HNWO	HN N N O N O N O N O N O N O N O N O N O	H ₂ N S 1 NH NH	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.588

Model Applicability

Predicted Endpoint

Distance

Reference

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

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

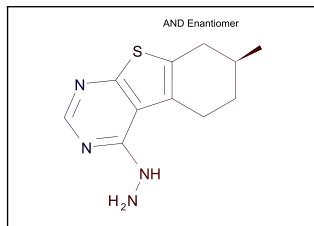
0.593

- 1. OPS PC19 out of range. Value: -2.8409. Training min, max, SD, explained variance: -2.8152, 4.6113, 1.185, 0.0147.
- 2. OPS PC30 out of range. Value: -2.7012. Training min, max, SD, explained variance: -2.5648, 3.3898, 0.9347, 0.0092.

Feature Contribution

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	-1441604640	AND Enantiomer N N N N N N H 2 N C [*]:[c]1:[*]C[C@@H](C)CC1	0.77	4 out of 5	

FCFP_6	-773983804	**No Enantioner** **No Enantion	0.409	10 out of 24
FCFP_6	-1043339860	AND Enantiomer S N N N H O'S H ₂ [*]CC(C)C[*]	0.383	24 out of 61
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	AND Enantiomer S N N H H S (*) (*) (*) (*) (*) (*) (*) (*) (*) (*)	-0.731	1 out of 12
FCFP_6	1293778554	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.719	0 out of 4
FCFP_6	-475316933	AND Enantiomer S N N N N (*) [*] [c] 1: [c] :n:[cH]:n:1	-0.423	0 out of 2



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.387
Enrichment: 1.32
Bayesian Score: 2.64
Mahalanobis Distance: 14.1

Mahalanobis Distance p-value: 4.16e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Lamotrigine	Danthron	Phenazopyridine
Structure	H ₂ N ^r r NH ₂	ОМОН	H ₂ N ^M NH ₂
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.561	0.572	0.574
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	-1441604640	AND Enantiomer S N NH H ₂ N [*]:[e]1:[*]C[C@@H](C)CC1	0.77	4 out of 5	

FCFP_6	1294344583	AND Enantiomer S N NH H ₂ N	0.517	2 out of 3
FCFP_6	-1043339860	[*]NN[c](:[*]):[*] AND Enantiomer S NH H ₂ N [*]CC(C)C[*]	0.383	24 out of 61
	Top Feat	ures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	AND Enantiomer S N NH H ₂ N [*]:n:[cH]:n:[*]	-0.731	1 out of 12
FCFP_6	-475316933	AND Enantiomer S NH H ₂ N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	-0.423	0 out of 2
FCFP_6	907007053	AND Enantiomer N N NH H ₂ N [*]C([*])C[c](:[*]):[*]	-0.366	11 out of 62

 $|C_{28}H_{28}N_6S_2|$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.286
Enrichment: 0.971
Bayesian Score: -0.759
Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 1.06e-008

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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

Name	Emetine	Fluticasone	Bromocriptine
Structure	on the state of th	HO to the first to	Br Man No
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.875	0.879	0.906
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

1. OPS PC19 out of range. Value: -3.2042. Training min, max, SD, explained variance: -2.8152, 4.6113, 1.185, 0.0147.

Feature Contribution

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	-1441604640	AND Enandomer N N N N N N N N N N N N N N N N N N N	0.77	4 out of 5	

FCFP_6	-773983804	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.409	10 out of 24
FCFP_6	-1043339860	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.383	24 out of 61
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.731	1 out of 12
FCFP_6	1293778554	AND Enantiomer N	-0.719	0 out of 4
FCFP_6	-475316933	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.423	0 out of 2

 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.181 Enrichment: 0.614 Bayesian Score: -6.29 Mahalanobis Distance: 16

Mahalanobis Distance p-value: 1.07e-009

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Mycophenolate	Nicardipine	Nimodipine
Structure	HO	HANDO ON THE PROPERTY OF THE P	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.615	0.655	0.668
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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: 902193919: [*]:[c](:[*])C#C

Feature Contribution					
	Top fe	atures for positive c	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	131784192	[*]C#C	0.983	8 out of 9	

FCFP_6	-773983804	N. ~ 0	0.409	10 out of 24
		[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1		
FCFP_6	-771557733	[*]N[c]1:[cH]:[*]:[cH]:[*]	0.38	2 out of 4
	Top Feat	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-124685461	[*]:n:[cH]:n:[*]	-0.731	1 out of 12
FCFP_6	1674955425	[*]O[c]1:[cH]:[c](:[c](:[r]):[*]:[c]:1[*]	-0.719	0 out of 4
FCFP_6	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4

AND Enantiomer S N NH CI

 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.146
Enrichment: 0.486
Bayesian Score: -8.59
Mahalanobis Distance: 15.7

Mahalanobis Distance p-value: 7.7e-007

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Mestranol	Levonorgestrel
Structure	OH	HO WO	O CH
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.638	0.659	0.663
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

- 1. OPS PC7 out of range. Value: -5.152. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
- 2. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution					
	Top fea	tures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
FCFP_12	-1441604640	AND Enantiomer S N N N N (N) (N) (P):[c]1:[*]C[C@@H](C (C) (C)	0.75	4 out of 4	

FCFP_12	-2002183168	AND Enantiomer N	0.616	4 out of 5
FCFP_12	-124685461	AND Enantomer N N N H CI [*]:n:[cH]:n:[*]	0.4	1 out of 1
	Top Fea	tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1151884458	AND Ensentioner S N N N N N N N N (c) (in:[*]):[c](: [*]):[*]	-1.11	0 out of 6
FCFP_12	590925877	AND Enantioner N N N H H CI [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	-1508180856	AND Enantiomer S N N N H CI [*][c]1:[cH]:[cH]:[c] (CI):[cH]:[cH]:1	-0.859	0 out of 4

AND Enantiomer S N N N N F

 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.152
Enrichment: 0.505
Bayesian Score: -6.97
Mahalanobis Distance: 18.1

Mahalanobis Distance p-value: 9.82e-009

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Nafenopin	Levonorgestrel	Mestranol		
Structure	OH	OH CHAPTER SHAPE OF THE SHAPE O	HO THO		
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen		
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen		
Distance	0.627	0.656	0.658		
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

- 1. OPS PC7 out of range. Value: -4.7493. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
- 2. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution					
	Top fea	tures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
FCFP_12	-1441604640	AND Enantiomer S N N H F [*]:[c]1:[*]C[C@@H](C)CC1	0.75	4 out of 4	

FCFP_12	-2002183168	AND Enantioner N	0.616	4 out of 5
FCFP_12	-124685461	AND Enantiomer NH NH [*]:n:[cH]:n:[*]	0.4	1 out of 1
	Top Feat	tures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1151884458	AND Equationer S N N H F [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6
FCFP_12	590925877	AND Enantiomer N N H F [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	-773983804	AND Enantiomer S N=N H F [*]N[c]1:[cH]:[c] ([*]):[cH]:[cH]:1	-0.789	1 out of 10

 $C_{17}H_{18}N_{4}S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.163
Enrichment: 0.54
Bayesian Score: -5.43
Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 0.000159

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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

	<u> </u>		
Name	Phenolphthalein	Oxazepam	Danthron
Structure	НО	CI TO NOT NOT NOT NOT NOT NOT NOT NOT NOT	O HOH
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.573	0.658	0.721
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

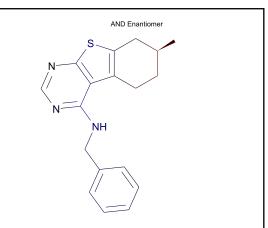
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
FCFP_12	-1441604640	AND Enantiomer N N N N H ₂ N [*]:[c]1:[*]C[C@@H](C)CC1	0.75	4 out of 4		

FCFP_12	-2002183168	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.616	4 out of 5
FCFP_12	-124685461	AND Enantomer S N N H ₂ N [*]:n:[cH]:n:[*]	0.4	1 out of 1
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	1069584379	AND Enantiomer S N N N H H 2N [*]:[c](:[*])N	-1.11	0 out of 6
FCFP_12	-1151884458	AND Enantiomer S N H ₂ N [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6
FCFP_12	590925877	AND Enantiomer N N N H H 2N [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13



 $C_{18}H_{19}N_3S$

Molecular Weight: 309.42856

ALogP: 4.678
Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.169
Enrichment: 0.563
Bayesian Score: -4.74
Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.000442

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Nafenopin	Oxaprocin	Suprofen		
Structure	OH	HO HO	OH OH		
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen		
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen		
Distance	0.566	0.622	0.665		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997		

Model Applicability

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

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

Feature Contribution					
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
FCFP_12	-1441604640	AND Enantiomer N N H (*):[c]1:[*]C[C@@H](C)CC1	0.75	4 out of 4	

FCFP_12	-2002183168	AND Enantiomer N	0.616	4 out of 5
FCFP_12	427906732	AND Enantiomer N	0.4	1 out of 1
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	1294255210	AND Enantiomer S NH NH [*]CN[c](:[*]):[*]	-1.63	0 out of 12
FCFP_12	-1151884458	AND Enantioner N	-1.11	0 out of 6
FCFP_12	991735244	AND Enantiomer N H H (*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.562	5 out of 28

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Eval.& Res./Off. Testing &

Res.) Sept. 1997

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.249
Enrichment: 0.828
Bayesian Score: -0.38
Mahalanobis Distance: 13.8

Mahalanobis Distance p-value: 2.12e-005

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

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

cutoff.

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds						
Name	Nithiazide	Sulfamethazine	Phenobarbital			
Structure	O NH	HN LT	HN O			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen			
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen			
Distance	0.757	0.771	0.799			
Reference	US FDA (Centre for Drug	US FDA (Centre for Drug	US FDA (Centre for Drug			

Model Applicability

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

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

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
FCFP_12	-1441604640	AND Enantiomer N N N N N H H N NH ₂ S [*]:[c]1:[*]C[C@@H](C)CC1	0.75	4 out of 4		

FCFP_12	-885461129	AND Enantiomer S NH NH NH NH ₂	0.683	3 out of 3
FCFP_12	-2002183168	[*]NNC(=[*])[*] AND Enantiomer NNN NH HN NH S NH2 S [*][C@@H]1[*][c]2:[*]	0.616	4 out of 5
	Top Feat	cures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1151884458	AND Enantiomer S N N N N N N N N N N N N	-1.11	0 out of 6
FCFP_12	17	AND Enantiomer S NH HN NH S NH S [*]:S:[*]	-0.53	5 out of 27
FCFP_12	-1564473960	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.519	0 out of 2

C₁₇H₁₈N₄O₂S₂ Molecular Weight: 374.48042

ALogP: 3.541
Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Prediction: Single-Carcinogen

Probability: 0.146
Enrichment: 0.484
Bayesian Score: -9.03
Mahalanobis Distance: 14

Mahalanobis Distance p-value: 1.58e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds						
Name	Phenolphthalein	Bicalutamide	Sulfamethazine			
Structure	НО	HO H	HN MH 2			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen			
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen			
Distance	0.712	0.715	0.719			
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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
FCFP_12	-1441604640	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.75	4 out of 4		

FCFP_12	-2002183168	AND Enantiomer N S N N N N N N N N N N N N N N N N N	0.616	4 out of 5
FCFP_12	-124685461	AND Enantiomer S NH S NH S NH () (*) (*) (*) (*) (*) (*)	0.4	1 out of 1
	Top Feat	tures for negative		1
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1151884458	AND Enantiomer S NH ₂ [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6
FCFP_12	590925877	AND Enantiomer S N N N N N N N N N N N N N N N N N N	-0.998	1 out of 13
FCFP_12	-1096219292	AND Enantiomer N N N N N N N N N N N N N	-0.859	0 out of 4

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

AND Enantiomer	
N NH	
$0=S$ H_2N 0	

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.146 Enrichment: 0.485 Bayesian Score: -8.75 Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.00077

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

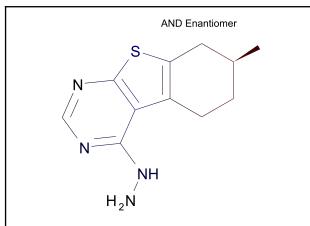
Structural Similar Compounds						
Structure	Sulfamethazine No. 10 to 10 t	Bicalutamide	Phenolphthalein			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen			
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen			
Distance	0.711	0.716	0.723			
Reference	US 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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1441604640	AND Enantiomer N N N N H H 2N C[*]:[c]1:[*]C[C@@H](C)CC1	0.75	4 out of 4

FCFP_12	-2002183168	AND Enantiomer N N N N N H N ("](C@@H)1["](c]2:["] :["](c](("]))(c):2 CC1	0.616	4 out of 5		
FCFP_12	-124685461	AND Enantioner S N N H H Parity (*) (*) (*) (*) (*)	0.4	1 out of 1		
	Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
FCFP_12	-1151884458	AND Enantiomer S N N N N N (*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6		
FCFP_12	590925877	AND Enantiomer N N H N H N (N N H N (N N H N (N N N H N (N N N H N (N N N H N N N N	-0.998	1 out of 13		
FCFP_12	-1096219292	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.859	0 out of 4		



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561
Rotatable Bonds: 1
Acceptors: 4

Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.201
Enrichment: 0.666
Bayesian Score: -2.53
Mahalanobis Distance: 15.1

Mahalanobis Distance p-value: 2.29e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Danthron	Hydralazine	Phenobarbital
Structure	O MOH	N N NH 2 H	HN O
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.576	0.636	0.643
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 Co	ntribution			
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1441604640	AND Enantiomer S N NH H ₂ N [*]:[c]1:[*]C[C@@H](C)CC1	0.75	4 out of 4

FCFP_12	-2002183168	AND Enantiomer S N H ₂ NH ["][C@@H]1["][c]2:["] :["]:[c]:["]:[c]:2 CC1	0.616	4 out of 5			
FCFP_12	-124685461	AND Enantiomer S NH H ₂ N [*]:n:[cH]:n:[*]	0.4	1 out of 1			
	Top Features for negative contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set			
FCFP_12	-1151884458	AND Enantlomer S NH H ₂ N [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6			
FCFP_12	17	AND Enantiomer S NH H ₂ N [*]:s:[*]	-0.53	5 out of 27			
FCFP_12	-1564473960	AND Enantomer NH H ₂ N' [*]:n:[c]1:s:[*]: [c]:1:[*]	-0.519	0 out of 2			

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6 Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.159 Enrichment: 0.527 Bayesian Score: -5.92 Mahalanobis Distance: 14.5

Mahalanobis Distance p-value: 5.83e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediction. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Glimepride	Simvastatin	Bicalutamide
Structure	NH NH		HO AND THE STATE OF THE STATE O

-	-		-
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.924	0.936	0.939
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

Multiple-Carcinogen

Single-Carcinogen

Model Applicability

Actual Endpoint

Structural Similar Compounds

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

- OPS PC7 out of range. Value: -4.4901. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
- Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Single-Carcinogen

Feature Contribution Top features for positive contribution **Fingerprint Bit/Smiles** Feature Structure Score Multiple-Carcinogen in training set FCFP 12 1441604640 0.75 4 out of 4

FCFP_12	-2002183168	AND Enantloner N S N N N N N N N N N N N N N N N N N	0.616	4 out of 5
FCFP_12	-124685461	AND Enantiomer NN N N N N N N N N N N N N N N N N N	0.4	1 out of 1
		tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1151884458	AND Enantiomer N N NH S N NH [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6
FCFP_12	590925877	AND Enantiomer N S N N NH S [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	-773983804	AND Enantiomer N S	-0.789	1 out of 10

AND Enantiomer S N NH NH CI

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Moderate_Severe

Probability: 0.833
Enrichment: 1.21
Bayesian Score: 0.771
Mahalanobis Distance: 8.73

Mahalanobis Distance p-value: 0.657

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	
Structure	H H	CI	O N N N N O O O O O O O O O O O O O O O	
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild	
Predicted Endpoint	Mild	Moderate_Severe	Mild	
Distance	0.582	0.596	0.597	
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP 2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

FCFP_10	-1508180856	AND Enantiomer N N H CI [*][c]1:[cH]:[cH]:[c] (CI):[cH]:[cH]:1	0.329	16 out of 17
FCFP_10	-745491832	AND Enantiomer AND Enantiomer N H CI[c]1:[cH]:[cH]:[*]: [cH]:[cH]:1	0.304	29 out of 32
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1293778554	AND Enantiomer N N H CI [*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19
FCFP_10	-773983804	AND Enantiomer N N N H CI [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.294	50 out of 102
FCFP_10	307419094	AND Enantiomer S	-0.29	21 out of 43

AND Enantiomer S N NH F

C₁₇H₁₆FN₃S

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Moderate_Severe

Probability: 0.826 Enrichment: 1.2

Bayesian Score: 0.219 Mahalanobis Distance: 8.63

Mahalanobis Distance p-value: 0.708

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	N;S-DIBENZOYL-O- AMINOTHIOPHENOL	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	
Structure	T Z H	H N N N O	CI	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance	0.566	0.573	0.600	
Reference	28ZPAK-;175;72	28ZPAK 89;72	CIGET* -;-;77	

Model Applicability

Unknown features are fingerprint features in 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: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1441604640	AND Enantiomer S N N H F [*]:[c]1:[*]C[C@@H](C)CC1	0.344	6 out of 6

FCFP_10	-1539132615	AND Enantiomer N N N N F [*]C[c]1:s:[*]:[c]	0.224	11 out of 13
FCFP_10	-2002183168]:1[*] AND Enantiomer S N N N H F [*](C@@H]1[*](c]2:[*] :[*]:[c](:[*]).[c]:2 CC1	0.207	23 out of 28
	Top Feat	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1293778554	AND Enantiomer N N N N N N N N N	-0.304	9 out of 19
FCFP_10	-773983804	AND Enantomer N	-0.294	50 out of 102
FCFP_10	307419094	AND Enantiomer N N N H F [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.29	21 out of 43

N NH NH

C₁₇H₁₇N₃OS

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Mild Probability: 0.756 Enrichment: 1.1

Bayesian Score: -2.35

Mahalanobis Distance: 7.33

Mahalanobis Distance p-value: 0.992

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	N;S-DIBENZOYL-O- AMINOTHIOPHENOL	1-BENZOYLAMINO-4- METHOXY-5- CHLORANTHRAQUINONE	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	
Structure	N I N I N I N I N I N I N I N I N I N I	HN AND THE O	H N N O	
Actual Endpoint	Mild	Mild	Mild	
Predicted Endpoint	Mild	Mild	Mild	
Distance	0.574	0.609	0.622	
Reference	28ZPAK-;175;72	28ZPAK-;90;72	28ZPAK 89;72	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1539132615	[*]C[c]1:s:[*]:[*]:[c]:1[*]	0.224	11 out of 13

FCFP_10	-2002183168	N N N N N N N N N N	0.207	23 out of 28
FCFP_10	-1438301350	[*]:[c]1:[*]CCCC1	0.204	14 out of 17
	Top Feat	tures for negative of	contribution	า
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1977641857	[*]:[cH]:[c](OC):[cH] :[*]	-0.78	4 out of 15
FCFP_10				
. 6.10	-1909820884	[*]:[c]1:[*]:[c]2 CCCC[c]:1:2	-0.507	0 out of 1

[*][c]1:[cH]:[cH]:[c] (OC):[cH]:[cH]:1

 $C_{17}H_{18}N_{4}S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Mild Probability: 0.805 Enrichment: 1.17

Bayesian Score: -0.828 Mahalanobis Distance: 7.81

Mahalanobis Distance p-value: 0.956

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

	•		
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	1-AMINO-4-HYDROXY-5- CHLORANTHRAQUINONE
Structure	HO 11 NH 2	HN MH 2	HO M NH 2
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.596	0.609	0.654
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;83;72

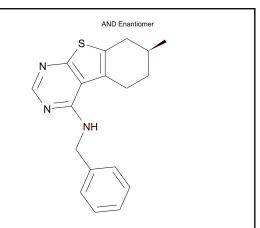
Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-1441604640	AND Enantiomer N N N H ₂ N [*]:[c]1:[*]C[C@@H](C)CC1	0.344	6 out of 6	
	•	•	•	•	

FCFP_10	-1539132615	AND Enantiomer N N H H 2 [*]C[c]1:s:[*]:[*]:[c]:1[*]	0.224	11 out of 13
FCFP_10	-2002183168	AND Enantlomer N N H H [*][C@@H]1[*][c]2:[*] :[*]:[c](:[*]):[c]:2 CC1	0.207	23 out of 28
	Top Fea	tures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-200702388	AND Enantiomer N N H, N [*]:[c](:[*])N[c]1:[c H]:[cH]:[c](N):[cH]: [cH]:1	-0.361	2 out of 5
FCFP_10	1293778554	AND Enantiomer N=N=N+ H ₂ N [*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19
FCFP_10	-773983804	AND Enantiomer N=N+1 H ₂ N [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.294	50 out of 102



 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Moderate_Severe

Probability: 0.833 Enrichment: 1.21 Bayesian Score: 0.713 Mahalanobis Distance: 7.54

Mahalanobis Distance p-value: 0.982

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sillinar Compounds					
Name	N;S-DIBENZOYL-O- AMINOTHIOPHENOL	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER		
Structure		O N	CI		

	ö		, , , , , , , , , , , , , , , , , , ,
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.529	0.568	0.591
Reference	28ZPAK-;175;72	28ZPAK 89;72	CIGET* -;-;77

Model Applicability

Structural Similar Compounds

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe
				in training set
FCFP_10	-1441604640	AND Enantiomer N N N N N (*):[c]1:[*]C[C@@H](C)CC1	0.344	6 out of 6

FCFP_10	907096426	AND Enantiomer S N N N N N N N (*]NC[c](:[*]):[*]	0.332	5 out of 5
FCFP_10	427906732	AND Enantioner N N N N N N N N N N N N N N N N N N	0.294	3 out of 3
	Top Fea	tures for negative o	ontributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	307419094	AND Enantiomer N	-0.29	21 out of 43
FCFP_10	-1698724694	AND Enantiomer N	-0.284	53 out of 107
FCFP_10	1294255210	AND Enantiomer NH NH [*]CN[c](:[*]):[*]	-0.218	20 out of 38

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Mild Probability: 0.813 Enrichment: 1.18 Bayesian Score: -0.44 Mahalanobis Distance: 7.72

Mahalanobis Distance p-value: 0.967

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	1;2;4-TRIHYDROXY ANTHRAQUINONE	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	ANILINE;2;2'DITHIOBIS-	
Structure	O OH WOH	OH OH NH	H ₂ N _H	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance	0.678	0.685	0.740	
Reference	28ZPAK-;103;7	28ZPAK 190;72	28ZPAK-;172;72	

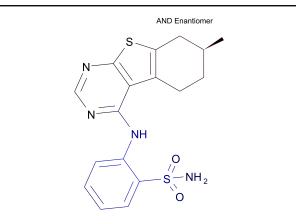
Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP 2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

Feature Co	Feature Contribution				
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-1441604640	AND Enantiomer S N N N N H H H N NH ₂ S [*]:[c]1:[*]C[C@@H](C)CC1	0.344	6 out of 6	

FCFP_10	1070061035	AND Enantiomer S NH HN NH S NH ₂ S [*]C(=[*])N	0.239	284 out of 338
FCFP_10	-1539132615	AND Enantioner N N N N N N N N N N N N N N N N N N	0.224	11 out of 13
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1294344583	AND Enantomer S NH HN S NH ₂ [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	1499521844	AND Enantiomer NH NH HN NH ₂ [*]NC(=S)N	-0.4	1 out of 3
FCFP_10	307419094	AND Enantiomer N N N N N H H N N H N	-0.29	21 out of 43



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild Probability: 0.706 Enrichment: 1.02 Bayesian Score: -3.42

Mahalanobis Distance: 7.33

Mahalanobis Distance p-value: 0.992

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	ANILINE;2;2'DITHIOBIS-
Structure	HO to NH 2	HN _{Mt} NH ₂	H ₂ N _H

Model Applicability

Actual Endpoint

Distance

Reference

Predicted Endpoint

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

Mild

Mild

0.651

28ZPAK-;124;72

Moderate_Severe

Moderate Severe

28ZPAK-;172;72

0.702

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Mild

Mild

0.611

28ZPAK 239;72

4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1441604640	AND Enantiomer S N N N N N N N N N N N N	0.344	6 out of 6

FCFP_10	1070061035	AND Enantioner S NH O NH O S NH O O (T)	0.239	284 out of 338
FCFP_10	-1539132615	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.224	11 out of 13
	Top Feat	tures for negative of	contribution	ı
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-377913290	AND Enantiomer S N N N N H H NH ₂ [*]N[c]1:[cH]:[*]:[cH]:[cH]:[c]:1S(=O)(=O)N	-1.03	1 out of 7
FCFP_10	-103236997	AND Enantionner N N N N N N N N N N (NH ₂ [*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]:[cH]: [c]:1N[c](:[*]):[*]	-0.6	1 out of 4
FCFP_10	-605671248	AND Enantiomer S N=N N+V NH ₂ [*]S(=[*])(=[*])[c]1: [cH]:[*]:[cH]:[cH]:[c]:1N[c](:[*]):[*]	-0.341	3 out of 7

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541
Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Mild Probability: 0.792 Enrichment: 1.15 Bayesian Score: -1.29

Mahalanobis Distance: 7.33

Mahalanobis Distance p-value: 0.992
Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false

negative rate

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	2- NAPHTHALENESULFONI C ACID;5-AMINO-6- ETHOXY-	
Structure	HO MANU NA	HN nt NH 2	HO NH ₂	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Mild	
Distance	0.620	0.660	0.700	
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;191;72	

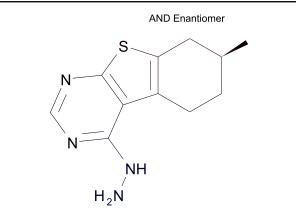
Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Feature Contribution Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1441604640	AND Enantiomer S N N H H 2N [*]:[c]1:[*]C[C@@H](C)CC1	0.344	6 out of 6

FCFP_10	1070061035	AND Enantiomer N N N N N N N N N N N N N	0.239	284 out of 338
FCFP_10	-1539132615	AND Enantiomer N H N (*] C [c] 1:s:[*]:[c]]:1[*]	0.224	11 out of 13
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-174657146	[*]S(=[*])(**[*])[c]1: [cH]:[cH]:[cH]:[cH]: 1	-0.4	1 out of 3
FCFP_10	1293778554	AND Enantiomer N N N N N N N N N N N N N	-0.304	9 out of 19
FCFP_10	-773983804	AND Enantiomer N S N H N (*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.294	50 out of 102



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561
Rotatable Bonds: 1
Acceptors: 4

Donors: 2

Model Prediction

Prediction: Mild Probability: 0.81 Enrichment: 1.18

Bayesian Score: -0.572

Mahalanobis Distance: 7.88

Mahalanobis Distance p-value: 0.946

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds Name PHENOL; 4;4'- SUI FONYI DI- ACID: 2-HYDROXY- FURIL; DIOXIME					
	SULFONYLDI-	ACID;2-HYDROXY-			
Structure	HO OH	O OH	O — N N HO		
Actual Endpoint	Moderate_Severe	Moderate_Severe	Moderate_Severe		
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Moderate_Severe		
Distance	0.542	0.551	0.554		
Reference	BIOFX* 601-05501;74	28ZPAK-;186;72	28ZPAK		

Model Applicability

Unknown features are fingerprint features in 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: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1441604640	AND Enantiomer S N NH H ₂ N [*]:[c]1:[*]C[C@@H](C)CC1	0.344	6 out of 6

FCFP_10	-1539132615	NH H ₂ N [*]C[c]1:s:[*]:[c]:1[*]	0.224	11 out of 13
FCFP_10	-2002183168	AND Enantomer N N N N N N N N N N N N N N N N N N N	0.207	23 out of 28
	Top Fea	tures for negative of	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1294344583	AND Enantiomer S NH H ₂ N [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	1070150408	AND Enantiomer S NH H ₂ N [*]NN	-0.507	0 out of 1
FCFP_10	307419094	AND Enantiomer S N N H ₂ NH [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.29	21 out of 43

 $|C_{28}H_{28}N_6S_2|$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Mild Probability: 0.808 Enrichment: 1.17

Bayesian Score: -0.679 Mahalanobis Distance: 9.84

Mahalanobis Distance p-value: 0.135

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	DINAPHTHO(1;2;3- CD:3';2';1'-IM)PERYLENE- 5;10-DIONE;16;17- DIHYDROXY	Anthraquinone; 1;1'- (anthraquinon-1;5- ylenediimino)di-	Anthraquinone; 1;1'- (anthraquinon-1;4- ylenediimino)di-	
Structure	ОН	H H H		
Actual Endpoint	Mild	Mild	Mild	
Predicted Endpoint	Mild	Mild	Mild	
Distance	0.767	0.772	0.774	
Reference	28ZPAK-;104;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

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

FCFP_10	-1441604640	AND Enantiomer N S 1 N N N N N N N N N N N N (*):[c]1:[*]C[C@@H](C)CC1	0.344	6 out of 6
FCFP_10	-1539132615	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.224	11 out of 13
FCFP_10	-2002183168	AND Enantiomer N S N N N N NH S [*][C@@H]1[*][c]2:[*] :[*]:[c](c[*]):[c]2 CC1	0.207	23 out of 28
	Ton Feat	ures for negative	contribution	1

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-200702388	AND Enantiomer S N N N N N N N N N N N N	-0.361	2 out of 5
FCFP_10	1293778554	AND Enantiomer N S N N NH S [*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19

FCFP_10	-773983804	AND Enantiomer	-0.294	50 out of 102
		N-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V-V		
		EN 💇		
		N NH S		
		[*]N[c]1:[cH]:[*]:[c]		
		([*]):[cH]:[cH]:1		

 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Mild Probability: 0.766 Enrichment: 1.11 Bayesian Score: -2.1

Mahalanobis Distance: 10.4

Mahalanobis Distance p-value: 0.0368

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Benzoic acid; p-(N-butyl- 2-(butylamino)acetamido)- ; butyl ester;	COLCHICINE	Benzoic acid; p-(N-butyl- 2-(piperidino)acetamido)-; butyl ester;		
Structure	H	HN 2tr	N N N N N N N N N N N N N N N N N N N		
Actual Endpoint	Moderate_Severe	Moderate_Severe	Moderate_Severe		
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Moderate_Severe		
Distance	0.648	0.707	0.750		
Reference	Arzneimittel-Forschung 8;609;58	AJOPAA 31;837;48	Arzneimittel-Forschung 8;609;58		

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 3. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 902193919: [*]:[c](:[*])C#C

FCFP_10	365650923	[*]OCCO[c](:[*]):[*]	0.386	17 out of 17
FCFP_10	-1716224640	[*][c]1:[]:[cH]:[c]2 °· :[c]([*]):[*]:[cH]:n :[c]:2:[cH]:1	0.294	3 out of 3
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1699003333	[*]N[c]1:[cH]:[cH]:[c H]:[c](:[cH]:1)C#[*]	-1.09	2 out of 12
FCFP_10	1679603620	[*]O[c]1:[cH]:[c](:n: [*]):[c](:[*]):[*]:[c]:1[*]	-0.507	0 out of 1
FCFP_10	341504799	[*]CCOC	-0.425	7 out of 17

TOPKAT_Ocular_Irritancy_Moderate_vs_Severe

AND Enantiomer S N NH CI

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Prediction: Moderate
Probability: 0.602
Enrichment: 0.971
Bayesian Score: -2.21
Mahalanobis Distance: 9.8

Mahalanobis Distance p-value: 0.0717

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	4;4'-DICHLORO-2- NITROBIPHENYL ETHER	
Structure	The state of the s	CI	CI NO CI	
Actual Endpoint	Moderate	Severe	Moderate	
Predicted Endpoint	Moderate	Severe	Moderate	
Distance	0.588	0.605	0.722	
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK-;84;72	

Model Applicability

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

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

Feature Contribution					
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set	
SCFP_12	-1272709286	AND Enantiomer N N N N H CI [*]C([*])C[c](:[*]):[*]	0.231	24 out of 31	

SCFP_12	341480432	AND Enantomer N N N N H CI [*]N[c]1:[cH]:[cH]:[c](CI):[cH]:[cH]:1	0.218	1 out of 1
SCFP_12	-1378360678	AND Enantiomer N N N N H (*) [*][c]1:[*]:[cH]:[c](CI):[cH]:[cH]:1	0.211	22 out of 29
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1310748454	AND Enantiomer N N N N N H CI [*][c]1:[*]:[c](: [*]):s:1	-1.04	0 out of 3
SCFP_12	951581613	AND Enantiomer S N N N N N N N N N N N N	-0.769	2 out of 9
SCFP_12	1334669481	AND Enantromer S N N H CI [*]N[c](:[cH]:[*]):[c H]:[*]	-0.685	28 out of 93

AND Enantiomer S N N NH

 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Moderate
Probability: 0.452
Enrichment: 0.729
Bayesian Score: -4.89
Mahalanobis Distance: 9.1

Mahalanobis Distance p-value: 0.287

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	ACETIC ACID; 2-(- CHLOROMETHY-1- NAPHTHYLTHIO)-	
Structure	The state of the s	CI	OH OH	
Actual Endpoint	Moderate	Severe	Severe	
Predicted Endpoint	Moderate	Severe	Severe	
Distance	0.621	0.626	0.679	
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK-;173;72	

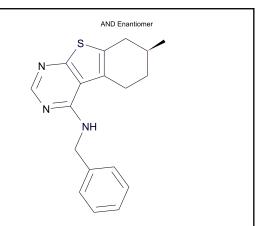
Model Applicability

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

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

Feature Contribution					
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set	
SCFP_12	-1272709286	AND Enantioner N	0.231	24 out of 31	

SCFP_12	-1849867720	AND Enantiomer N N N N H F [*][C@@H]1[*][c]2:[*] :[*]:[c](:[*)):[c]2 CC1	0.168	16 out of 22
SCFP_12	-1272798659	AND Enantiomer S N NH F [*]CCC([*])[*]	0.139	216 out of 309
	Top Featu	res for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1310748454	AND Enantiomer N H F [*][c]1:[*]:[*]:[c](: [*]):s:1	-1.04	0 out of 3
SCFP_12	951581613	AND Enantiomer S N H F [*]:[c](:[*])N[c](:[*]):[*]	-0.769	2 out of 9
SCFP_12	1334669481	AND Enantiomer N H F [*]N[c](:[cH]:[*]):[c H]:[*]	-0.685	28 out of 93



C₁₈H₁₉N₃S

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Moderate
Probability: 0.641
Enrichment: 1.03
Bayesian Score: -1.34

Mahalanobis Distance: 8.92

Mahalanobis Distance p-value: 0.375

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	ACETIC ACID; 2-(- CHLOROMETHY-1- NAPHTHYLTHIO)-	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	
Structure	CI	OH CI	O H N N N N N N N N N N N N N N N N N N	
Actual Endpoint	Severe	Severe	Moderate	
Predicted Endpoint	Severe	Severe	Moderate	
Distance	0.598	0.635	0.655	
Reference	CIGET* -;-;77	28ZPAK-;173;72	28ZPAK-;242;72	

Model Applicability

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

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

Feature Contribution						
	Top fea	tures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set		
SCFP_12	-1272709286	AND Enantiomer S	0.231	24 out of 31		

SCFP_12	-1849867720	AND Enantiomer S N N (H (*)[C@@H]1[*][c]2:[*] :[*]:[c]:[c]:[*]:[c]:2 CC1	0.168	16 out of 22
SCFP_12	-1272798659	AND Enantiomer S NH NH [*]CCC([*])[*]	0.139	216 out of 309
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1310748454	AND Enantiomer N N N N N H N (H H) (E) (E) (E) (E) (E) (E) (E) (E) (E) (-1.04	0 out of 3
SCFP_12	18117904	AND Exantiomer S NH NH [*]CN[c](:[*]):[*]	-0.481	3 out of 9
SCFP_12	10	AND Enantioner N N N N N N N N N N N N N N N N N N	-0.451	43 out of 112

28ZPAK-;175;72

AND Enantiomer S N NH CI

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.98

Mahalanobis Distance: 7.8

Mahalanobis Distance p-value: 0.958

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	N;S-DIBENZOYL-O- AMINOTHIOPHENOL
Structure	The second secon	O N N N N N N N N N N N N N N N N N N N	O O O O O O O O O O O O O O O O O O O
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.584	0.584	0.585

Model Applicability

Feature Contribution

Reference

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

28ZPAK 89;72

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

28ZPAK-;242;72

- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

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

[*][c]1:[*]:[*]:[c](:

[*]):s:1

FCFP_12	-1508180856	AND Enantiomer S N N N N (I) (I) (CI):[cH]:[cH]:[c] (CI):[cH]:[cH]:1	0.2	17 out of 17
FCFP_12	-1539132615	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.197	13 out of 13
	Top Feat	ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	0	AND Enantiomer S N N H CI [*]C[*]	0	1184 out of 1397
FCFP_12	-1272798659	AND Enantiomer S N N H CI [*]CCC([*])[*]	0	517 out of 643
FCFP_12	-1043339860	AND Enantioner N N N H CI [*]CC(C)C[*]	0	146 out of 186

Irritant

0.595

CIGET* -;-;77

AND Enantiomer S N NH F

C₁₇H₁₆FN₃S

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18

Bayesian Score: 1.75

Mahalanobis Distance: 7.67

Mahalanobis Distance p-value: 0.972

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	N;S-DIBENZOYL-O- AMINOTHIOPHENOL	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER
Structure	O O O O O O O O O O O O O O O O O O O	O NAME O O O O O O O O O O O O O O O O O O O	CION
Actual Endpoint	Irritant	Irritant	Irritant

Model Applicability

Predicted Endpoint

Distance

Reference

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

Irritant

0.566

28ZPAK 89;72

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Irritant

0.557

28ZPAK-:175:72

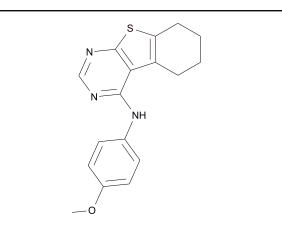
4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	AND Examtomer N N F [*][c]1:[*]:[c](: [*]):S:1	0.208	44 out of 44

FCFP_12	-1539132615	AND Enantiomer N N N N F [*]C[c]1:s:[*]:[c]]:1[*]	0.197	13 out of 13
FCFP_12	17	AND Enantiomer NH NH [*]:S:[*]	0.189	48 out of 49
	Top Feat	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	0	AND Enantiomer S N NH F [*]C[*]	0	1184 out of 1397
FCFP_12	-1272798659	AND Enantiomer S N NH F [*]CCC([[*])[*]	0	517 out of 643
FCFP_12	-1043339860	AND Enantiomer N N N N N N F [*]CC(C)C[*]	0	146 out of 186

0.618

28ZPAK 89;72



 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18

Bayesian Score: 0.616
Mahalanobis Distance: 7.11

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	N;S-DIBENZOYL-O- AMINOTHIOPHENOL	1-BENZOYLAMINO-4- METHOXY-5- CHLORANTHRAQUINONE	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-		
Structure	O T T T T T T T T T T T T T T T T T T T	HN MU	O N N N C I		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Irritant	Irritant	Irritant		

Model Applicability

Distance

Reference

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

0.604

28ZPAK-:90:72

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

0.572

28ZPAK-:175:72

4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	1747237384	[*][c]1:[*]:[c](: [*]):s:1	0.208	44 out of 44	

FCFP_12	-1539132615	[*]C[c]1:s:[*]:[c]:1[*]	0.197	13 out of 13
FCFP_12	17	N N N N N N N N N N N N N N N N N N N	0.189	48 out of 49
	-	tures for negative		1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1060187936	[*]N[c]1:[cH]:[cH]:[c][(OC):[cH]:[cH]:1	-0.344	2 out of 4
FCFP_12	414792462	[*]O[c]1:[cH]:[cH]:[c](N[c](:[*]):[*]):[c H]:[cH]:1	-0.268	1 out of 2
FCFP_12	-1909820884	[*]:[c]1:[*]:[c]2 CCCC[c]:1:2	-0.268	1 out of 2

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.13
Mahalanobis Distance: 7.73

Mahalanobis Distance p-value: 0.966

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	Anthraquinone; 1-amino- 2-bromo-4-hydroxy-	
Structure	HO stands of the NH 2	O HN MH 2	HO MANH 2	
Actual Endpoint	Irritant	Irritant	Non-Irritant	
Predicted Endpoint	Irritant	Irritant	Non-Irritant	
Distance	0.590	0.595	0.618	
Reference	28ZPAK 239;72	28ZPAK-;124;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 535;86	

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

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

		[*][c]1:[*]:[c](: [*]):s:1		
FCFP_12	-1539132615	AND Enantiomer N N H ₂ H ₂ [*]C[c]1:s:[*]:[*]:[c]:1[*]	0.197	13 out of 13
FCFP_12	17	AND Enantiomer N N H H S (*):S:[*]	0.189	48 out of 49

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	880832703	AND Enantiomer S N H 2N [*][c]1:[cH]:[cH]:[c] (N):[cH]:[cH]:1	-0.0697	15 out of 20
FCFP_12	-1272798659	AND Enantiomer S N N H ₂ N [*]CCC([*])[*]	0	517 out of 643

FCFP_12	-1043339860	AND Enantiomer	þ	146 out of 186
		, N=(N		
		H₂N		
		[*]CC(C)C[*]		

AND Enantiomer S N NH

 $|C_{18}H_{19}N_3S$

Molecular Weight: 309.42856

ALogP: 4.678
Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18

Bayesian Score: 1.35

Mahalanobis Distance: 7.49

Mahalanobis Distance p-value: 0.986

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly linaccurate.

Structural Similar Compounds

	<u> </u>		
Name	N;S-DIBENZOYL-O- AMINOTHIOPHENOL	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER
Structure	N H N H	O N N N N N N N N N N N N N N N N N N N	CI
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.526	0.564	0.590
Reference	28ZPAK-;175;72	28ZPAK 89;72	CIGET* -;-;77

Model Applicability

Unknown features are fingerprint features in 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: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

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

FCFP_12	-1539132615	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.197	13 out of 13
FCFP_12	17	AND Enantioner N NH NH [*]:S:[*]	0.189	48 out of 49
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	AND Enantiomer N= S N H H (H) (F)C[c]1:[cH]:[cH]:[c H]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	-1272798659	AND Enantomer S NH NH [*]CCC([*])[*]	0	517 out of 643
FCFP_12	-1043339860	AND Enantomer NH NH [*]CC(C)C[*]	0	146 out of 186

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4 Donors: 3

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18

Bayesian Score: 1.04

Mahalanobis Distance: 7.39

Mahalanobis Distance p-value: 0.99

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	1;2;4-TRIHYDROXY ANTHRAQUINONE	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	1-Naphthalenesulfonic acid; 4-amino-5-hydroxy-	
Structure	O OH O OH	OH NH OH	HO S NH 2	
Actual Endpoint	Irritant	Irritant	Non-Irritant	
Predicted Endpoint	Irritant	Irritant	Non-Irritant	
Distance	0.664	0.669	0.695	
Reference	28ZPAK-;103;7	28ZPAK 190;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1057;86	

Model Applicability

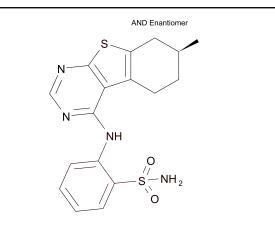
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

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

FCFP_12	1747237384	AND Enantlomer N N N N N N H H N NH ₂ S [*][c]1:[*]:[c](: [*]):S:1	0.208	44 out of 44
FCFP_12	-1539132615	AND Enantomer N N N N N N N N N H N N N 1 N N 1 N N 1 N 1	0.197	13 out of 13
FCFP_12	17	AND Enantiomer S NH NH HN S NH ₂ [*]:S:[*]	0.189	48 out of 49

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1872154524	AND Enantiomer S NH NH HN NH ₂ [*]C(=S)[*]	0	563 out of 690
FCFP_12	0	AND Enantiomer S NH HN NH S [*]C[*]	O	1184 out of 1397

FCFP_12	1	AND Enantiomer	0	872 out of 1051
		N=\ NH		
		HN NH		
		S S		
		[*]O[*]		



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.68

Mahalanobis Distance: 7.29 Mahalanobis Distance p-value: 0.993

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	Anthraquinone; 1-amino- 2-bromo-4-hydroxy-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE
Structure	HO the MANUAL MA	HO MANH 2	HN MAN NH 2
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.612	0.646	0.649
Reference	28ZPAK 239;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 535;86	28ZPAK-;124;72

Model Applicability

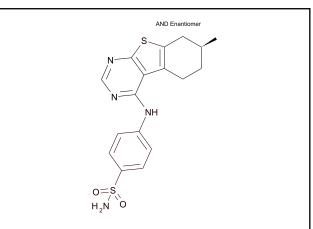
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

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

FCFP_12	1747237384	AND Enantiomer N N N N N N N N N N N N N 1 (*) (*) (*) (*) (*) (*) (*) (*)	0.208	44 out of 44
FCFP_12	-1539132615	AND Enantiomer N N N N N N N N N N N N N N 1 N N N 1 N N N 1 N N N 1 N N N 1 N N N 1 N N N N 1 N N N N 1 N	0.197	13 out of 13
FCFP_12	17	AND Enantiomer NH S-NH ₂ [*]:s:[*]	0.189	48 out of 49

Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	AND Enantiomer S N N NH ₂ [*]C[c]1:[cH]:[cH]:[c H]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	1	AND Enantiomer S NH O S NH ₂ O [*]O[*]	0	872 out of 1051

FCFP_12	-1272798659	AND Enantiomer	0	517 out of 643
		NH NH S-NH ₂		
		[*]CCC([*])[*]		



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1 Enrichment: 1.18 Bayesian Score: 1.9

Mahalanobis Distance: 7.29

Mahalanobis Distance p-value: 0.993

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	Anthraquinone; 1-amino- 2-bromo-4-hydroxy-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE		
Structure	HO MANH 2	HO MINH 2	O HN MAN TO THE NH 2		
Actual Endpoint	Irritant	Non-Irritant	Irritant		
Predicted Endpoint	Irritant	Non-Irritant	Irritant		
Distance	0.621	0.657	0.659		
Reference	28ZPAK 239;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 535;86	28ZPAK-;124;72		

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

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

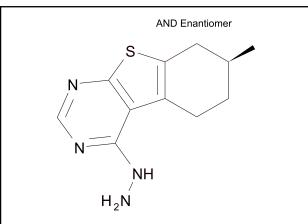
FCFP_12	1747237384	AND Enantioner N N H; N (*)[c]1:[*]:[c](: [*]):S:1	0.208	44 out of 44
FCFP_12	-1539132615	AND Enantiomer N N N H N N (*)C[c]1:s:[*]:[*]:[c]:1[*]	0.197	13 out of 13
FCFP_12	17	AND Enantiomer S N N H O'S H ₂ N [*]:S:[*]	0.189	48 out of 49

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-453677277	AND Enantiomer S N N N N N N N (*)C[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	0	264 out of 323
FCFP_12	-773983804	AND Enantomer N N N N N N N N N N N N N N N N N N	0	102 out of 121

FCFP_12	0	AND Enantiomer	0	1184 out of 1397
		N=(N H		
		O' 9 H ₂ N		
		[*]C[*]		

0.547

28ZPAK



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561
Rotatable Bonds: 1
Acceptors: 4

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1 Enrichment: 1.18 Bayesian Score: 1

Mahalanobis Distance: 7.79

Mahalanobis Distance p-value: 0.96

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	PHENOL; 4;4'- SULFONYLDI-	NAPHTHENESULFONIC ACID;2-HYDROXY-	FURIL;DIOXIME		
Structure	HO OF SOME	O OH	O — N HO		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Irritant	Irritant	Irritant		

Model Applicability

Distance

Reference

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

0.542

28ZPAK-;186;72

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

BIOFX* 601-05501:74

- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

0.540

4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	1747237384	AND Enantiomer N NH H ₂ N [*][c]1:[*]:[c](: [*]):s:1	0.208	44 out of 44	

FCFP_12	-1539132615	AND Enantiomer NH H ₂ N [*]C[c]1:s:[*]:[c]:1[*]	0.197	13 out of 13
FCFP_12	17	AND Enantiomer S NH H ₂ N [*]:S:[*]	0.189	48 out of 49
	Top Fe	atures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	307419094	AND Enantiomer N N H ₂ N ¹ [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	O	43 out of 52
FCFP_12	203677720	AND Enantiomer S N NH H ₂ N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0	319 out of 382
FCFP_12	136597326	AND Enantiomer S N NH H ₂ N [*]C([*])C	0	612 out of 753

Irritant

Irritant

0.753

28ZPAK-;104;72

 $|C_{28}H_{28}N_6S_2|$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18

Bayesian Score: 0.917

Mahalanobis Distance: 9.37

Mahalanobis Distance p-value: 0.315

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	ANILINE;N;N'-1;4- ANTHRAQUINONYLENEB IS(4-PHENOXY-	BENZANILIDE;2';2'"- DITHIOBIS-	DINAPHTHO(1;2;3- CD:3';2';1'-IM)PERYLENE- 5;10-DIONE;16;17- DIHYDROXY
Structure	T N N N N N N N N N N N N N N N N N N N	O H H H H H H H H H H H H H H H H H H H	ОН

Model Applicability

Actual Endpoint

Distance

Reference

Predicted Endpoint

Structural Similar Compounds

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

Non-Irritant

Non-Irritant

28ZPAK-;173;72

0.737

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Non-Irritant

Non-Irritant

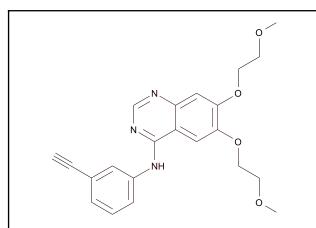
28ZPAK-;114;72

0.719

4. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Feature Struc		Irritant in training set
	ture Score	I . •
		PC:
[*][c]1:[*]:[c](0.208	44 out of 44
	[*][c]1:[*]:[c]	N NH S NH

FCFP_12	-1539132615	AND Enantiomer N N NH S NH [*]C[c]1:s:[*]:[c]]:1[*]	0.197	13 out of 13
FCFP_12	17	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.189	48 out of 49
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	0	AND Enantiomer S N N N N N N N N N N N N	0	1184 out of 1397
FCFP_12	-1272798659	AND Enantiomer N N N N N N N N N N N N N N N N N N	O	517 out of 643
FCFP_12	-1043339860	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0	146 out of 186



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

	Prediction
MOGEL	Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18

Bayesian Score: 2.13

Mahalanobis Distance: 9.75

Mahalanobis Distance p-value: 0.161

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Benzoic acid; p-(N-butyl- 2-(butylamino)acetamido)- ; butyl ester;	COLCHICINE	Cinchoninamide; 2- butoxy-N-(2- (diethylamino)ethyl)-; monohydrochloride		
Structure	H N N N N N N N N N N N N N N N N N N N	HN 2r	NH NH		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Non-Irritant	Irritant	Irritant		
Distance	0.637	0.682	0.747		
Reference	Arzneimittel-Forschung 8;609;58	AJOPAA 31;837;48	Arzneimittel-Forschung 8;181;58		

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 3. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 902193919: [*]:[c](:[*])C#C

Feature Contribution Top features for positive contribution						
FCFP_12	1747237384	[*][c]1:[*]:[c](: [*]):s:1	0.208	44 out of 44		

FCFP_12	178336375	oʻ	0.202	19 out of 19
		N N N N N N N N N N N N N N N N N N N		
		[*]:[cH]:[c](:n:[*]): [c](:[*]):[*]		
FCFP_12	17	[c](:[*]):[*]	0.189	48 out of 49
TOTF_12		ا	0.109	40 Out 01 49
		N N N		
		[*]:s:[*]		
	Top Feat	ures for negative of	ontribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	341504799	ĵ	0	17 out of 22
		N N N N N N N N N N N N N N N N N N N		
FCFP_12	-1272768868	[*]CCOC	0	396 out of 514
		N • O		
		N H		
		[*]CCO[*]		
FCFP_12	1	ر ا	0	872 out of 1051
		N N N N N N N N N N N N N N N N N N N		
		\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \		
		[*]0[*]		

AND Enantiomer S N NH CI

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.288
Enrichment: 0.893
Bayesian Score: -0.553
Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 0.00845

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Meclofenamate	Levonorgestrel
Structure	OH	OH HN CI	O H
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.638	0.651	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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	ntribution			
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	85262808	AND Enantomer N N N N H CI [*][c]1:[*]:[c](: [*]):S:1	0.851	5 out of 5

ECFP_12	553181281	AND Enantiomer S N N H CI [*][C@@H]1[*][c]2:[*] :[*]:[c](:[*]):[c]:2 CC1	0.613	2 out of 2
ECFP_12	-177077903	AND Enantiomer N N N	0.529	6 out of 10
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	99947387	AND Enantiomer S N N N N N H CI [*]:[c](:[*])CI	-0.817	8 out of 62
ECFP_12	1854732111	AND Enantiomer N N N N H H C [*][c]1:[c](;[cH]:[c](;[cH]:[cH]:1	-0.816	4 out of 33
ECFP_12	-769655848	AND Enantiomer S N N N N H CI [*][c]1:[cH]:[cH]:[c] (CI):[cH]:[cH]:1	-0.797	1 out of 11

N NH NH

 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.371
Enrichment: 1.15
Bayesian Score: 3.05
Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.169

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simil	ar Compounds		
Name	Nafenopin	Meclofenamate	Levonorgestrel
Structure	OH OH	OH HN CI	OH NAME OF THE OWN OWN OF THE OWN OWN OF THE OWN
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.622	0.622	0.648
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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Contribution					
	Top fe	atures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	85262808	AND Enantiomer N N N F [*][c]1:[*]:[*]:[c](: [*]):s:1	0.851	5 out of 5	

ECFP_12	553181281	AND Enantiomer N N H F	0.613	2 out of 2
ECFP_12	914325265	[*][C@@H]1[*][c]2:[*] :[*]:[c](:[*]):[c]:2 CC1 AND Enantiomer	0.516	8 out of 14
		NH NH		
	Ton Fea	[*]:s:[*] tures for negative (contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	292958156	AND Enantiomer N NH F	-0.485	0 out of 2
ECFP_12	220735655	[*]CC(C)C[*] AND Enantiomer S NH NH [*]:[c](:[*])F	-0.327	4 out of 19
ECFP_12	-1311285389	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.318	2 out of 10

[*][c](:[*]):[c](F):[cH]:[*]

N NH NH

 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568
Rotatable Bonds: 3
Acceptors: 4

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.432
Enrichment: 1.34
Bayesian Score: 5.15
Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 0.00124

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Indomethacin	Estrogens; conjug.	Nafenopin
Structure	HO	O IN	

Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.592	0.637	0.638
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

Structural Similar Compounds

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	85262808	[*][c]1:[*]:[c](: [*]):s:1	0.851	5 out of 5		

ECFP_12	-177077903	[*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10
ECFP_12	914325265	N N N N N N N N N N N N N N N N N N N	0.516	8 out of 14
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	552088469	[*]:[c]1:[*]:[c]2 [*]CCC[c]:1:2	-0.272	0 out of 1
ECFP_12	-1672647522	[*]C[c]1:s:[*]:[c]:1[*]	-0.272	0 out of 1
ECFP_12	864909220	N=N N=N H N=N H N=N (*)OC	-0.147	12 out of 45

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.474
Enrichment: 1.47
Bayesian Score: 6.43
Mahalanobis Distance: 12.1

Mahalanobis Distance p-value: 0.00357

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Niclosamide	Phenolphthalein	Pyrimethamine
Structure	CIAN OH	НО	NH ₂ N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.573	0.584	0.638
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

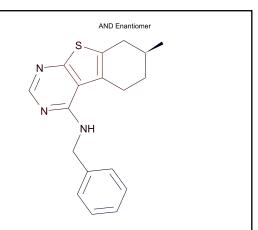
Model Applicability

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: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	85262808	AND Enantomer N N N H N (*)[c]1:[*]:[*]:[c](: [*]):s:1	0.851	5 out of 5	

ECFP_12	553181281	AND Enantiomer N N N N N H ₂ N [*][C@@H]1[*][c]2:[*] :[*]:(c]:(-1):[c]:2 CC1	0.613	2 out of 2
ECFP_12	773910207	AND Enantitioner N	0.553	4 out of 6
		ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	292958156	AND Enantioner S N N H 2 N [*]CC(C)C[*]	-0.485	0 out of 2
ECFP_12	-1672647522	AND Enantiamer N N N N H ₂ N [*]C[c]1:s:[*]:[c]]:1[*]	-0.272	0 out of 1
ECFP_12	865482986	AND Enantioner N N H H T T T T T T T T T T T	-0.233	4 out of 17



C₁₈H₁₉N₃S

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.362
Enrichment: 1.13
Bayesian Score: 2.75
Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 9.36e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Nafenopin	Meclofenamate	Pergolide	
Structure	OH O	OH HNV	No. of the state o	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen	
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen	
Distance	0.577	0.624	0.627	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	85262808	AND Enantiomer N N N H H (*][c]1:[*]:[*]:[c](: [*]):s:1	0.851	5 out of 5	

ECFP_12	553181281	AND Enantiomer N N N H H	0.613	2 out of 2
ECFP_12	914325265	[*][C@@H]1[*][c]2:[*] :[*]:[c](:[*]):[c]:2 CC1	0.516	8 out of 14
	Top For	[*]:s:[*] tures for negative (contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-281505363	AND Enantiomer N N N H H (*)[c]1:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	1571214559	AND Enantiomer N N H H (*)1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	491100606	AND Enantiomer	-0.485	0 out of 2

[*]CN[c](:[*]):[*]

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.328
Enrichment: 1.02
Bayesian Score: 1.35
Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.00572

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly linaccurate.

Name	Furosemide	Phenazopyridine	Niclosamide
Structure	HO O NH2 NH2 O S O O O O O O O O O O O O O O O O O	N N N N N N N N N N N N N N N N N N N	CI AND HAND ON THE STATE OF THE
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.724	0.738	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.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
- 4. Unknown ECFP_2 feature: -571028867: [*]NC(=S)N

Top features for positive contribution				
Fingerprint	Bit/Smiles Feature Structure		Score	Carcinogen in training set
ECFP_12	85262808	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.851	5 out of 5

ECFP_12	553181281	AND Enantiomer N N N N N H H N N 1 S [*][C@@H]1(*][c]2-[t*] :(*]:[o](-1*)]:[o]:2 CC1	0.613	2 out of 2
ECFP_12	914325265	AND Enantiomer NH HN NH S NH S [*]:s:[*]	0.516	8 out of 14
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	292958156	AND Enantiomer S N NH HN NH S NH ₂ [*]CC(C)C[*]	-0.485	0 out of 2
ECFP_12	1979182050	AND Enantiomer NH NH HN NH ₂ [*]C(=S)[*]	-0.485	0 out of 2
ECFP_12	-845108448	AND Enantiomer S NH NH HN NH ₂ [*]=S	-0.485	0 out of 2

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

AND Enantiomer
N NH
O S-NH ₂

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.332
Enrichment: 1.03
Bayesian Score: 1.53
Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.00246

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Indapamide	Niclosamide	Metolazone
Structure	H ₂ N _C O	CI AND HAND ON THE CONTRACT OF	O I I I NH NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint Non-Carcinogen		Non-Carcinogen	Non-Carcinogen
Distance	0.598	0.625	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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	85262808	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.851	5 out of 5		

ECFP_12	553181281	AND Enandomer S N N N N N N N N N N N N	0.613	2 out of 2
ECFP_12	914325265	AND Enantiomer S NH NH S NH ₂ [*]:S:[*]	0.516	8 out of 14
	Top Featur	es for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-934226723	AND Enantiomer NH NH S-NH O [*]S(=[*])(=[*])N	-1.16	0 out of 7
ECFP_12	-450780499	AND Enantionner S NH NH NH (*)[C](:[*]):[C](:[CH]:[*])S(=O)(=O)N	-1.06	0 out of 6
ECFP_12	-2121766239	AND Enantiomer S N N N N N N N N N N N N N N N N N N	-1.06	0 out of 6

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.341
Enrichment: 1.06
Bayesian Score: 1.92
Mahalanobis Distance: 12.8

Mahalanobis Distance p-value: 0.000276

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Indapamide	Niclosamide	Metolazone		
Structure	H ₂ N _O O	CI MOH	H ₂ N ² S th		
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Distance	0.614	0.629	0.650		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &		

Model Applicability

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

Res.) Sept. 1997

Res.) Sept. 1997

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

Res.) Sept. 1997

2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	Feature Contribution					
	Top fe	atures for positive of	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	85262808	AND Enantomer N N N N N (*)[c]1:[*]:[c](: [*]):s:1	0.851	5 out of 5		

ECFP_12	553181281	AND Enantiomer N N H N (*][C@@H]1[*][c]2:[*] :[*]:[c]:[*]:[c]:2 CC1	0.613	2 out of 2
ECFP_12	-177077903	AND Enantiomer S N N N N (*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10
		ures for negative of	-	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-934226723	AND Enantioner N N N N N N N N N	-1.16	0 out of 7
ECFP_12	-2121766239	AND Enuntriomer S N N N H N (*):[c](:[*])S(=0)(=0)N	-1.06	0 out of 6
ECFP_12	292958156	AND Enantioner N N N N N H N H 2N C 19 C C C C C C C C C C C C C C C C C	-0.485	0 out of 2

US FDA (Centre for Drug

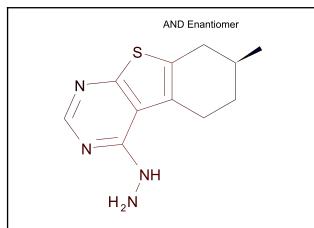
Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.43 Enrichment: 1.34 Bayesian Score: 5.09 Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.00578

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediction. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Lamotrigine	Phenazopyridine	Dapsone		
Structure	H ₂ N ^M NH ₂	N N N H ₂ N ^M NH ₂	H ₂ N O S NH ₂		
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen		
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen		
Distance	0.578	0.594	0.606		

Model Applicability

Reference

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Unknown ECFP 2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	Feature Contribution					
	Top fe	atures for positive of	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	85262808	AND Enantlomer NH H ₂ N [*][c]1:[*]:[*]:[c](: [*]):s:1	0.851	5 out of 5		

ECFP_12	553181281	AND Enantiomer S N H ₂ NH [*][C@@H]1[*][c]2:[*] :[*]:[c]:[2 CC1	0.613	2 out of 2
ECFP_12	914325265	AND Enantiomer S NH H ₂ N [*]:s:[*]	0.516	8 out of 14
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	292958156	AND Enantiomer S NH H ₂ N [*]CC(C)C[*]	-0.485	0 out of 2
ECFP_12	-1672647522	AND Enantiomer S NH H ₂ N [*]C[c]1:s:[*]:[c]:1[*]	-0.272	0 out of 1
ECFP_12	865482986	AND Enantiomer S NH H ₂ N [*]C([*])C	-0.233	4 out of 17

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

AND Enantiomer S N
NH N N H
S

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.401
Enrichment: 1.25
Bayesian Score: 4.13
Mahalanobis Distance: 13.2

Mahalanobis Distance p-value: 7.35e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Probucol	Emetine	Bromocriptine		
Structure	HO	The state of the s	Br H N N N N N N N N N N N N N N N N N N		
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen		
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen		
Distance	0.878	0.906	0.927		
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: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set ECFP_12 85262808 0.851 5 out of 5

ECFP_12	553181281	AND Enantioner N S N N N N N N N N N N	0.613	2 out of 2
ECFP_12	-177077903	AND Enantioner N S N N NH S [*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	292958156	AND Enantioner N N N N N N N N N N N N N N N N N N N	-0.485	0 out of 2
ECFP_12	-1672647522	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.272	0 out of 1
ECFP_12	865482986	AND Enantioner N N N N N N N N N N N N N N N N N N N	-0.233	4 out of 17

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

N	
N NH O O	

 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.285
Enrichment: 0.885
Bayesian Score: -0.692
Mahalanobis Distance: 16.3

Mahalanobis Distance p-value: 1.65e-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 esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Mycophenolate	Nicardipine	Nimodipine
Structure	HO		H N PO
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.644	0.660	0.675
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

- 1. OPS PC28 out of range. Value: 3.6564. Training min, max, SD, explained variance: -2.8936, 3.5771, 1.028, 0.0111.
- 2. Unknown ECFP_2 feature: 1139738044: [*]:[c](:[*])C#C

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1545539812	[*]C#C	0.78	8 out of 10

ECFP_12	-1939823063	, N, O, O	0.78	8 out of 10
		N H OO		
		[*]#C		
ECFP_12	-177077903	[*]N[c](:[cH]:[*]):[c	0.529	6 out of 10
		H]:[*]		
	Top Featu	res for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-2063600634	[*]O[c]1:[cH]:[c](:[c]([*]):[*]):[c](:[*]):[*]:[c]:1[*]	-0.661	0 out of 3
ECFP_12	-1253653003	[*]COC	-0.661	0 out of 3
ECFP_12	-2063202154	[*]O[c]1:[cH]:[c](:n: [*]):[c](:[*]):[*]:[c]:1[*]	-0.661	0 out of 3

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

AND Enantiomer	_
N NH	
CI	

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3 Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.441 Enrichment: 1.18 Bayesian Score: -2.22 Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.029

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	ilar Compounds Mestranol	Pergolide	Brotizolam
Structure	HO HO	- S H	Br N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.622	0.637	0.694

Model Applicability

Reference

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.

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Feature Co	ntribution				
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
SCFP_4	-1065373877	AND Enantiomer S N N N N C I [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.721	3 out of 3	

SCFP_4	-1181430618	AND Enantiomer S N N H CI [*]:n:[cH]:n:[*]	0.663	4 out of 5
SCFP_4	17	AND Enantiomer N N N H CI [*]:S:[*]	0.548	10 out of 17
	Top Feat	tures for negative	contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantomer N N N N N H H C([*])C([c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	-1849867720	AND Enantiomer N N N N N CI [*][C@@H]1[*][c]2:[*] :[*]:[c]:[:"]:[c]:2 CC1	-0.73	1 out of 10
SCFP_4	112346096	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.73	1 out of 10

AND Enantiomer S N NH

 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.496 Enrichment: 1.33

Bayesian Score: -0.556 Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 0.00242

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Pergolide	Mestranol	Flurbiprofen
Structure	N H	HO THO	OH OH
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.612	0.621	0.647
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1065373877	AND Enantiomer S N H F [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.721	3 out of 3

SCFP_4	-1181430618	AND Enantiomer S NH NH [*]:n:[cH]:n:[*]	0.663	4 out of 5
SCFP_4	-730654023	AND Enunioner N	0.571	3 out of 4
		es for negative c		
•		Feature Structure		Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enumidomer S N N N H F [*]C([*])C[c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	-1849867720	AND Enantiomer S N N N H F [*](C@@H 1*[*](c 2;*[*]) :[*]:(c 2;*[*]) CC1	-0.73	1 out of 10
SCFP_4	112346096	AND Enantiomer N N N N N N N N	-0.73	1 out of 10

N NH NH

C₁₇H₁₇N₃OS

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.467
Enrichment: 1.25
Bayesian Score: -1.5
Mahalanobis Distance: 14.1

Mahalanobis Distance p-value: 2.16e-005

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Estrogens; conjug.	Omeprazole	Cytembena
Structure	ON SON	N N N N N N N N N N N N N N N N N N N	OH Br
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.598	0.622	0.636
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Contribution						
	Top fea	tures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
SCFP_4	-1065373877	[*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.721	3 out of 3		

SCFP_4	-1181430618	S N	0.663	4 out of 5
SCFP_4	17	[*]:n:[cH]:n:[*]	0.548	10 out of 17
SCFP_4		N N N N N N N N N N	0.548	TO out of 17
	Top Feat	ures for negative	contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	[*]C([*])C[c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	112346096	[*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.73	1 out of 10
SCFP_4	-1849867720		-0.73	1 out of 10

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.436
Enrichment: 1.17
Bayesian Score: -2.32
Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 0.0214

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly linaccurate.

Name	<u> </u>	Fatranana, aanius	Davefaranam
Name	Phenolphthalein	Estrogens; conjug.	Doxefazepam
Structure	НО	ON OH	OH OH
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.556	0.651	0.661
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

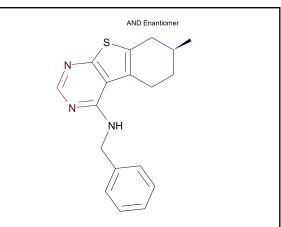
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

i eature contribution							
Top features for positive contribution							
Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set				
-1065373877	**Note: The second of the seco	0.721	3 out of 3				
	Top fea	Top features for positive of Bit/Smiles Feature Structure -1065373877 AND Enantiomer S H H 2, N [*][c]1:[*]:[c](:[*])	Top features for positive contribution Bit/Smiles Feature Structure Score -1065373877 0.721				

SCFP_4	-1181430618	AND Enantiomer S N N H ₂ N [*]:n:[cH]:n:[*]	0.663	4 out of 5
SCFP_4	17	AND Enantiomer N N N N H ₂ N [*]:s:[*]	0.548	10 out of 17
	Top Feat	tures for negative	contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantiomer S N N N N N N N N N N N N	-1.16	1 out of 17
SCFP_4	-1849867720	AND Enantiomer N N N N N H ₂ N [*][C@@H]1(*)[c]2:[*] :[*]:[c](:[*)]:[c]:2 CC1	-0.73	1 out of 10
SCFP_4	112346096	AND Enumirorer N	-0.73	1 out of 10



 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.45
Enrichment: 1.2
Bayesian Score: -1.97
Mahalanobis Distance: 8.82

Mahalanobis Distance p-value: 0.283

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Pergolide	Flurbiprofen	Mestranol		
Structure	S N N N N N N N N N N N N N N N N N N N	P _A OH	HO THOUSAND		
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen		
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen		
Distance	0.552	0.630	0.632		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &		

Model Applicability

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

Res.) Sept. 1997

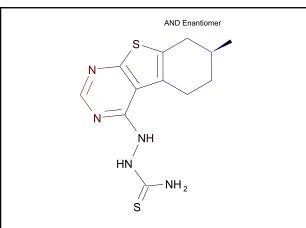
Res.) Sept. 1997

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

Res.) Sept. 1997

Feature Contribution							
	Top fea	tures for positive o	ontribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set			
SCFP_4	-1065373877	**************************************	0.721	3 out of 3			

SCFP_4	-1181430618	AND Enantiomer S NH NH [*]:n:[cH]:n:[*]	0.663	4 out of 5
SCFP_4	17	AND Enantiomer NH NH [*]:S:[*]	0.548	10 out of 17
	Top Feat	ures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantiomer N N N N N H	-1.16	1 out of 17
SCFP_4	112346096	AND Enantiomer N N N N N N N N N (H (E) [c] [:[c] 1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.73	1 out of 10
SCFP_4	-1849867720	AND Enandomer N N N N N N N (*) [*](C@@H)1[*](c]2:[*] :[*]:[c](:[*)):[c]:2 CC1	-0.73	1 out of 10



 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4 Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.496 Enrichment: 1.33 Bayesian Score: -0.532

Mahalanobis Distance: 14.7

Mahalanobis Distance p-value: 5.9e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Propylthiouracil	Phenazopyridine	Torsemide
Structure	HN H	H ₂ N ^M NH ₂	HN NH NH NH
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.718	0.731	0.748
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Co	ntribution			
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1065373877	AND Enantiomer S N N N N N N H H N NH ₂ S [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.721	3 out of 3

SCFP_4	-1181430618	AND Enantiomer S NH HN NH S NH ₂ S	0.663	4 out of 5
SCFP_4	17	[*]:n:[cH]:n:[*] AND Enantiomer NH HN NH S NH S [*]:s:[*]	0.548	10 out of 17
	Top Feat	tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantiomer N N N N N H H NH ₂ S [*]C([*])C[c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	112346096	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.73	1 out of 10
SCFP_4	-1849867720	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.73	1 out of 10

Molecular Weight: 374.48042

ALogP: 3.541
Rotatable Bonds: 3

Acceptors: 5
Donors: 2

 $C_{17}H_{18}N_4O_2S_2$

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.312 Enrichment: 0.834 Bayesian Score: -4.89 Mahalanobis Distance: 13

Mahalanobis Distance p-value: 0.000225

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Torsemide	Estrogens; conjug.	Sulfamethazine
Structure	HN NH NH	ON O	HN M
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Predicted Endpoint Single-Carcinogen		Single-Carcinogen
Distance	0.631	0.646	0.668
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Co	ntribution			
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1065373877	**ND Enantiomer **S **NH2 [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.721	3 out of 3

SCFP_4	-1181430618	AND Enantiomer NH NH S-NH ₂ [*]:n:[cH]:n:[*]	0.663	4 out of 5
SCFP_4	17	AND Enantiomer S NH NH S NH ₂ [*]:S:[*]	0.548	10 out of 17
		res for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantiomer S N N N N N N N N N N N N	-1.16	1 out of 17
SCFP_4	112346096	AND Enantioner N N N N N N N N N 1 (*1):[c]:[c]:[c] ((*1):[*]:[c]:1: [*]	-0.73	1 out of 10
SCFP_4	-1849867720	AND Enantiomer N	-0.73	1 out of 10

TOPKAT_Rat_Female_FDA_Single_vs_Multiple

AND Enantiomer
N NH
$O=S$ H_2N O

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

IIVIOUEI FIEUICIIOII	M	odel	Prediction	1
----------------------	---	------	-------------------	---

Prediction: Single-Carcinogen

Probability: 0.305
Enrichment: 0.816
Bayesian Score: -5.01
Mahalanobis Distance: 11.6

Mahalanobis Distance p-value: 0.00379

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Estrogens; conjug.	Torsemide	Sulfamethazine
Structure	O THE STATE OF THE	MH NH	HN NH 2
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.644	0.649	0.655
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Co	ntribution			
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1065373877	AND Enantiomer N N N N H 2N (*)[c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.721	3 out of 3

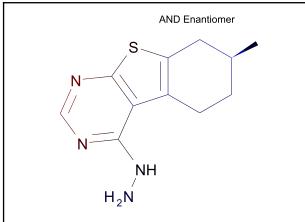
SCFP_4	-1181430618	AND Enantiomer N N N H O'S H ₂ N [*]:n:[cH]:n:[*]	0.663	4 out of 5
SCFP_4	17	AND Enantiomer N N N N H H O'S H ₂ N [*]:S:[*]	0.548	10 out of 17
	Top Feat	tures for negative	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantiomer N N N N H H N (*) (*) (*) (*) (*) (*) (*) (*)	-1.16	1 out of 17
SCFP_4	-1849867720	AND Enantiomer N N N H 2 N 1"[IC@@H]1["]0[2:["] :["]:[0]:["]:[0]:2 CC1	-0.73	1 out of 10
SCFP_4	112346096	AND Enuritioner N N N N N N N N N N N N N N N N N N N	-0.73	1 out of 10

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Eval.& Res./Off. Testing &

Res.) Sept. 1997



 $C_{11}H_{14}N_{4}S$

Donors: 2

Molecular Weight: 234.32065

ALogP: 2.561
Rotatable Bonds: 1
Acceptors: 4

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.427
Enrichment: 1.14
Bayesian Score: -2.54
Mahalanobis Distance: 13.4

Mahalanobis Distance p-value: 0.000104

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Propylthiouracil	Phenazopyridine	Hydralazine	
Structure	HN H	N N N N N N N N N N N N N N N N N N N	N N NH 2	
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen	
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen	
Distance	0.560	0.570	0.595	
Reference	US FDA (Centre for Drug	US FDA (Centre for Drug	US FDA (Centre for Drug	

Model Applicability

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

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

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Feature Co	ntribution			
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1065373877	AND Enantiomer S NH H ₂ [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.721	3 out of 3

SCFP_4	-1181430618	AND Enantiomer S NH H ₂ N [*]:n:[cH]:n:[*]	0.663	4 out of 5
SCFP_4	17	AND Enantiomer NH H ₂ N [*]:s:[*]	0.548	10 out of 17
	Top Featu	res for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantioner N N N N H ₂ N [*]C([*])C[c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	-1849867720	AND Enantiomer N N N N H ₂ N [†] [*][C@@H]1[*][c]2:[*] :[*]:[c]:[*]:[c]:2 CC1	-0.73	1 out of 10
SCFP_4	112346096	AND Enantioner S N N N H ₂ NH [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.73	1 out of 10

 $|C_{28}H_{28}N_6S_2|$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.387
Enrichment: 1.04
Bayesian Score: -3.43
Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.00213

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Bromocriptine	Deserpidine	Simvastatin		
Structure	Br - V				

			ОН
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.893	0.903	0.922
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

Structural Similar Compounds

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

Feature Co	Feature Contribution					
	Top fea	ntures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
SCFP_4	-1065373877	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.721	3 out of 3		

SCFP_4	-1181430618	AND Enantiomer N S N N N N N N N N N N N N N N N N N	0.663	4 out of 5
SCFP_4	17	AND Exantiomer N N N N N N N N N N N N N N N N N N N	0.548	10 out of 17
	Top Feat	ures for negative	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1272709286	AND Enantioner N N N N N N N *]C([*])C[c](:[*]):[*]	-1.16	1 out of 17
SCFP_4	112346096	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.73	1 out of 10
SCFP_4	-1849867720	AND Enantionner N S N N N N N N N N N N N N N N N N N	-0.73	1 out of 10

N O O
NH

 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.423
Enrichment: 1.13
Bayesian Score: -2.63
Mahalanobis Distance: 15.2

Mahalanobis Distance p-value: 2.08e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Nicardipine	Diltiazem	Moricizine	
Structure	H N N N N N N N N N N N N N N N N N N N	N N N N N N N N N N N N N N N N N N N	N N N N N N N N N N N N N N N N N N N	
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen	
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen	
Distance	0.632	0.748	0.749	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	

Model Applicability

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

Feature Co		tures for positive o	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1065373877	[*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.721	3 out of 3

SCFP_4	-1181430618	N Z H	0.663	4 out of 5
SCFP_4	2142015375	[*]:n:[cH]:n:[*]	0.433	2 out of 3
301 F <u>+</u>	2142013373	[*]:[cH]:[c]:1:[*]	0.433	2 out of 3
		tures for negative o	-	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	622342378	[*][c]1:[]:[cH]:[c]2 °· :[c]([*]):[*]:[cH]:n :[c]:2:[cH]:1	-0.816	0 out of 4
SCFP_4	112346096	[*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1:	-0.73	1 out of 10
SCFP_4	1242547645	[*]CO[c]1:[cH]:[c](:[*]):[*]:[cH]:[c]:10[*]	-0.489	0 out of 2

AND Enantiomer S N NH CI

 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.373
Enrichment: 1.11
Bayesian Score: 0.553
Mahalanobis Distance: 17.3

Mahalanobis Distance p-value: 5.48e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Meclofenamate	Nafenopin	Indomethacin		
Structure	OH HN C	OH OH	HO		
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen		
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen		
Distance	0.614	0.623	0.641		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997		

Model Applicability

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

Feature Contribution					
	Top fea	atures for positive c	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	1310748454	AND Enantiomer N N N N N (i) (i) (i') (i') (i') (i') (i') (i') (0.437	7 out of 13	

SCFP_6	-55982834	AND Enantiomer N S N N N H CI [*]:[c]1:[*]C[C@@H](C)CC1	0.432	4 out of 7
SCFP_6	-1065373877	AND Enantiomer N N N N H CI [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.429	3 out of 5
	Top Feat	ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	341480432	AND Enantiomer S N N N N N N N N N N N N	-0.674	0 out of 3
SCFP_6	1905487031	AND Enantiomer S N=N N H CI [*][c]1:[cH]:[c] (CI):[cH]:[cH]:1	-0.48	2 out of 12
SCFP_6	-1272709286	AND Enantiomer N N N N N N N N N N N N N	-0.459	12 out of 61

US FDA (Centre for Drug

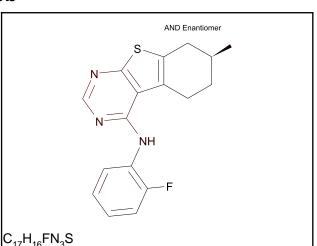
Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.463
Enrichment: 1.39
Bayesian Score: 3.32
Mahalanobis Distance: 17.6

Mahalanobis Distance p-value: 1.34e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Meclofenamate	Pergolide
Structure	OH OH	OH HN CI	S N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.610	0.612	0.643

Model Applicability

Reference

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	1310748454	AND Enantiomer N N H F [*][c]1:[*]:[c](: [*]):s:1	0.437	7 out of 13	

SCFP_6	-55982834	AND Exantioner N	0.432	4 out of 7
SCFP_6	-1065373877	AND Enantiomer S N N F [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.429	3 out of 5
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1551011249	AND Enantiomer S N H F [*]N[c]1:[cH]:[*]:[cH]:[cH]:[6]:1F	-0.674	0 out of 3
SCFP_6	-1272709286	AND Enantioner S N H F [*]C([*])C[c](:[*]):[*]	-0.459	12 out of 61
SCFP_6	815479470	AND Enantioner S N N H F [*][C@H]1[*]C[c]2:[c] (C1):s:[c](;[1]):[c] :2:[*]	-0.278	0 out of 1

Non-Carcinogen

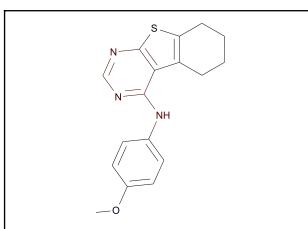
Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.618



 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.454
Enrichment: 1.36
Bayesian Score: 3.05
Mahalanobis Distance: 17.5

Mahalanobis Distance p-value: 1.5e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Indomethacin	Nafenopin	Estrogens; conjug.
Structure	HO		H O II OH

Carcinogen

Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.600

Model Applicability

Actual Endpoint

Distance

Reference

Predicted Endpoint

Structural Similar Compounds

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Non-Carcinogen

Non-Carcinogen

Res.) Sept. 1997

0.561

Feature Contribution Top features for positive contribution				
SCFP_6	1310748454	[*][c]1:[*]:[c](: [*]):s:1	0.437	7 out of 13

SCFP_6	494226440		0.432	4 out of 7
SCFP_6	-1065373877	[*]:[c]1:[*]CCCC1	0.429	3 out of 5
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1272709286	[*]C([*])C[c](:[*]):[*]	-0.459	12 out of 61
SCFP_6	1287669168	[*][c]1:[cH]:[cH]:[c] (OC):[cH]:[cH]:1	-0.38	1 out of 6
SCFP_6	815479470	[*][C@H]1[*]C[e]2:[e] (C1):s:[e](:[*]):[e] :2:[*]	-0.278	0 out of 1

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.497 Enrichment: 1.49 Bayesian Score: 4.26 Mahalanobis Distance: 16

Mahalanobis Distance p-value: 4.48e-009

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Niclosamide	Phenolphthalein	Pyrimethamine
Structure	I A OH	НО	NH 2 N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.561	0.573	0.606
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing

Model Applicability

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

Res.) Sept. 1997

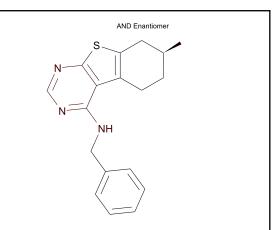
Res.) Sept. 1997

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

Res.) Sept. 1997

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	1181470699	AND Enandiomer N N H N (*) [*][c]1:[cH]:[cH]:[c] (N):[cH]:[cH]:1	0.536	4 out of 6	

SCFP_6	-192131231	AND Enantiomer N	0.52	5 out of 8
SCFP_6	1310748454	AND Enantiomer N H H 2N [*][c]1:[*]:[c](: [*]):s:1	0.437	7 out of 13
	Top Featu	res for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1272709286	AND Enantiomer N N N N N H 2N (**]C([**])C[c](:[**]):[**]	-0.459	12 out of 61
SCFP_6	815479470	AND Enantiomer N N N H 2 N (*][C@H]]1[*]C[c]2:[c] (C1):s:[c](:[¹]):[c] :2:[¹]	-0.278	0 out of 1
SCFP_6	2109165795	AND Enantiomer S N N H ₂ N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.226	42 out of 165



 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.438
Enrichment: 1.31
Bayesian Score: 2.59
Mahalanobis Distance: 16.2

Mahalanobis Distance p-value: 2.02e-009

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Pergolide	Meclofenamate
Structure		OH OH	OH HN N

Non-Carcinogen

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.587

Non-Carcinogen

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.612

Model Applicability

Actual Endpoint

Distance

Reference

Predicted Endpoint

Structural Similar Compounds

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Carcinogen

Carcinogen

Res.) Sept. 1997

0.550

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	1310748454	AND Enantiomer N N N (H (E) (C) (1:[*]:[*]:[c](: (*)):S:1	0.437	7 out of 13	

SCFP_6	-55982834	AND Enantiomer S N N N H I']:[e]1:["]C[C@@H][C)CC1	0.432	4 out of 7
SCFP_6	-1065373877	AND Enantiomer S N (H (H (E) (E) (E) (E) (E) (E)	0.429	3 out of 5
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	18117904	AND Enantomer S NH NH [*]CN[c](:[*]):[*]	-0.578	1 out of 8
SCFP_6	1653911926	AND Enantiomer S N N N H (*)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.504	12 out of 64
SCFP_6	-1272709286	AND Enantiomer N N N N H H [*]C([*])C[c](:[*]):[*]	-0.459	12 out of 61

Molecular Weight: 293.411

 $C_{12}H_{15}N_5S_2$

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4 Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.427 Enrichment: 1.28 Bayesian Score: 2.28 Mahalanobis Distance: 16.8

Mahalanobis Distance p-value: 1.02e-010

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediction. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Furosemide	Phenazopyridine	Lamotrigine
Structure	HO O NH2 NH2 S O S	N N NH 2	CI N N N N N N N N N N N N
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.722	0.722	0.729
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

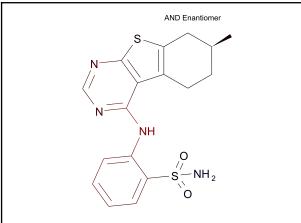
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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	
SCFP_6	1310748454	AND Enantomer N N N N N H N NH S NH ₂ S [*][c]1:[*]:[c](: [*]):s:1	0.437	7 out of 13	

SCFP_6	-55982834	AND Enantiomer S N N N N N N N N N N N N N N N N N N	0.432	4 out of 7
SCFP_6	-1065373877	AND Enantiomer S N N N H H H N NH2 S [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.429	3 out of 5
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1272709286	AND Examinmer N N N N N N N H H N NH ₂ S [*]C([*])C[c](:[*]):[*]	-0.459	12 out of 61
SCFP_6	384861283	AND Enantiomer S NH HN NH S NH ₂ [*]C(=[*])N	-0.38	1 out of 6
SCFP_6	815479470	AND Enantiomer N N N N N N H H N NH S S ["](C@H)1["]C[c]2:[c] (C1):s:[c](:[")]:[c] :2:["]	-0.278	0 out of 1



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.434
Enrichment: 1.3
Bayesian Score: 2.49
Mahalanobis Distance: 17.6

Mahalanobis Distance p-value: 1.37e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	ilar Compounds	Metolazone	Niclosamide
Structure	HN ZZ NO CI	H ₂ N NH	Cland OH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.572	0.612	0.613
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	1310748454	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.437	7 out of 13	

SCFP_6	-55982834	AND Examtomer S N N N N N N N NH ₂ [*]:[e]1:[*]C[C@@H](C)CC1	0.432	4 out of 7
SCFP_6	-1065373877	AND Enantioner S N N N N NH ₂ [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.429	3 out of 5
		es for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1358544872	AND Enantiomer NH NH S NH S NH (*) S(=[*])(=[*])N	-0.484	1 out of 7
SCFP_6	-1272709286	AND Enantioner N N N N N N N N N N N N N N N N N N	-0.459	12 out of 61
SCFP_6	-1463646519	AND Enantioner S N N N N N N N N N N N N	-0.38	1 out of 6

AND Enantiomer H₂N

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

 $C_{17}H_{18}N_4O_2S_2$

Model Prediction

Prediction: Carcinogen

Probability: 0.397 Enrichment: 1.19 Bayesian Score: 1.36 Mahalanobis Distance: 17.4

Mahalanobis Distance p-value: 3.69e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

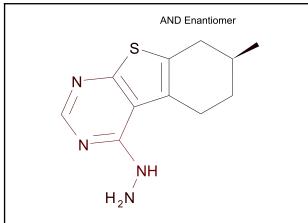
Name	Indapamide	Niclosamide	Metolazone
Structure	HN 2N O	CI AND CO	H ₂ N Structure NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.581	0.612	0.616
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	1310748454	AND Enantiomer N N N N N H H N (*) (*) (*) (*) (*) (*) (*) (*) (*) (*)	0.437	7 out of 13	

SCFP_6	-55982834	AND Enantlomer N N N N N H N N (*):[e]1:[*]C[C@@H](C)CC1	0.432	4 out of 7
SCFP_6	-1065373877	**ND Enantomer **ND Enantomer	0.429	3 out of 5
		ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1358544872	AND Enantioner N N N N N H H 2N N (*)S(=[*])(=[*])N	-0.484	1 out of 7
SCFP_6	-1272709286	AND Enantiomer N N N N N N N N N N N N N	-0.459	12 out of 61
SCFP_6	-1463646519	AND Enuntiomer N N N N N H H N [*][c](:[*]):[c](:[cH]:[*])S(=O)(=O)N	-0.38	1 out of 6



 $C_{11}H_{14}N_4S$

Donors: 2

Molecular Weight: 234.32065

ALogP: 2.561
Rotatable Bonds: 1
Acceptors: 4

Model Prediction

Prediction: Carcinogen

Probability: 0.485 Enrichment: 1.45 Bayesian Score: 3.95 Mahalanobis Distance: 18

Mahalanobis Distance p-value: 1.27e-013

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Lamotrigine	Phenazopyridine	Pyrimethamine	
Structure	CI N N N N N N N N N N N N	H ₂ N ^M NH ₂	NH 2 N N	
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen	
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen	
Distance	0.566	0.578	0.585	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	

Model Applicability

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

Res.) Sept. 1997

Res.) Sept. 1997

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

Res.) Sept. 1997

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	1310748454	NH H ₂ N [*][c]1:[*]:[c](: [*]):s:1	0.437	7 out of 13		

SCFP_6	-55982834	AND Enantiomer S NH H ₂ N [*]:[c]1:[*]C[C@@H](C	0.432	4 out of 7
SCFP_6	-1065373877	OCC1 AND Enantiomer NH H ₂ N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.429	3 out of 5
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1272709286	AND Enantiomer N N NH H ₂ N [*]C([*])C[c](:[*]):[*]	-0.459	12 out of 61
SCFP_6	815479470	AND Enantiomer N N N H ₂ NH [*][C@H]1[*]C[e]2:[e] (C1):s:[e]([*]).[e] :2:[*]	-0.278	0 out of 1
SCFP_6	5	AND Enantiomer S N NH H ₂ N [*]N	-0.232	41 out of 162

 $|C_{28}H_{28}N_6S_2|$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.418
Enrichment: 1.25
Bayesian Score: 1.99
Mahalanobis Distance: 17.3

Mahalanobis Distance p-value: 5.38e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Probucol	Emetine	Carbenicillin
Structure	HO real transfer of the control of t	D. R. L.	OHO OH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.846	0.878	0.909
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

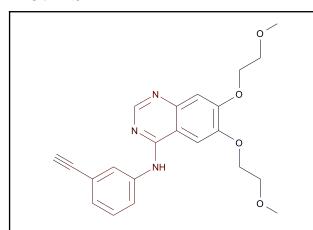
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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	1310748454	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.437	7 out of 13	

SCFP_6	-55982834	AND Enantiomer S N N N N N N (*]:[c]1:["]C[C@@H](C)CC1	0.432	4 out of 7
SCFP_6	-1065373877	AND Enantlower S N N N N N N N N N N N N	0.429	3 out of 5
	Top Feat	ures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1272709286	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.459	12 out of 61
SCFP_6	815479470	AND Enantiomer S N N N N N N N N N N N N	-0.278	0 out of 1
SCFP_6	2109165795	AND Enartiomer N S N N N N N N N N N N N N N N N N N	-0.226	42 out of 165



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.52
Enrichment: 1.56
Bayesian Score: 4.91
Mahalanobis Distance: 17.4

Mahalanobis Distance p-value: 2.91e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Mycophenolate	Nicardipine	Nimodipine		
Structure	HO	I N N N N N N N N N N N N N N N N N N N	H N N N N N N N N N N N N N N N N N N N		
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen		
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen		
Distance	0.624	0.667	0.676		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997		

Model Applicability

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

Feature Contribution					
	Top fea	ntures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	-1065373877	[*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.429	3 out of 5	

SCFP_6	123285475	[*]O[c]1:[cH]:[c](:n: [*]):[c](:[*]):[*]:[0.429	3 out of 5
SCFP_6	-2147171373	c]:1[*]	0.415	1 out of 1
	Top Feat	ures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	125474664	[*]O[c]1:[cH]:[c](:[c]):[*]):[*]:[c]:1[*]	-0.484	1 out of 7
SCFP_6	-417738003	[*]occoc	-0.264	1 out of 5
SCFP_6	951581613	[*]:[c](:[*])N[c](:[*]):[*]	-0.132	1 out of 4

AND Enantiomer	
N NH	
CI	

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.562
Enrichment: 1.36
Bayesian Score: -0.202
Mahalanobis Distance: 24.2

Mahalanobis Distance p-value: 7.52e-014

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Mestranol	Chlorpromazine
Structure	OH	HO THO	S w C
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.637	0.656	0.720
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-1065373877	AND Enantiomer S N N H CI [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.649	3 out of 3

SCFP_8	-1181430618	AND Enantomer S N N N N N (I [*]:n:[cH]:n:[*]	0.453	4 out of 6
SCFP_8	-2147171373	AND Enantlomer S N N (*) (*) (*) (*) (*) (*) (*	0.383	1 out of 1
		tures for negative		1
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-52074512	AND Enantiomer S N N N N (r) (r) (r) (r) (r) (-0.707	2 out of 14
SCFP_8	-601571304	AND Enantomer S N N N H C [*]:[cH]:[c](CI):[cH] :[*]	-0.707	2 out of 14
SCFP_8	-1378360678	AND Enantomer S N N N H CI [*][c]1:[*]:[cH]:[c](CI):[cH]:[cH]:1	-0.58	2 out of 12

AND Enantiomer S N N NH F

 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.566 Enrichment: 1.37 Bayesian Score: 0.263 Mahalanobis Distance: 23.6

Mahalanobis Distance p-value: 2.79e-013

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Mestranol	Norethindrone
Structure	OH	HO WANTED TO THE REAL PROPERTY OF THE PROPERTY	O OH
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.622	0.653	0.698
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Contribution					
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
SCFP_8	-1065373877	AND Enantiomer S N H F [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.649	3 out of 3	

SCFP_8	-1181430618	AND Enantiomer S NH NH [*]:n:[cH]:n:[*]	0.453	4 out of 6
SCFP_8	-2147171373	AND Enuntiomer S N H F [*]N[c]1:n:[cH]:n:[*] :[c]:1:[*]	0.383	1 out of 1
		es for negative c		
Fingerprint	Bit/Smiles	Feature Structure		Multiple- Carcinogen in training set
SCFP_8	1655894345	AND Enantiomer S N N H F [*][c]1:[cH]:[cH]:[cH]:[cH]:[c]:1F	-0.546	0 out of 2
SCFP_8	-1794884847	AND Enantiomer NH NH F [*]:[c](:[*])F	-0.463	1 out of 6
SCFP_8	-1381307546	AND Enuntiomer S N H F [*][c]1:[*]:[cH]:[cH] :[cH]:[c]:1F	-0.463	1 out of 6

Multiple-Carcinogen

Multiple-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.683

N NH NH

 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.569
Enrichment: 1.37
Bayesian Score: 1.66
Mahalanobis Distance: 22.3

Mahalanobis Distance p-value: 3.23e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Phenolphthalein	Flutamide
Structure		НО	O II.

Multiple-Carcinogen

Multiple-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Model Applicability

Actual Endpoint

Distance

Reference

Predicted Endpoint

Structural Similar Compounds

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

0.671

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

Multiple-Carcinogen

Multiple-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.613

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-1065373877	[*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.649	3 out of 3

SCFP_8	-1181430618	s N N N H O N H (*]:n:[cH]:n:[*]	0.453	4 out of 6
SCFP_8	951581613	[*]:[c](:[*])N[c](:[*]):[*]	0.383	1 out of 1
		ures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	136239834	s in the second	-0.358	3 out of 13
SCFP_8	803350837	[*]C[c]1:s:[*]:[c]:1[*]	-0.31	0 out of 1
SCFP_8	1287669168	[*][c]1:[cH]:[c] (OC):[cH]:[cH]:1	-0.31	0 out of 1

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.558 Enrichment: 1.35 Bayesian Score: 3.3

Mahalanobis Distance: 22.1

Mahalanobis Distance p-value: 4.4e-012

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Phenolphthalein	Oxazepam	Doxefazepam
Structure	НО	CI N H OH	OH OH
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.572	0.683	0.684
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

catal C Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
SCFP_8	-1065373877	AND Enantiomer S N H 2N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.649	3 out of 3	

SCFP_8	-1181430618	AND Enantlomer N N H H 2N [*]:n:[cH]:n:[*]	0.453	4 out of 6
SCFP_8	951581613	AND Enantiomer N N H ₂ N [*]:[c](:[*])N[c](:[*]):[*]	0.383	1 out of 1
		tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	803350837	**NO Enantiomer **No E	-0.31	0 out of 1
SCFP_8	1310748454	AND Enantioner N N H H ₂ N [*][c]1:[*]:[*]:[c](: [*]):s:1	-0.166	2 out of 7
SCFP_8	8	AND Enantiomer N N N H H H 1:n:[*]	-0.154	12 out of 39

Multiple-Carcinogen

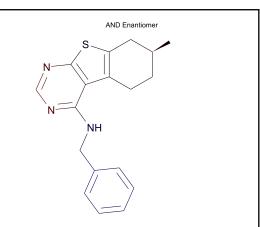
Multiple-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.685



C₁₈H₁₉N₃S

Molecular Weight: 309.42856

ALogP: 4.678
Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.554 Enrichment: 1.34

Bayesian Score: -0.906 Mahalanobis Distance: 19.9

Mahalanobis Distance p-value: 3.76e-010

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nafenopin	Mestranol	Flutamide
Structure		HO OH	O II. O NAME OF THE PARTY OF TH

Single-Carcinogen

Res.) Sept. 1997

0.677

Multiple-Carcinogen

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Multiple-Carcinogen

Multiple-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.560

Model Applicability

Actual Endpoint

Distance

Reference

Predicted Endpoint

Structural Similar Compounds

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

1. OPS PC20 out of range. Value: 3.0614. Training min, max, SD, explained variance: -2.6601, 2.9766, 1.03, 0.0144.

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Multiple-Carcinogen in training set SCFP_8 -1065373877 - -106537877 - -106537877 - -106537877 - -106537877 - -10653787 - -10653787 - -106537877 - -10653787 - -10653787 - -10653787 - -10653787 - -10653787 - -10653787 - -10653787 - -10653787 - -10653787 - -10653787 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -1065378 - -106578 - -106578 - -106578 - -106578 -

SCFP_8	-1181430618	AND Enantiomer S NH NH (*]:n:[cH]:n:[*]	0.453	4 out of 6
SCFP_8	-2147171373	AND Enantiomer N N N H H [*]N[c]1:n:[cH]:n:[*] :[c]:1:[*]	0.383	1 out of 1
	Top Feat	tures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	1653911926	AND Enantiomer S N N H H (*)[c]1:[cH]:[cH]:[cH]]:[cH]:[cH]:1	-0.985	1 out of 12
SCFP_8	1155330592	AND Enantiomer N NH NH [*]NC[c](:[*]):[*]	-0.737	0 out of 3
SCFP_8	1271460694	AND Enantioner N N H H (*)NC[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.31	0 out of 1

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.559
Enrichment: 1.35
Bayesian Score: 3.26
Mahalanobis Distance: 23.3

Mahalanobis Distance p-value: 4.48e-013

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Furosemide	Phenazopyridine	Torsemide
Structure	HO O NH2 NH2 SO	H ₂ N ^M NH ₂	HN ON NH ON NH ON NH
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.742	0.758	0.774
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

1. OPS PC20 out of range. Value: 3.0652. Training min, max, SD, explained variance: -2.6601, 2.9766, 1.03, 0.0144.

Feature Contribution

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
SCFP_8	-1065373877	AND Enantomer N N N N H N NH S NH ₂ S [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.649	3 out of 3	

SCFP_8	18	AND Enantiomer S NH NH HN NH S NH S [*]=S	0.553	2 out of 2
SCFP_8	1435188938	AND Enantiomer S NH NH HN NH ₂ [*]C(=S)[*]	0.553	2 out of 2
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	384861283	AND Enantiomer S N NH NH HN NH S NH S [*]C(=[*])N	-0.31	0 out of 1
SCFP_8	803350837	AND Enantiomer N N N N N N H N NH S S [*]C[c]1:s:[*]:[c]:1[*]	-0.31	0 out of 1
SCFP_8	1310748454	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.166	2 out of 7

US FDA (Centre for Drug

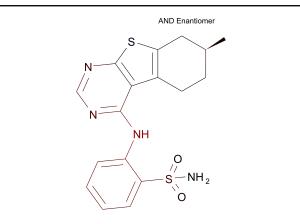
Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.558
Enrichment: 1.35
Bayesian Score: 3.32
Mahalanobis Distance: 24.5

Mahalanobis Distance p-value: 5.12e-014

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Torsemide	Doxefazepam	Bicalutamide		
Structure	HN NH NH	OH N N OH	HO HO HO NO THE FE		
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen		
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen		
Distance	0.684	0.719	0.722		

Model Applicability

Reference

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-1065373877	AND Enantiomer S N N NH ₂ [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.649	3 out of 3

SCFP_8	-1181430618	AND Enantiomer S NH O NH NH	0.453	4 out of 6
SCFP_8	-2147171373	AND Enandiomer N N N N N N N N N N N N N N N N N N	0.383	1 out of 1
		ures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	182991870	AND Enantiomer S N N N N N N N N N N N N N N N N N N	-0.342	1 out of 5
SCFP_8	-1463646519	AND Enantiomer S N N N N N N N N N N N N N N 1 (*][c](:[*]):[c](:[cH]:[*]]S(=O)(=O)N	-0.31	0 out of 1
SCFP_8	803350837	AND Enantiomer N N N N N N N N N N N N N	-0.31	0 out of 1

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.567 Enrichment: 1.37 Bayesian Score: 0.552 Mahalanobis Distance: 23.8

Mahalanobis Distance p-value: 1.59e-013

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Torsemide	Bicalutamide	Doxefazepam
Structure	HN NH NH	HO HO HN AND THE F	OH OH
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.714	0.721	0.723
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

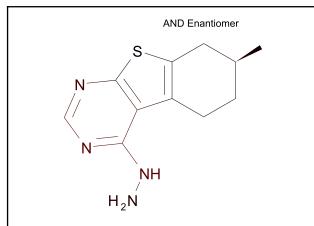
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
SCFP_8	-1065373877	AND Enantiomer S N N N N N N N N N N N N N N N N N N	0.649	3 out of 3	

SCFP_8	-1181430618	AND Enantiomer N N N H O'S H ₂ N [*]:n:[cH]:n:[*]	0.453	4 out of 6
SCFP_8	-300914917	[*]:[c](*[*])N[c]1:[c H]:[cH]:[*]:[cH]:[cH	0.383	1 out of 1
	Top Feat	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-1247518081	AND Enantlomer N	-0.737	0 out of 3
SCFP_8	1892918731	[*]N[c] [†] [cH]:[cH]:[c](:[cH]:[cH]:0) (=O)N	-0.546	0 out of 2
SCFP_8	182991870	AND Enantiomer N N N N H; N ["]:[o](:["])S(=O)(=O))N	-0.342	1 out of 5



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.568
Enrichment: 1.37
Bayesian Score: 2.08
Mahalanobis Distance: 23.2

Mahalanobis Distance p-value: 5.29e-013

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Danthron	Danthron Phenazopyridine			
Structure	O MOH	H ₂ N ^M NH ₂	H ₂ N O O O O O O O O O O O O O O O O O O O		
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen		
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen		
Distance	0.589	0.591	0.621		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997		

Model Applicability

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

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-1065373877	AND Enantomer S NH H ₂ N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.649	3 out of 3

SCFP_8	-1181430618	AND Enantiomer S NH H ₂ N [*]:n:[cH]:n:[*]	0.453	4 out of 6
SCFP_8	-915345805	AND Enantiomer S N NH H ₂ N [*]:[c](:[*])NN	0.383	1 out of 1
	Top Feat	ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	5	AND Enantiomer S NH H ₂ N [*]N	-0.463	9 out of 41
SCFP_8	803350837	AND Enantiomer N NH H ₂ N [*]C[c]1:s:[*]:[c]:1[*]	-0.31	0 out of 1
SCFP_8	1310748454	AND Enantiomer NH H ₂ N [*][c]1:[*]:[c](: [*]):s:1	-0.166	2 out of 7

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.567
Enrichment: 1.37
Bayesian Score: 2.33
Mahalanobis Distance: 23.1

Mahalanobis Distance p-value: 6.96e-013

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Deserpidine	Simvastatin	Lovastatin
Structure	Oran I The Control of		

	H	ОН	ОН
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.968	0.971	1.006
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

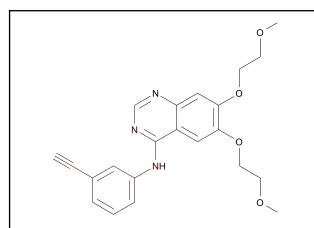
Structural Similar Compounds

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

1. ALogP out of range. Value: 7.842. Training min, max, mean, SD: -3.329, 7.238, 2.2886, 2.006.

Fingerprint Bit/Smiles Feature Structure Score Multiple-Carcinogen in training set SCFP_8 -1065373877 - O.649 3 out of 3

SCFP_8	-1181430618	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.453	4 out of 6
SCFP_8	-2147171373	[*]:n:[cH]:n:[*] AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.383	1 out of 1
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	803350837	AND Enantomer N N N N N N N N N N N N N N N N N N N	-0.31	0 out of 1
SCFP_8	1310748454	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.166	2 out of 7
SCFP_8	8	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.154	12 out of 39



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.568
Enrichment: 1.37
Bayesian Score: 0.821
Mahalanobis Distance: 16.5

Mahalanobis Distance p-value: 4.64e-007

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Nicardipine	Nimodipine	Felodipine
Structure	H N N N N N N N N N N N N N N N N N N N	H N N N O	O CI III
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.685	0.700	0.721
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Co	ntribution			
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-1065373877	[*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.649	3 out of 3

SCFP_8	2	N N N N N N N N N N N N N N N N N N N	0.584	6 out of 8
SCFP_8	-1181430618	[*]C#[*]	0.453	4 out of 6
	Top Feat	ures for negative c	ontribution	
Fingerprint	Bit/Smiles		Score	Multiple- Carcinogen in training set
SCFP_8	136239834	[*]OC	-0.358	3 out of 13
SCFP_8	446954673	[*]CO[c]1:[cH]:[c]2:n :[cH]:[*]:[c]([*]):[c]:2:[cH]:[o]:10[*]	-0.31	0 out of 1
SCFP_8	949482168	[*]ccoc	-0.31	0 out of 1

 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mild Probability: 0.382 Enrichment: 1.04

Bayesian Score: -0.773 Mahalanobis Distance: 7.93

Mahalanobis Distance p-value: 0.808

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Aniline, 2,4-bis(o- methylphenoxy)-	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	Salicylic acid, p-tolyl ester	
Structure	H ₂ N ^N	O THAT I WANT THE TANK THE TAN	MOH OH	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance	0.627	0.664	0.687	
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucich Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,242,1	FCTOD7 Food and Chemical Toxicology. (Pergamon Press Inc., Maxwell House, Fa irview Park, Elmsford, NY 10523) V.20- 1982- Volume(issue)/page/year: 21,835,1983	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Co	ntribution			
	Top fe	eatures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set

FCFP_12	-1441604640	AND Enantiomer S N N H H C [*]:[o]1:[*]C[C@@H](C)CC1	0.385	1 out of 1
FCFP_12	-1731231884	AND Enantiomer N N N H H (*):[c]1:[*]CC[C@H](C)C1	0.385	1 out of 1
FCFP_12	76292238	AND Enantiomer S N H (*]N[c]1:n:[cH]:n:[*] :[c]:1:[*]	0.385	1 out of 1

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severon in training set
FCFP_12	307419094	AND Enantiomer S N N N N N (H C) ([*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	-0.915	2 out of 18
FCFP_12	1294255210	AND Enantiomer NH NH [*]CN[c](:[*]):[*]	-0.504	2 out of 11

FCFP_12	17	AND Enantiomer S N NH	-0.332	8 out of 32
		[*]:s:[*]		

AND Enantiomer S N N NH CI

 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.918
Enrichment: 0.997
Bayesian Score: -2.53
Mahalanobis Distance: 8.77

Mahalanobis Distance p-value: 0.546

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simil	lar Compounds		
Name	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	s-Triazine, 2,4-dichloro-6- (o-chloroanilino)-	Aniline, 2,4-bis(o- methylphenoxy)-
Structure	O H H W	HN N CI	H ₂ N ^N
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.622	0.640	0.655
Reference	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,242,1	34ZIAG* -,235,69	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986

Model Applicability

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

Irritant in training
set

FCFP_12	-1539132615	AND Enantioner N N N N H CI [*]C[c]1:s:[*]:[*]:[c]:1[*]	0.0795	9 out of 9
FCFP_12	-124685461	AND Enantiorner S N N H CI [*]:n:[cH]:n:[*]	0.0734	5 out of 5
FCFP_12	-475316933	AND Enantiomer N N H CI [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.0703	4 out of 4

		atures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	AND Enantiomer S	-0.708	4 out of 10
FCFP_12	-773983804	AND Enantomer N N N N N N N N N N N N N N N N N N	-0.444	46 out of 79

590925877	AND Enantiomer	-0.434	56 out of 95
	N=(N		
	C [*]N [o](·[oH]·[*])·[o		
	H]:[*]		
	590925877	[*]N[c](:[cH]:[*]):[c	[*]N[c](:[cH]:[*]):[c

AND Enantiomer S N N NH F

C₁₇H₁₆FN₃S

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.895
Enrichment: 0.971
Bayesian Score: -2.82
Mahalanobis Distance: 8.71

Mahalanobis Distance p-value: 0.58

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds						
Name	s-Triazine, 2,4-dichloro-6- (o-chloroanilino)-	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	Aniline, 2,4-bis(o- methylphenoxy)-			
Structure	HN A NOTE OF THE PART OF THE P	O H H W H H H H H H H H H H H H H H H H	H ₂ N ₂ N ₂ N ₃ N ₄			
Actual Endpoint	Irritant	Irritant	Irritant			
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant			
Distance	0.560	0.644	0.646			
Reference	34ZIAG* -,235,69	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,242,1	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986			

Model Applicability

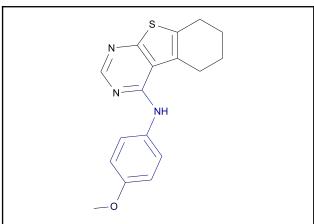
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

FCFP_12	-1539132615	AND Enantiomer N N N H F [*]C[c]1:s:[*]:[*]:[c]:1[*]	0.0795	9 out of 9
FCFP_12	-124685461	AND Enantomer S NH NH F [*]:n:[cH]:n:[*]	0.0734	5 out of 5
FCFP_12	-475316933	AND Enantomer N N N N N N N N N N N N N	0.0703	4 out of 4

Fin manneled		atures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1783756416	AND Enantiomer S N H F [*]N[c]1:[cH]:[*]:[cH]:[cH]:[c]:1F	-0.509	4 out of 8
FCFP_12	-1724769936	AND Enantomer S N H F [*]N[c]1:[cH]:[cH]:[c H]:[cH]:[c]:1[*]	-0.475	11 out of 20

FCFP_12	-773983804	AND Enantiomer	-0.444	46 out of 79
		N N		
		F		
		[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1		



 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.709
Enrichment: 0.769
Bayesian Score: -3.91
Mahalanobis Distance: 7.41

Mahalanobis Distance p-value: 0.98

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds						
Name	s-Triazine, 2,4-dichloro-6- (o-chloroanilino)-	Benzoic acid, 5- (chlorosulfonyl)-2,4- dichloro-	Aniline, 2,4-bis(o- methylphenoxy)-			
Structure	HN N N C C I MAN	CI S OH	H ₂ N ^N			
Actual Endpoint	Irritant	Irritant	Irritant			
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant			
Distance	0.627	0.701	0.705			
Reference	34ZIAG* -,235,69	FCTOD7 Food and Chemical Toxicology. (Pergamon Press Inc., Maxwell House, Fa irview Park, Elmsford, NY 10523) V.20- 1982- Volume(issue)/page/year: 20,563,1982	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986			

Model Applicability

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

Feature Contribution Top features for positive contribution							
				,			

FCFP_12	-1539132615	[*]C[c]1:s:[*]:[c]:1[*]	0.0795	9 out of 9		
FCFP_12	-124685461	s N N H H -0 [*]:n:[cH]:n:[*]	0.0734	5 out of 5		
FCFP_12	-475316933	[*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.0703	4 out of 4		
Top Features for negative contribution						

Top Features for negative contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	839741273	[*]:[c](:[*])N[c]1:[c H]:[cH]:[*]:[cH]:[cH	-0.708	4 out of 10		
FCFP_12	414792462	[*]O[c]1:[cH]:[cH]:[c](N[c](:[*]):[*]):[c H]:[cH]:1	-0.65	0 out of 1		

FCFP_12	-773983804	N⊸S T	-0.444	46 out of 79
		N N		
		_o [*]N[c]1:[cH]:[*]:[c]		
		([*]):[cH]:[cH]:1		

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4 Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.183 Enrichment: 0.199 Bayesian Score: -5.47

Mahalanobis Distance: 8.06

Mahalanobis Distance p-value: 0.867

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	1-Amino-2-bromo-4- hydroxyanthraquinone	1-Amino-4-hydroxy-5- chloroanthraquinone	Phenol, 4,4'-sulfonyldi-	
Structure	HO MNH 2	HO the NH 2	HO OI OI OH	
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant	
Distance	0.650	0.669	0.696	
Reference	28ZPAK -,83,72	28ZPAK -,83,72	BIOFX* BIOFAX Industrial Bio-Test Laboratories, Inc., Data Sheets. (1810 Fro ntage Rd., Northbrook, IL 60062) Volume(issue)/page/year: 601-05501,1974	

Model Applicability

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

Feature Contribution					
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	-1539132615	AND Enantiomer N N H H 2 [*]C[c]1:s:[*]:[*]:[c]:1[*]	0.0795	9 out of 9	

FCFP_12	-124685461	AND Enantiomer S N H H H 1:n:[cH]:n:[*]	0.0734	5 out of 5
FCFP_12	-475316933	AND Enantiomer N N H ₂ N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.0703	4 out of 4
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-970693500	AND Enantiomer N H N [*]N[c]1:[cH]:[cH]:[c](N):[cH]:[cH]:1	-1.04	0 out of 2
FCFP_12	-200702388	AND Enantiomer S N H (*):[c](:[*])N[c]1:[c H]:[cH]:[c](N):[cH]: [cH]:1	-0.846	1 out of 4
FCFP_12	839741273	[*]:[c](-[*]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH	-0.708	4 out of 10

 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 0.974
Enrichment: 1.06
Bayesian Score: -0.706

Mahalanobis Distance: 7.74

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	s-Triazine, 2,4-dichloro-6- (o-chloroanilino)-	Aniline, 2,4-bis(o- methylphenoxy)-	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	
Structure	HN N CI	H ₂ N ^N r	O THE WAY THE	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant	
Distance	0.626	0.630	0.668	
Reference	34ZIAG* -,235,69	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,242,1	

Model Applicability

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

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

FCFP_12	-1539132615	AND Enantiomer N N N H H (*]C[c]1:s:[*]:[c]]:1[*]	0.0795	9 out of 9
FCFP_12	907096426	AND Enantiomer NH NH [*]NC[c](:[*]):[*]	0.0772	7 out of 7
FCFP_12	427906732	AND Enantiomer N H H (*]NC[c]1:[cH]:[cH]:[*]:[cH]:[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	1294255210	AND Exantomer NH NH [*]CN[c](:[*]):[*]	-0.486	12 out of 22
FCFP_12	1618154665	AND Enantomer N N N N N N (H H) [*][c](:[*]):[cH]:[cH]:[*]	-0.0845	412 out of 490

FCFP_12	16	AND Enantiomer N N NH	-0.0843	423 out of 503
		[*][c](:[*]):[*]		

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Irritant

Probability: 0.966 Enrichment: 1.05 Bayesian Score: -1.27

Mahalanobis Distance: 8.34

Mahalanobis Distance p-value: 0.763

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Benzenesulfonic acid, 2- anilino-5-nitro-	8-Methylamino-4-hydroxy- 2-naphthalene sulfonic acid	Aniline, 2,2-dithiobis-	
Structure	O NH OH	HO WN H	H ₂ N _{th}	
Actual Endpoint	Irritant	Non-Irritant	Non-Irritant	
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant	
Distance	0.770	0.772	0.781	
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1061,1986	28ZPAK -,190,72	28ZPAK -,512,72	

Model Applicability

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

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

FCFP_12	-1539132615	AND Enantiomer N N N N N N N N H H N NH 2 S [*]C[c]1:s:[*]:[c]:1[*]	0.0795	9 out of 9
FCFP_12	-124685461	AND Enantiomer NH HN NH S NH [*]:n:[cH]:n:[*]	0.0734	5 out of 5
FCFP_12	-475316933	AND Enunitomer N N N N N N N N N N N N N N N 1 N N N 1 N N N 1 N	0.0703	4 out of 4

		atures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	AND Enantiomer S NH HN NH ₂ [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	16	AND Enantiomer S NH NH HN NH ₂ [*][c](:[*]):[*]	-0.0843	423 out of 503

FCFP_12	3	AND Enantiomer -0.1	.0812	291 out of 345
		s T		
		NH HN		
		S NH ₂		
		[*]N[*]		

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541
Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.866
Enrichment: 0.94
Bayesian Score: -3.08
Mahalanobis Distance: 7.74

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Benzenesulfonic acid, 2- anilino-5-nitro-	2-Anthracenesulfonic acid, 1-amino-9,10- dihydro-9,10-dioxo-4- (2,4,6 -trimethylanilino)-, monosodium salt	Urea, 1,3-bis(2- benzothiazolylthiomethyl)-	
Structure	O N N N N N N N N N N N N N N N N N N N	H ₂ N ₂ N ₂ N ₃ N ₄ N ₄ N ₄ N ₅	HN N	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant	
Distance	0.657	0.728	0.769	
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1061,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	AMIHBC AMA Archives of Industrial Hygiene and Occupational Medicine. (Chicago, IL) V.2-10, 1950-54. For publisher information, see AEHLAU. Volume(issue)/pag e/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.

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

FCFP_12	-1539132615	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.0795	9 out of 9
FCFP_12	-605671248	AND Enantomer N N N N N N N N N N N 1 N N 1 N 1 N 1	0.0772	7 out of 7
FCFP_12	-124685461	AND Enantiomer NH S-NH ₂ [*]:n:[cH]:n:[*]	0.0734	5 out of 5

Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1724769936	AND Enantiomer S N N NH ₂ [*]N[c]1:[cH]:[cH]:[c H]:[cH]:[c]:1[*]	-0.475	11 out of 20
FCFP_12	-773983804	AND Enantiomer N N N N N N N N N N N N N N N N (*) NH ₂ (*) N[c] 1: [cH]: [*]: [c] ([*]): [cH]: [cH]: 1	-0.444	46 out of 79

FCFP_12	590925877	AND Enantiomer	-0.434	56 out of 95
		N=\		
		NH ₂		
		[*]N[c](:[cH]:[*]):[c		
		H]:[*]		

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.663
Enrichment: 0.719
Bayesian Score: -4.07
Mahalanobis Distance: 7.74

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simi	lar Compounds		
Name	Benzenesulfonic acid, 2- anilino-5-nitro-	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	Urea, 1,3-bis(2- benzothiazolylthiomethyl)-
Structure	O NH OH	H ₂ N ₂ N ₄ N ₄ N ₄ N ₄ N ₄ N ₄ N ₅	HN HN N
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.676	0.741	0.774
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1061,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	AMIHBC AMA Archives of Industrial Hygiene and Occupational Medicine. (Chicago, IL) V.2-10, 1950-54. For publisher information, see AEHLAU. Volume(issue)/pag e/year: 5,311,1952

Model Applicability

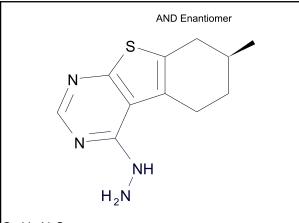
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

FCFP_12	-1539132615	AND Enantiomer N N N N H 1,N N (*]C[c]1:s:[*]:[*]:[c]:1[*]	0.0795	9 out of 9
FCFP_12	-124685461	AND Enantiomer N N N H N H 2:N (1:2) (1:1	0.0734	5 out of 5
FCFP_12	-475316933	AND Examtomer N N N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.0703	4 out of 4

Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	AND Enuntioneer N=N N=N N=N N=N N=N N=N N=N N=N N=N N	-0.708	4 out of 10
FCFP_12	-773983804	**No Enantioner** **No Enantion	-0.444	46 out of 79

FCFP_12	590925877	AND Enantiomer	-0.434	56 out of 95
		, N		
		N=(N		
		เ*]N[c](:[cH]:[*]):[c		
		H]:[*]		



 $C_{11}H_{14}N_{4}S$

Donors: 2

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Model Prediction

Prediction: Non-Irritant

Probability: 0.95
Enrichment: 1.03
Bayesian Score: -1.89
Mahalanobis Distance: 8.12

Mahalanobis Distance p-value: 0.846

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simil	Structural Similar Compounds				
Name	1-Naphthalenesulfonic acid, 2-hydroxy-	Phenol, 4,4'-sulfonyldi-	m-Toluene sulfonic acid, 6-amino-4-chloro-		
Structure	O O OH	HO OE S	HO O O O O O O O O O O O O O O O O O O		
Actual Endpoint	Irritant	Irritant	Non-Irritant		
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant		
Distance	0.559	0.566	0.585		
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1059,1986	BIOFX* BIOFAX Industrial Bio-Test Laboratories, Inc., Data Sheets. (1810 Fro ntage Rd., Northbrook, IL 60062) Volume(issue)/page/year: 601-05501,1974	28ZPAK -,526,72		

Model Applicability

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

Feature Co	Feature Contribution			
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1539132615	AND Enantomer S N NH H ₂ N [*]C[c]1:s:[*]:[c]]:1[*]	0.0795	9 out of 9

FCFP_12	-124685461	AND Enantiomer S NH H ₂ N [*]:n:[cH]:n:[*]	0.0734	5 out of 5
FCFP_12	-475316933	AND Enantiomer NH H ₂ N [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.0703	4 out of 4
	-	ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	AND Enantiomer S N NH H ₂ N [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	1070150408	AND Enantiomer S N NH H ₂ N [*]NN	-0.347	1 out of 2
FCFP_12	16	AND Enantiomer S NH H ₂ N [*][c](:[*]):[*]	-0.0843	423 out of 503

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.844
Enrichment: 0.916
Bayesian Score: -3.25
Mahalanobis Distance: 10

Mahalanobis Distance p-value: 0.0563

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Benzenesulfonic acid, 2,2'-(4,4'- biphenylylenedivinylene)d i-, disod ium salt	Anthraquinone, 1,1'- iminodi-	Urea, 1,3-bis(2-benzothiazolylthiomethyl)-	
Structure	OH CONTRACTOR OF THE CONTRACTO	T H H H H H H H H H H H H H H H H H H H	HN N	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant	
Distance	0.699	0.938	0.953	
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986	AMIHBC AMA Archives of Industrial Hygiene and Occupational Medicine. (Chicago, IL) V.2-10, 1950-54. For publisher information, see AEHLAU. Volume(issue)/pag e/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.

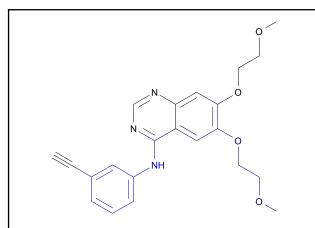
Feature Co	ntribution			
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
				•

FCFP_12	-1539132615	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.0795	9 out of 9
FCFP_12	-124685461	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.0734	5 out of 5
FCFP_12	-475316933	AND Enantiomer NNN NN NN NH S NH S [*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.0703	4 out of 4

Top Features for negative contribution			
Bit/Smiles	Feature Structure	Score	Irritant in training set
-200702388	AND Enantomer N N N N N N N N N N N N N N N N N N N	-0.846	1 out of 4
839741273	AND Enantiomer N S N N N N N N N N N N N N N N N N N	-0.708	4 out of 10
	-200702388	-200702388 AND Enantiomer S (*):[c](:[*))N[c]1:[c H]:[cH]:[c](N):[cH]: [cH]:1 AND Enantiomer AND Enantiomer (*):[c](:[*))N[c]1:[c H]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[c	-200702388 AND Enuntiomer [*]:[c](:[*])M[c]1:[c H]:[cH]:[c](N):[cH]: [cH]:1 AND Enuntiomer -0.846 -0.846 -0.846 -0.708

FCFP_12	-773983804	AND Enantiomer	-0.444	46 out of 79
		N N N		
		N NH		
		S & C		
		[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1		

Erlotinib



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.119
Enrichment: 0.13
Bayesian Score: -5.76
Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.0333

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Propanoic acid, 2-(4-((5- (trifluoromethyl)-2- pyridinyl)oxy)phenoxy) -, butyl ester	Carbamic acid, ((dibutylamino)thio)methy I-, 2,2-dimethyl-2,3-dihydr o-7-benzofuranyl ester	1,4-Pentadien-3-one, 1,5- bis(p-azidophenyl)-	
Structure				
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.776	0.777	0.803	
Reference	NNGADV Nippon Noyaku Gakkaishi. Journal of the Pesticide Science Society of Japan. (Nippon Noyaku Gakkai, 1-43-11, Komagome, Toshima-ku, Tokyo 170, Japan) V.1- 1976- Volume(issue)/page/year: 15,305,1990	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: OTS0539690	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,733,1986	

Model Applicability

Unknown features are fingerprint features in 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: 902193919: [*]:[c](:[*])C#C

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

FCFP_12	-124685461	[*]:n:[cH]:n:[*]	0.0734	5 out of 5
FCFP_12	-475316933	[*][c]1:[*]:[c](:[*]) :n:[cH]:n:1	0.0703	4 out of 4
FCFP_12	-1385265413	[*]:[cH]:[c]1:n:[cH]: n:[*]:[c]:1:[*]	0.0658	3 out of 3

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	411414971	[*][o]1:[cH]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-1.31	1 out of 7
FCFP_12	-1059904848	C):[cH]:[.] [.][c](:[.]):[c](occo	-1.04	0 out of 2

FCFP_12	839741273	o ʻ	-0.708	4 out of 10
		N O		
		[*]:[c](:[*])N[c]1:[c o		
		H]:[cH]:[*]:[cH]:[cH]:1		
		J		

AND Enantiomer S N NH CI

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 3.09

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 4.72e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Phenolphthalein	646	1-Amino-2,4- dibromoanthra-qui-none		
Structure	НО	OH OH	Br O		
Actual Endpoint (-log C)	2.43468	0.937339	2.90243		
Predicted Endpoint (-log C)	3.66084	3.26294	3.50402		
Distance	0.663	0.701	0.750		
Reference	CPDB	CPDB	CPDB		

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]
- 3. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 4. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	AND Enantiomer N N N H CI [*]:n:[*]	0.229

ECFP_6	-817402818	AND Enantiomer S N N N N N H	0.129
ECFP_6	-167460056	[*]CI AND Enantiomer S N N N (*) (*) (*) (*) (*) (*)	0.0596
		or negative contributio	
Fingerprint ECFP_6	Bit/Smiles 1996767644	Feature Structure	Score -0.251
	1000767671	AND Enantiomer S N N H CI [*][c](:[*]):[cH]:[cH]:[*]	
ECFP_6	642810091	AND Enantiomer S N N N H CI [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	AND Enantiomer S N N N CI [*]:[cH]:[*]	-0.232

AND Enantiomer S N NH F

 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 9.29

Unit: mg/kg_body_weight/day Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.00248

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Phenolphthalein	646	1-Amino-2,4- dibromoanthra-qui-none		
Structure	НО	OH OH	Br O		
Actual Endpoint (-log C)	2.43468	0.937339	2.90243		
Predicted Endpoint (-log C)	3.66084	3.26294	3.50402		
Distance	0.636	0.691	0.731		
Reference	CPDB	CPDB	CPDB		

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]
- 3. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 4. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 5. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):c:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	655739385	AND Enantiomer S NH NH [*]:n:[*]	0.229		

ECFP_6	-167460056	AND Enantiomer	0.0596
		NH	
		*	
ECFP_6	734603939	[*]C([*])[*] AND Enantiomer	0.0424
LOFF_0	734003939	N S	0.0424
		N= NH	
		F	
		[*]C	
	Top Features f	or negative contributio	on
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	AND Enantiomer	-0.251
		N N	
		N H F	
		[*][c](:[*]):[cH]:[cH	
ECFP_6	642810091]:[*]	-0.247
ECFF_0	042010091	AND Enantiomer	-0.247
		\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	
		F	
		[*][c](:[*]):[*]	
ECFP_6	-182236392	AND Enantiomer	-0.232
		NH NH	
		F	
		[*]:[cH]:[*]	

3.17837

0.702

CPDB

N NH NH

 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: 7.19

Unit: mg/kg_body_weight/day Mahalanobis Distance: 12.3

Mahalanobis Distance p-value: 4.73e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

	•		
Name	Phenolphthalein	646	C.I. pigment red 3
Structure	НО	OH OIN	N N N N N N N N N N N N N N N N N N N
Actual Endpoint (-log C)	2.43468	0.937339	0.937339

3.26294

0.648

CPDB

Model Applicability

Predicted Endpoint (-log

Distance

Reference

Structural Similar Compounds

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

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

3.66084

0.626

CPDB

- 2. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 3. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	655739385	S N N N N N N N N N N N N N N N N N N N	0.229		

ECFP_6	-176455838 734603939	[*]O[c](:[cH]:[*]):[c H]:[*]	0.0818
	Ton Features 6	for negative contributio	in and
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	[*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	S N=N H H -0 [*]:[cH]:[*]	-0.232

CPDB

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: 3.25

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11.4

Mahalanobis Distance p-value: 0.000343

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Phenolphthalein	646	C.I. pigment red 3
Structure	НО	OH OH	
Actual Endpoint (-log C)	2.43468	0.937339	0.937339
Predicted Endpoint (-log C)	3.66084	3.26294	3.17837
Distance	0.596	0.686	0.725

Model Applicability

Reference

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

CPDB

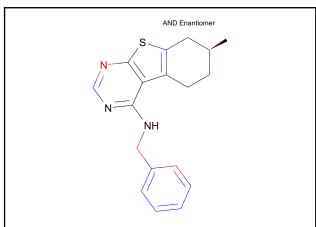
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]

CPDB

Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
 Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

reature Cont	Hibution				
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	655739385	AND Enantiomer S N N H ₂ N [*]:n:[*]	0.229		

ECFP_6	1572579716	AND Enantiomer S N N H ₂ N [*]N	0.225
ECFP_6	-167460056	AND Enantiomer S N N H ₂ N [*]C([*])[*]	0.0596
Fig. v. a. u.u. visa 4		or negative contributio	
Fingerprint ECFP_6	Bit/Smiles 1996767644	Feature Structure	Score -0.251
		AND Enantismer N N N H ₂ N [*][c](:[*]):[cH]:[cH]:[*]	
ECFP_6	642810091	AND Enantiomer N N N N H H 2N [*][C](:[*]):[*]	-0.247
ECFP_6	-182236392	AND Enantiomer S N N H ₂ N [*]:[cH]:[*]	-0.232



 $C_{18}H_{19}N_3S$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 4.81

Unit: mg/kg_body_weight/day Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.00592

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	Phenolphthalein	646	1077	
Structure	НО	OH OIL	NH ₂	
Actual Endpoint (-log C)	2.43468	0.937339	3.82962	
Predicted Endpoint (-log C)	3.66084	3.26294	3.933	
Distance	0.634	0.689	0.715	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]
- 3. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 4. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	AND Enantiomer NH NH [*]:n:[*]	0.229	
		[*]:n:[*]		

ECFP_6	1559650422	AND Enantiomer N N N N N N [*]C[*]	0.203
ECFP_6	-2024255407	AND Enantiomer S	0.172
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*][c](:[*]):[cH]:[cH	-0.251
ECFP_6	642810091	AND Enantioner NH NH [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	AND Enantiomer S NH NH [*]:[cH]:[*]	-0.232

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: 2.92

Unit: mg/kg_body_weight/day Mahalanobis Distance: 12.9

Mahalanobis Distance p-value: 1.8e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	2-Hydrazino-4-(p- nitrophenyl) thiazole	Furosemide	2-Hydrazino-4-(5-nitro-2-furyl) thi-azole	
Structure	N-NH ₂	H ₂ N ₂ S CI N ₁ O OH	H ₂ N NH	
Actual Endpoint (-log C)	4.34807	2.65498	4.13967	
Predicted Endpoint (-log C)	4.37858	3.60472	4.34552	
Distance	0.685	0.687	0.699	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]
- 3. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 4. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 5. Unknown ECFP_2 feature: -571028867: [*]NC(=S)N

Fingerprint Bit/Smiles Feature Structure Score ECFP_6 655739385 SH, NH, NH, NH, NH, SR, NH, S

ECFP_6 ECFP_6	-932108170	AND Enantiomer NH NH NH S [*]N AND Enantiomer	0.225
	Top Features f	[*]C(=[*])N	on .
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	AND Enantiomer S NH NH NH S NH S [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	AND Enantlomer S NH NH NNH NNH S NH S [*]:[cH]:[*]	-0.232
ECFP_6	-992506539	AND Enantiomer S NH NH HN NH ₂ [*]C[*]	-0.0852

N NH NH 2

C₁₇H₁₈N₄O₂S₂

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 8.21

Unit: mg/kg_body_weight/day Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 8.8e-008

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	542	Ochratoxin A	470		
Structure	AND Enantiomer AND Enantiomer HN HN HO HN HO HO HN HO HO HN HO HO	OH HO CI	OH OO O		
Actual Endpoint (-log C)	4.79932	4.79932	4.62839		
Predicted Endpoint (-log C)	3.6353	3.6353	3.93264		
Distance	0.728	0.728	0.764		
Reference	CPDB	CPDB	CPDB		

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]
- 3. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 4. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Foaturo	Contribution
reature	Contribution

Top features for positive contribution					
ECFP_6	655739385	AND Enantiomer NH Sinite NH [*]:n:[*]	0.229		

ECFP_6	1572579716	AND Enantiomer S NH O NH O (*)NH O (*)	0.225
ECFP_6	-167460056	AND Enantiomer S NH O NH O O [*]C([*])[*]	0.0596
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	AND Enantiomer N N N N N N N N N N N S NH ₂ [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	AND Enantiomer S NH O S NH ₂ [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	AND Enantiomer S NH O S NH ₂ O [*]:[cH]:[*]	-0.232

AND Enantiomer S N NH NH O=S H₂N O

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: 3.51

Unit: mg/kg_body_weight/day Mahalanobis Distance: 12.9

Mahalanobis Distance p-value: 1.99e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	542	Ochratoxin A	470		
Structure	AND Enantioner AND Enantioner HN HO	OH HO CI	OH ON		
Actual Endpoint (-log C)	4.79932	4.79932	4.62839		
Predicted Endpoint (-log C)	3.6353	3.6353	3.93264		
Distance	0.735	0.735	0.763		
Reference	CPDB	CPDB	CPDB		

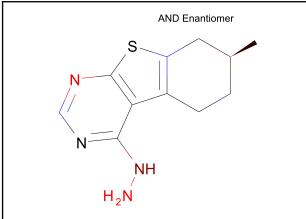
Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]
- Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
 Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

reature Continuution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score			
ECFP_6	655739385	AND Enantiomer S N N N N N N N N N N N N	0.229			
<u> </u>	L	I	I			

ECFP_6	1572579716	AND Enantiomer S N N N H O N H 2 N T N H N N T N N N N N N N N N N N N N	0.225
ECFP_6	-167460056	AND Enantiomer N N N N N H OI H ₂ N [*]C([*])[*]	0.0596
		for negative contributio	
Fingerprint ECFP_6	Bit/Smiles 1996767644	Feature Structure	Score -0.251
2011_0	1330707044	[*][c](:[*]):[cH]:[cH	0.201
ECFP_6	642810091	AND Enantiomer S N N N N H N (): H ₂ N [*][C](:[*]):[*]	-0.247
ECFP_6	-182236392	AND Enantiomer N N N N H N H 2 N H R R H R H R H R H R H R H R H R H R	-0.232



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: 2.64

Unit: mg/kg_body_weight/day Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.00657

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	Chrysazin	215	4,4´-Thiodianiline
Structure	но	H ₂ N ₁ N ₁ N ₁ N ₁ N ₁ CI	H ₂ N NH ₂
Actual Endpoint (-log C)	3.0774	3.477	3.81392
Predicted Endpoint (-log C)	3.07832	3.97642	3.72747
Distance	0.568	0.618	0.626
Reference	CPDB	CPDB	CPDB

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]
- 3. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 4. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	AND Enantiomer S NH H ₂ N [*]:n:[*]	0.229	

ECFP_6 ECFP_6	-934039951	AND Enantiomer S NH H ₂ N [*]N AND Enantiomer S NH H ₂ N	0.225
		[*]NN	
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	AND Enantiomer S N NH H ₂ N [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	AND Enantiomer S NH H ₂ N [*]:[cH]:[*]	-0.232
ECFP_6	-992506539	AND Enantiomer S NH H ₂ N [*]C[*]	-0.0852

TOPKAT_Carcinogenic_Potency_TD50_Mouse

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.169

Unit: mg/kg_body_weight/day Mahalanobis Distance: 13.8

Mahalanobis Distance p-value: 1.09e-009

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simila	r Compounds		
Name	223	Phenolphthalein	646
Structure	AND Enastronies	НО	OH OH
Actual Endpoint (-log C)	5.08368	2.43468	0.937339
Predicted Endpoint (-log C)	5.08273	3.66084	3.26294
Distance	1.084	1.139	1.147
Reference	CPDB	CPDB	CPDB

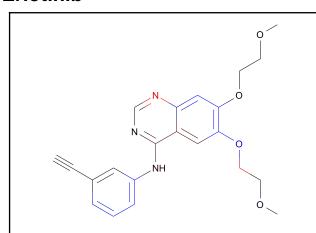
Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 292958156: [*]CC(C)C[*]
- 3. Unknown ECFP_2 feature: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 4. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]

Fingerprint	Bit/Smiles	miles Feature Structure		
ECFP_6	655739385	AND Enantomer N N N N N N N N N N N N N N N N N N	0.229	

ECFP_6	-167460056	AND Enantiomer S N N N N N N N N N N N N	0.0596
ECFP_6	734603939	[*]C([*])[*] AND Enantiomer N N N N N N N N N **	0.0424
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	AND Enantiomer N S N NH S [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.247
ECFP_6	-182236392	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.232



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: 39.8

Unit: mg/kg_body_weight/day Mahalanobis Distance: 16.1

Mahalanobis Distance p-value: 1.15e-016

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Simila	ar Compounds		
Name	Compound LY171883	832	5,5'-(1,1'-Biphenyl)-2,5- dyl-bis (oxy)(2,2- dimethylpentanoic acid)
Structure	DH OH	OH NH NH NH	OH HO
Actual Endpoint (-log C)	3.45372	3.45372	3.90166
Predicted Endpoint (-log C)	2.84749	2.80429	2.75893
Distance	0.772	0.782	0.796
Reference	CPDB	CPDB	CPDB

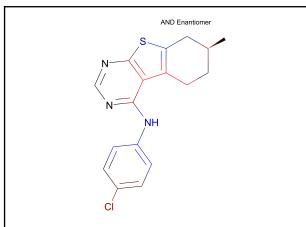
Model Applicability

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

- 1. OPS PC13 out of range. Value: -3.2209. Training min, max, SD, explained variance: -3.068, 3.6909, 1.329, 0.0220.
- 2. Unknown ECFP_2 feature: -182178874: [*]#C[c](:c:[*]):c:[*]
- 3. Unknown ECFP_2 feature: 1139738044: [*]:[c](:[*])C#C
- 4. Unknown ECFP_2 feature: -1253653003: [*]COC

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

ECFP_6	1559650422	[*]C[*]	0.203
ECFP_6	1333660716	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.0746
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*][c](:[*]):[cH]:[cH]:[cH]:[t*]	-0.251
ECFP_6	642810091	[*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	[*]:[cH]:[*]	-0.232



C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 3.75

Unit: mg/kg_body_weight/day Mahalanobis Distance: 18.1

Mahalanobis Distance p-value: 1.65e-021

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Indomethacin	45	Phenolphthalein		
Structure	OH OH	OH N	НО		
Actual Endpoint (-log C)	5.49293	3.92659	2.54766		
Predicted Endpoint (-log C)	4.9569	3.28325	3.7508		
Distance	0.621	0.646	0.651		
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.

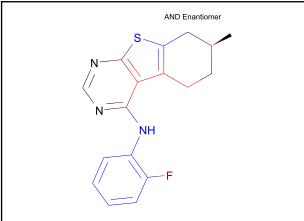
1. All properties and OPS components are within expected ranges.

Feature Cont	ribution			
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	32	AND Enantiomer S N N N H C I [*]CI	0.154	

FCFP_6	203677720	AND Enantiomer N N N N H CI [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0.137
FCFP_6	307419094	[*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1:	0.121
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	AND Enantomer N N N H CI [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	AND Enantlomer S S N H CI [*]N[c](:[cH]:[*]):[c H]:[*]	-0.323
FCFP_6	17	AND Enantiomer S N N N H	-0.149

0.630

CPDB



 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 10.5

Unit: mg/kg_body_weight/day Mahalanobis Distance: 18.6

Mahalanobis Distance p-value: 1.03e-023

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Indomethacin	45	Phenolphthalein		
Structure	OH OH	OH N N N	НО		
Actual Endpoint (-log C)	5.49293	3.92659	2.54766		
Predicted Endpoint (-log C)	4.9569	3.28325	3.7508		

Model Applicability

Distance

Reference

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

0.628

CPDB

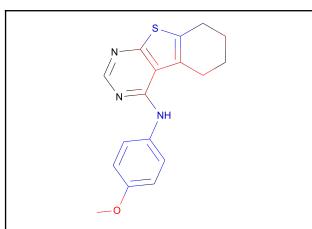
1. All properties and OPS components are within expected ranges.

0.627

CPDB

Feature Conti	ribution			
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	32	AND Enantiomer S N N N N F [*]CI	0.154	

FCFP_6	203677720	N	0.137
FCFP_6	307419094	N	0.121
		egative contribution	
Fingerprint		Feature Structure	Score
FCFP_6	991735244	AND Enandomer N N N N N H F [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422
FCFP_6	16	AND Enantiomer NH NH [*][C](:[*]):[*]	-0.354
FCFP_6	590925877	AND Enantiomer N S N N N N H F [*]N[c](:[cH]:[*]):[c H]:[*]	-0.323



 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4 Donors: 1

Model Prediction

Prediction: 1.42

Unit: mg/kg_body_weight/day Mahalanobis Distance: 16.7

Mahalanobis Distance p-value: 1.26e-016

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Indomethacin	Phenolphthalein	646		
Structure	OH OH	НО	OH OH		
Actual Endpoint (-log C)	5.49293	2.54766	2.41938		
Predicted Endpoint (-log C)	4.9569	3.7508	3.77987		
Distance	0.556	0.608	0.627		
Reference	CPDB	CPDB	CPDB		

Model Applicability

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

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

Feature Cont	ribution			
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	136627117	s in the second	0.69	

FCFP_6	1	N N N N N N N N N N N N N N N N N N N	0.234
FCFP_6	203677720	[*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0.137
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	[*][c](:[*]):[*]	-0.354
FCFP_6	590925877	[*]N[c](:[cH]:[*]):[c H]:[*]	-0.323
FCFP_6	1674451008	[*]O[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.233

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: 2.42

Unit: mg/kg_body_weight/day Mahalanobis Distance: 16.1

Mahalanobis Distance p-value: 1.25e-014

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Phenolphthalein	Indomethacin	646		
Structure	НО		-CI OH		

	но	OH OH	, o 2
Actual Endpoint (-log C)	2.54766	5.49293	2.41938
Predicted Endpoint (-log C)	3.7508	4.9569	3.77987
Distance	0.570	0.664	0.682
Reference	CPDB	CPDB	CPDB

Model Applicability

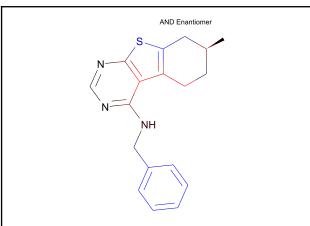
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_6	-1247245120	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.375			

FCFP_6	203677720	AND Enantiomer N N N N H H 2N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0.137
FCFP_6	307419094	AND Enantiomer S N N N N N N (*][c](^t , t, t)[c](1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.121
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	AND Enantiomer N N N N H H 2N [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	AND Enantiomer N N N N H N N N N N N N	-0.323
FCFP_6	17	AND Enantiomer S N N N H H S I I I I I I I I I I I I	-0.149

CPDB



 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 29.9

Unit: mg/kg_body_weight/day Mahalanobis Distance: 17.2

Mahalanobis Distance p-value: 3.34e-018

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds						
Phenolphthalein	Nafenopin s	Indomethacin				
НО	OH	OH OH				
2.54766	4.45051	5.49293				
3.7508	3.8403	4.9569				
0.613	0.617	0.622				
	Phenolphthalein	Phenolphthalein Nafenopin s				

Model Applicability

Reference

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

CPDB

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

CPDB

Feature Cont	Feature Contribution					
	Top features	for positive contributio	n			
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_6	203677720	AND Enantiomer N N H H S 1 (*)C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0.137			

FCFP_6	307419094	[*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1:	0.121
FCFP_6	-1272798659	AND Enantiomer S N NH NH [*]CCC([*])[*]	0.11
		or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	**************************************	-0.422
FCFP_6	-2093839777	**ND Enantomer **ND Enantomer	-0.378
FCFP_6	16	AND Enantiomer NH NH [*][c](:[*]):[*]	-0.354

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4 Donors: 3

Model Prediction

Prediction: 10

Unit: mg/kg_body_weight/day Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 1.05e-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 prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	1,2-Dihydro-2-(5-nitro-2- thi-enyl) quinazolin-4(3H)- one	2-Hydrazino-4-(p- nitrophenyl) thiazole	Purpurin		
Structure	O-NO HNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	N N NH 2 H	HO OH O		
Actual Endpoint (-log C)	5.25509	4.86687	2.57737		
Predicted Endpoint (-log C)	3.89291	5.14723	3.49183		
Distance	0.622	0.670	0.681		
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				
FCFP_6	1	AND Enantiomer S NH NH HN NH ₂ [*]O[*]	0.234	

FCFP_6	203677720	AND Enantiomer N N N H H N NH ₂ S [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0.137
FCFP_6	307419094	AND Enantiomer N N N N H H N N N N H H N N 1 ([*]):[*]:[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.121
-		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	AND Enantiomer NH HN NH S NH S [*][C](:[*]):[*]	-0.354
FCFP_6	17	AND Enantiomer S NH HN NH S NH ₂ [*]:S:[*]	-0.149
FCFP_6	0	AND Enantiomer S NH NH HN S NH S [*]C[*]	-0.115

F

┖

ᆮ

TOPKAT_Carcinogenic_Potency_TD50_Rat

N NH NH 2

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: 29.3

Unit: mg/kg_body_weight/day Mahalanobis Distance: 17

Mahalanobis Distance p-value: 1.11e-017

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

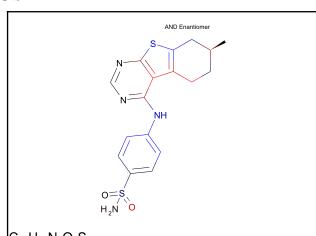
Structural Similar Compounds				
Name	4-(2-Hydroxyethylamino)- 2-(5-nitro-2- thienyl)quinazoline	1,2-Dihydro-2-(5-nitro-2- thi-enyl) quinazolin-4(3H)- one	542	
Structure	N N N N N N N N N N N N N N N N N N N	-O-N HN HN O	AND Enantomer AND Enantomer HN HO O	
Actual Endpoint (-log C)	5.22831	5.25509	6.59334	
Predicted Endpoint (-log C)	4.31976	3.89291	5.06501	
Distance	0.622	0.662	0.693	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

Feature Contr	ibution			
	Top features	for positive contributio	า	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	AND Enantiomer S NH S NH S NH S NH [*]O[*]	0.234	

	203677720	[*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0.137
FCFP_6	307419094	[*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.121
-	Top Features for ne		
Fingerprint		Feature Structure	Score
FCFP_6	991735244	AND Enantiomer S N N N N N N N N N N N N N N 1 (*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422
FCFP_6	16	AND Enantiomer NH NH NH S NH ₂ [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	AND Enantiomer S N N N N N N N N N N N N N N N N N N	-0.323



C₁₇H₁₈N₄O₂S₂

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: 10.7

Unit: mg/kg_body_weight/day Mahalanobis Distance: 16.8

Mahalanobis Distance p-value: 4.07e-017

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

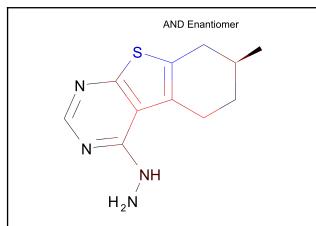
Structural Similar Compounds					
Name	4-(2-Hydroxyethylamino)- 2-(5-nitro-2- thienyl)quinazoline	1,2-Dihydro-2-(5-nitro-2- thi-enyl) quinazolin-4(3H)- one	Omeprazole		
Structure	N N N N N N N N N N N N N N N N N N N	-O-N HN N	O N N N N N N N N N N N N N N N N N N N		
Actual Endpoint (-log C)	5.22831	5.25509	3.4628		
Predicted Endpoint (-log C)	4.31976	3.89291	4.7324		
Distance	0.626	0.678	0.701		
Reference	CPDB	CPDB	CPDB		

Model Applicability

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

Feature Contr	ribution			
	Top features	for positive contribution	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	AND Enantiomer S N N N N H O H 2 N T T T T T T T T T T T T T T T T T T	0.234	

FCFP_6	203677720	AND Enantomer N N N N N N H,N N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0.137
FCFP_6	307419094	**No Enantiomer	0.121
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	AND Enantiomer N N N N H H 2N [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	AND Enantioner N=	-0.323
FCFP_6	17	AND Enantiomer N N N N H H N N H	-0.149



 $C_{11}H_{14}N_{4}S$

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: 9.15

Unit: mg/kg_body_weight/day Mahalanobis Distance: 16

Mahalanobis Distance p-value: 2.28e-014

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Chrysazin	215	1,2-Dihydro-2-(5-nitro-2- thi-enyl) quinazolin-4(3H)- one		
Structure	но	H ₂ N ₄ , N ₂ N ₂ N ₁ CI	S H N N N N N N N N N N N N N N N N N N		
Actual Endpoint (-log C)	2.99143	2.84742	5.25509		
Predicted Endpoint (-log C)	3.29868	3.32496	3.89291		
Distance	0.552	0.594	0.604		
Reference	CPDB	CPDB	CPDB		

Model Applicability

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

Feature Contr	Feature Contribution				
	Top features	for positive contributio	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_6	203677720	AND Enantiomer S NH H ₂ N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	0.137		

FCFP_6	307419094	AND Enantiomer S N H ₂ NH [*][c](:[*]):[c]1:[c] ([*]):[*]:[*]:[c]:1: [*]	0.121
FCFP_6	-1272798659	AND Enantiomer S NH H ₂ N [*]CCC([*])[*]	0.11
Fingerprint	Top Features f Bit/Smiles	or negative contributio Feature Structure	on Score
FCFP_6	16	AND Enantiomer	-0.354
		NH H ₂ N [*][c](:[*]):[*]	
FCFP_6	17	AND Enantiomer S NH H ₂ N [*]:s:[*]	-0.149
FCFP_6	0	AND Enantiomer S N NH H ₂ N [*]C[*]	-0.115

TOPKAT_Carcinogenic_Potency_TD50_Rat

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6 Donors: 2

C₂₈H₂₈N₆S₂

Model Prediction

Prediction: 0.347

Unit: mg/kg_body_weight/day Mahalanobis Distance: 18.2

Mahalanobis Distance p-value: 4.61e-022

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

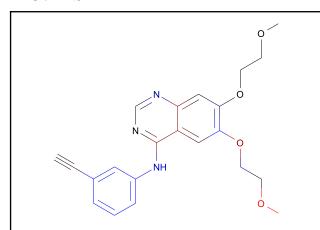
Structural Simila	r Compounds		
Name	223	411	Fluvastatin
Structure	AND Enerthoner	Na Na	HO HO
Actual Endpoint (-log C)	6.29867	3.06566	3.51742
Predicted Endpoint (-log C)	7.5657	4.8672	5.41573
Distance	1.050	1.054	1.066
Reference	CPDB	CPDB	CPDB

Model Applicability

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

Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score									
						FCFP_6	203677720	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.137

FCFP_6	307419094	[*][c](:[*]);[c]1:[c] ([*]):[*]:[*]:[c]:1:	0.121
FCFP_6	-1272798659	AND Enantiomer N S N N N N N N N N N N N N N N N N N	0.11
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	AND Enantomer N N N N N N N N N N N N N N N N N N N	-0.354
FCFP_6	590925877	AND Enantomer N S N N N N N N N N N N N N N	-0.323
FCFP_6	17	AND Enantiomer N S N NH S [*]:S:[*]	-0.149



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: 8.06

Unit: mg/kg_body_weight/day Mahalanobis Distance: 17.7

Mahalanobis Distance p-value: 3.99e-020

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

Structural Simila	ar Compounds		
Name	Loxtidine	C.I. direct brown 95	Omeprazole
Structure	NH NH N NH		
Actual Endpoint (-log C)	2.87532	5.31387	3.4628
Predicted Endpoint (-log C)	3.63996	4.30266	4.7324
Distance	0.685	0.715	0.741
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: 902193919: [*]:[c](:[*])C#C

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

FCFP_6	1	[*]O[*]	0.234
FCFP_6	203677720	[*]C[c]:1:[*]	0.137
		gative contribution	
•		Feature Structure	Score
FCFP_6	991735244	[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.422
FCFP_6	16	[*][c](:[*]):[*]	-0.354
FCFP_6	590925877	[*]N[c](:[cH]:[*]):[c H]:[*]	-0.323

0.615

NTP REPORT # 226

3.36361

AND Enantiomer N NH CI

 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.0375

Unit: g/kg_body_weight Mahalanobis Distance: 33.7

Mahalanobis Distance p-value: 2.01e-031

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	BROTIZOLAM	TRIAZOLAM	C.I. SOLVENT YELLOW 14
Structure	Br CI	N CI N N	N N N N N N N N N N N N N N N N N N N

3.83659

3.85527

UPJ-33030

0.593

Model Applicability

Actual Endpoint (-log C)

Predicted Endpoint (-log

Distance

Reference

Structural Similar Compounds

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

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

ARZNEI.FORSCH.36.592.

2. Unknown ECFP_6 feature: 914325265: [*]:s:[*]

2.99309

3.70649

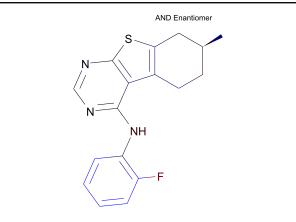
0.589

1986

- 3. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 4. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 5. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 6. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 7. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 8. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 9. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 10. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 11. Unknown ECFP_6 feature: 865482986: [*]C([*])C
- 12. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 13. Unknown ECFP 6 feature: -710237522: [*]:n:[cH]:n:[*]
- 14. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 15. Unknown ECFP 6 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
- 17. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](CI):[cH]:[*]
- 18. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

	ribution Top foatures	for positive contribution	n
Cinggrupint	Bit/Smiles	Feature Structure	
Fingerprint			Score
ECFP_6	-167460056	AND Enantiomer S N N N H H	0.136
		c í [*]C([*])[*]	
FCFP_6	32	AND Enantiomer S N N N N N N N N N N N N	0.101
		ci [*]CI	
FCFP_6	3	AND Enantiomer S N N N H CI	0.0924
		[*]N[*]	
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136597326	AND Enantiomer S N N N H CI [*]C([*])C	-0.0815

FCFP_6	203677720	AND Enantiomer S N N N N H CI [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0713
FCFP_6	-773983804	AND Enantiomer S N N H CI [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.0589



C₁₇H₁₆FN₃S

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.0746
Unit: g/kg_body_weight
Mahalanobis Distance: 33.4

Mahalanobis Distance p-value: 9.75e-031

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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

Name	BROTIZOLAM	TRIAZOLAM	C.I. SOLVENT YELLOW 14	
Structure	N N N N N N N N N N N N N N N N N N N	CIW N	OH N	
Actual Endpoint (-log C)	2.99309	3.83659	4.298	
Predicted Endpoint (-log C)	3.70649	3.85527	3.36361	
Distance	0.577	0.582	0.586	
Reference	ARZNEI.FORSCH.36.592. 1986	UPJ-33030	NTP REPORT # 226	

Model Applicability

Unknown features are fingerprint features in 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: 914325265: [*]:s:[*]
- 3. Unknown ECFP_6 feature: -1046436026: [*]F
- 4. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 5. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 6. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 7. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 8. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 9. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 10. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 11. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 12. Unknown ECFP_6 feature: 865482986: [*]C([*])C
- 13. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 14. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 15. Unknown ECFP 6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]
- 17. Unknown ECFP_6 feature: 1335108269: [*]N[c](:[cH]:[*]):[c]([*]):[*]
- 18. Unknown ECFP_6 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]
- 19. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

ECFP_6	1564392544	AND Enantiomer S N N H F [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133
FCFP_6	136597326	AND Enantiomer S N NH NH F [*]C([*])C	-0.0815

ARZNEI.FORSCH.36.592.

0.619

1986

N NH NH

 $C_{17}H_{17}N_3OS$

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: 0.0486

Unit: g/kg_body_weight Mahalanobis Distance: 31.8

Mahalanobis Distance p-value: 7.04e-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 prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	C.I. PIGMENT RED 3	C.I. SOLVENT YELLOW 14	BROTIZOLAM
Structure	ON OH ON ON O	OH N	Br N N
Actual Endpoint (-log C)	3.0252	4.298	2.99309

3.36361

0.614

NTP REPORT # 226

Model Applicability

Predicted Endpoint (-log

Distance

Reference

Structural Similar Compounds

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

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

NTP REPORT # 407

2. Unknown ECFP_6 feature: 914325265: [*]:s:[*]

3.34768

0.548

- 3. Unknown ECFP_6 feature: -1332781180: [*]CCC[*]
- 4. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 5. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[*]:[c]:1[*]
- 6. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 7. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 8. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 9. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 10. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 12. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 13. Unknown ECFP_6 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]
- 14. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
- 15. Unknown ECFP 6 feature: 1307307440: [*]:[c](:[*])OC

Feature Contribution

Top features for positive contribution

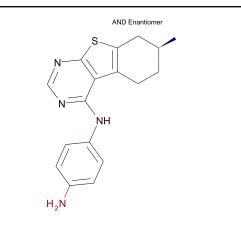
-176455838 3	[*]O[c](:[cH]:[*]):[c H]:[*]	0.106
3	s Y	0.0924
	N=N+ N=N+ -0 [*]N[*]	
1036089772	[*]:[o](:[*])OC	0.073
		1036089772

Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	N N N N N N N N N N N N N N N N N N N	-0.102	

FCFP_6	203677720	[*]C[c]1:[c]:(1:[*] :[*]:[c]:1:[*]	-0.0713
FCFP_6	-773983804	[*]N[c]1:[cH]:[c] ([*]):[cH]:[cH]:1	-0.0589

0.653

NTP REPORT # 222



 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: 0.063

Unit: g/kg_body_weight
Mahalanobis Distance: 33.1

Mahalanobis Distance p-value: 3.04e-030

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Sillillar Compounds			
Name	PYRIMETHAMINE	C.I. PIGMENT RED 3	HC BLUE 1
Structure	H ₂ NW	The state of the s	OH NH
Actual Endpoint (-log C)	3.99776	3.0252	3.0323

Model Applicability

Predicted Endpoint (-log

Distance

Reference

Structural Similar Compounds

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

3.34768

NTP REPORT # 407

0.621

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_6 feature: 914325265: [*]:s:[*]

4.25218

NTP 77 55

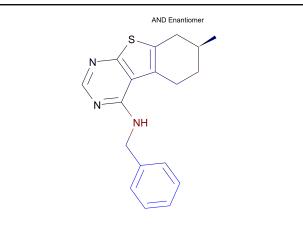
0.619

- 3. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 4. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 5. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 6. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[c]:1:[*]
- 7. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 8. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 9. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 10. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 11. Unknown ECFP 6 feature: 865482986: [*]C([*])C
- 12. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 13. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 14. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
- 17. Unknown ECFP_6 feature: 971820502: [*]:[cH]:[c](N):[cH]:[*]
- 18. Unknown ECFP_6 feature: -938530932: [*]:[c](:[*])N

Feature Contribution

Fingerprint	Bit/Smiles	for positive contribution Feature Structure	Score
ECFP_6	-167460056	AND Enantiomer S N N N H H ₂ N	0.136
		[*]C([*])[*]	
FCFP_6	3	AND Enuntiomer N N N N N H	0.0924
		H ₂ N [*]N[*]	
FCFP_6	1069584379	AND Enantiomer S N N N H H 2N	0.0717
		[*]:[c](:[*])N	
	Top Features	for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136597326	AND Enantiomer S N N N H ₂ N [*]C([*])C	-0.0815

FCFP_6	203677720	AND Enantiomer S N N N H H 2 (*]C[c]1:[c]((*]):[*] :[*]:[c]:1:[*]	-0.0713
FCFP_6	-773983804	AND Enantiomer S H ₂ N [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.0589



 $C_{18}H_{19}N_3S$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.0709

Unit: g/kg_body_weight Mahalanobis Distance: 34.6

Mahalanobis Distance p-value: 5.26e-033

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	C.I. SOLVENT YELLOW 14	BROTIZOLAM	TRIAZOLAM	
Structure	OH N N N	N N N N N N N N N N N N N N N N N N N	CIW N	
Actual Endpoint (-log C)	4.298	2.99309	3.83659	
Predicted Endpoint (-log C)	3.36361	3.70649	3.85527	
Distance	0.589	0.597	0.605	
Reference	NTP REPORT # 226	ARZNEI.FORSCH.36.592. 1986	UPJ-33030	

Model Applicability

Unknown features are fingerprint features in 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: 914325265: [*]:s:[*]
- 3. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 4. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 5. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 6. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 7. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 8. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[c](:[*]):s:1
- 9. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 10. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 11. Unknown ECFP_6 feature: 865482986: [*]C([*])C
- 12. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 13. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 14. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: 491100606: [*]CN[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: 769925792: [*]NC[c](:[*]):[*]
- 17. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

Feature Contribution

Top features for positive contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-167460056	AND Enantiomer N N N N N N N (*) (*) (*) (*) (*)	0.136
ECFP_6	1559650422	AND Enantiomer N N N (*]C[*]	0.129
FCFP_6	3	AND Enantiomer N NH NH [*]N[*]	0.0924
		for negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	AND Enantiomer N N H H (*)[c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134

ECFP_6	1564392544	AND Enardiomer S N N H H (*)[c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133
FCFP_6	-1698724694	AND Enantiomer N N N H H C And Enantiomer C C	-0.0944

NTP 77 55

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4 Donors: 3

Model Prediction

Prediction: 0.122

Unit: g/kg_body_weight
Mahalanobis Distance: 31.9

Mahalanobis Distance p-value: 5.05e-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 prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	FUROSEMIDE	DAPSONE	PYRIMETHAMINE
Structure	HO NH2 NH2 CI	H ₂ N O S NH ₂	NH 2 N N N N N N N N N N N N N N N N N N N
Actual Endpoint (-log C)	4.27645	3.6168	3.99776
Predicted Endpoint (-log C)	4.40005	3.43657	4.25218

Model Applicability

Distance

Reference

Churchinal Cimilar Campainale

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

0.702

NTP 20 47

- 1. OPS PC31 out of range. Value: 5.3583. Training min, max, SD, explained variance: -4.0208, 4.2529, 1.246, 0.0068.
- 2. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]

NTP REPORT #356

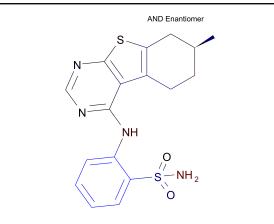
3. Unknown ECFP_6 feature: 914325265: [*]:s:[*]

0.623

- 4. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 5. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 6. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 7. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 8. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 9. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 10. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 11. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 12. Unknown ECFP_6 feature: 865482986: [*]C([*])C
- 13. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 14. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 15. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -1238415266: [*]NN[c](:[*]):[*]
- 17. Unknown ECFP_6 feature: 1635339976: [*]NNC(=[*])[*]
- 18. Unknown ECFP_6 feature: -571028867: [*]NC(=S)N
- 19. Unknown ECFP_6 feature: -932108170: [*]C(=[*])N

	Top features	for positive contribution	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	-167460056	AND Enantiomer	0.136	
		" NH HN NH S NH S [*]C([*])[*]		
ECFP_6	-845108448	AND Enantiomer	0.105	
		NH NH NH NH ₂		
		[*]=S		
FCFP_6	3	AND Enantiomer S NH NH HN NH ₂ S	0.0924	
		[*]N[*]		
		for negative contribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	AND Enantiomer S N NH NH NH NH ₂ S	-0.102	
		[*]0[*]		

FCFP_6	136597326	AND Enantiomer S N NH HN S NH ₂ [*]C([*])C	-0.0815
FCFP_6	203677720	AND Enantlomer N N N N N N N N N N N N N N N N N N	-0.0713



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.323

Unit: g/kg_body_weight Mahalanobis Distance: 35.3

Mahalanobis Distance p-value: 3.24e-034

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	D & C RED 9	CHLORSULFURON	FUROSEMIDE
Structure	ON THE STATE OF TH	HN N O	HO O NH2
Actual Endpoint (-log C)	3.87715	4.15566	4.27645
Predicted Endpoint (-log C)	3.6546	3.79771	4.40005
Distance	0.600	0.628	0.669
Reference	NTP REPORT # 225	EPA COVER SHEET 0027;880301;(1)	NTP REPORT # 356

Model Applicability

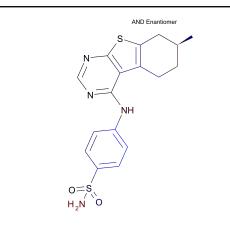
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_6 feature: 914325265: [*]:s:[*]
- 3. Unknown ECFP_6 feature: -797085356: [*]S(=[*])(=[*])[*]
- 4. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 5. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 6. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 7. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 8. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 9. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[c](:[*]):s:1
- 10. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 11. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 12. Unknown ECFP_6 feature: 865482986: [*]C([*])C
- 13. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 14. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 15. Unknown ECFP 6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]
- 17. Unknown ECFP_6 feature: 1335108269: [*]N[c](:[cH]:[*]):[c]([*]):[*]
- 18. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[t]
- 19. Unknown ECFP_6 feature: 1755253401: [*][c](:[cH]:[*])S(=[*])(=[*])[*]

- Unknown ECFP_6 feature: -2121766239: [*]:[c](:[*])S(=O)(=O)N Unknown ECFP_6 feature: 2102150379: [*]S(=[*])(=O)[*] Unknown ECFP_6 feature: -934226723: [*]S(=[*])(=[*])N 20.
- 21.
- 22.

Feature Contribu	Feature Contribution				
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	-167460056	AND Enantiomer S NH O S NH O [*]C([*])[*]	0.136		
FCFP_6	3	AND Enantiomer S NH NH S NH S NH S NH S NH S NH S NH	0.0924		
ECFP_6	1572579716	AND Enantiomer S NH NH S NH O [*]N	0.0576		
		egative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score		

FCFP_6	991735244	AND Enantiomer N NH2 NH2 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134
ECFP_6	1564392544	AND Enantiomer S N N N N N N N N N N N N N N N 1 C N N N 1 C N N 1 C N N 1 C N N 1 C N 1	-0.133
FCFP_6	1	AND Enantiomer S NH NH S NH ₂ [*]O[*]	-0.102



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.177

Unit: g/kg_body_weight
Mahalanobis Distance: 35.6

Mahalanobis Distance p-value: 1.1e-034

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	D & C RED 9	CHLORSULFURON	FUROSEMIDE	
Structure	NT _N N _T	HN HN N O	HO O NH ₂	
Actual Endpoint (-log C)	3.87715	4.15566	4.27645	
Predicted Endpoint (-log C)	3.6546	3.79771	4.40005	
Distance	0.609	0.641	0.680	
Reference	NTP REPORT # 225	EPA COVER SHEET 0027;880301;(1)	NTP REPORT # 356	

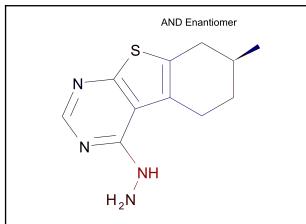
Model Applicability

- 1. OPS PC8 out of range. Value: -5.7967. Training min, max, SD, explained variance: -5.7428, 7.3359, 2.68, 0.0314.
- 2. Unknown ECFP_6 feature: 914325265: [*]:s:[*]
- 3. Unknown ECFP_6 feature: -797085356: [*]S(=[*])(=[*])[*]
- 4. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 5. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 6. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 7. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 8. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 9. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[c](:[*]):s:1
- 10. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 11. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 12. Unknown ECFP_6 feature: 865482986: [*]C([*])C
- 13. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 14. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 15. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]
- 17. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
- 18. Unknown ECFP_6 feature: -177264675: [*]S(=[*])(=[*])[c](:[cH]:[*]):[cH]:[*]

- Unknown ECFP_6 feature: -2121766239: [*]:[c](:[*])S(=O)(=O)N Unknown ECFP_6 feature: 2102150379: [*]S(=[*])(=O)[*] Unknown ECFP_6 feature: -934226723: [*]S(=[*])(=[*])N 19.
- 20.
- 21.

Feature Contribut	Feature Contribution				
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score		
ECFP_6	-167460056	AND Enantlomer N N N N N H H C! H Z! C([*])[*]	0.136		
FCFP_6	3	AND Enantiomer N N N H OI H Z N [*]N[*]	0.0924		
ECFP_6	1572579716	AND Enantiomer S N N N N N H H O H N [*] N	0.0576		
		egative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score		

FCFP_6	1	AND Enantiomer S N N N H H O'S H ₂ N [*]O[*]	-0.102
FCFP_6	-453677277	AND Enantiomer N N N N N H N N (N) N (N) (N) (N) (N) (N) (N) (N) (-0.0906
FCFP_6	136597326	AND Enantiomer N N N N N H 2N [*]C([*])C	-0.0815



 $C_{11}H_{14}N_4S$

Molecular Weight: 234.32065

ALogP: 2.561
Rotatable Bonds: 1
Acceptors: 4

Donors: 2

Model Prediction

Prediction: 0.129

Unit: g/kg_body_weight
Mahalanobis Distance: 36.4

Mahalanobis Distance p-value: 4.74e-036

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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

Name	PYRIMETHAMINE	4;4'-THIODIANILINE	DAPSONE
Structure	H ₂ N ^N	H ₂ N NH ₂	H ₂ N O S NH ₂
Actual Endpoint (-log C)	3.99776	3.46	3.6168
Predicted Endpoint (-log C)	4.25218	3.46805	3.43657
Distance	0.465	0.532	0.559
Reference	NTP 77 55	NTP 47 VIII	NTP 20 47

Model Applicability

- 1. OPS PC31 out of range. Value: 4.3094. Training min, max, SD, explained variance: -4.0208, 4.2529, 1.246, 0.0068.
- 2. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: 1070150408: [*]NN
- 4. Unknown ECFP_6 feature: 914325265: [*]:s:[*]
- 5. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 6. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 7. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 8. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 9. Unknown ECFP 6 feature: 51876938: [*]CC[c](:[*]):[*]
- 10. Unknown ECFP 6 feature: 85262808: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 11. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 12. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 13. Unknown ECFP_6 feature: 865482986: [*]C([*])C
- 14. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 16. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 17. Unknown ECFP_6 feature: -136863293: [*]:[c](:[*])NN
- 18. Unknown ECFP_6 feature: -934039951: [*]NN

Feature Cont		for positive contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-167460056	AND Enantiomer S N NH H ₂ N	0.136
		[*]C([*])[*]	
FCFP_6	3	AND Enantiomer	0.0924
		NH H ₂ N	
		[*]N[*]	
ECFP_6	1572579716	AND Enantiomer S N NH H ₂ N	0.0576
		[*]N	
	Top Features	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136597326	AND Enantiomer S N NH H ₂ N	-0.0815
		[*]C([*])C	

FCFP_6	203677720	NH H ₂ N	-0.0713
		[*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	
ECFP_6	-1331450522	AND Enantiomer S N NH H ₂ N	-0.0538
		[*]CCC([*])[*]	

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.0719

Unit: g/kg_body_weight Mahalanobis Distance: 33.8

Mahalanobis Distance p-value: 1.77e-031

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	C.I.PIGMENT RED 23	DIARYLANILIDE YELLOW	C.I. ACID RED 14
Structure	ON ON O		HO Sharm OH
Actual Endpoint (-log C)	2.28997	2.70208	2.8654
Predicted Endpoint (-log C)	3.52921	3.76154	3.29295
Distance	0.862	0.898	0.923
Reference	NTP 411 146	NTP 30 C-4	NTP REPORT # 220

Model Applicability

Unknown features are fingerprint features in 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: 914325265: [*]:s:[*]
- 3. Unknown ECFP_6 feature: 292958156: [*]CC(C)C[*]
- 4. Unknown ECFP_6 feature: 53207596: [*]C([*])C[c](:[*]):[*]
- 5. Unknown ECFP_6 feature: -1672647522: [*]C[c]1:s:[*]:[c]:1[*]
- 6. Unknown ECFP_6 feature: -1660340418: [*]C[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 7. Unknown ECFP_6 feature: 51876938: [*]CC[c](:[*]):[*]
- 8. Unknown ECFP_6 feature: 85262808: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 9. Unknown ECFP_6 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 10. Unknown ECFP_6 feature: -1661653144: [*][c](:[*]):[c]1:[c]([*]):[*]:[*]:[c]:1:[*]
- 11. Unknown ECFP 6 feature: 865482986: [*]C([*])C
- 12. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 13. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 14. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]

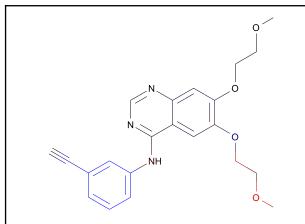
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	-167460056	AND Enantiomer S N N N N N N N N N N N N N N N N N N	0.136	
FCFP_6	3	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.0924	
ECFP_6	-992506539	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.0554	
	Top Features for negative contribution			
Fingerprint	Rit/Smiles	Feature Structure	Score	

	Top Features for negative contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	136597326	AND Enantomer N N N N N N N N N N N N N N N N N N	-0.0815	

FCFP_6	203677720	AND Enantiomer N S N N N N N N N N N N N N N N N N N	-0.0713
FCFP_6	-773983804	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.0589



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0359

Unit: g/kg_body_weight Mahalanobis Distance: 34.8

Mahalanobis Distance p-value: 2.91e-033

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

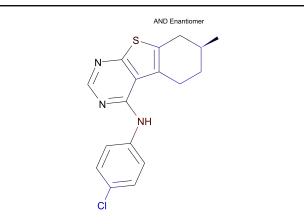
Structural Similar Compounds					
Name	ASSURE	DILTIAZEM	RHODAMINE 6G		
Structure	Clan				
Actual Endpoint (-log C)	5.00328	4.21961	4.54906		
Predicted Endpoint (-log C)	4.27671	4.005	4.6787		
Distance	0.611	0.676	0.689		
Reference	EPA COVER SHEET 0335;891001;(1)	NDA-18602	NTP 364 39		

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: 902193919: [*]:[c](:[*])C#C
- 3. Unknown FCFP 2 feature: 131784192: [*]C#C
- 4. Unknown ECFP_6 feature: -1114776580: [*]C#[*]
- 5. Unknown ECFP_6 feature: -1939823063: [*]#C
- 6. Unknown ECFP_6 feature: 1410041175: [*]:[cH]:[c](:n:[*]):[c](:[*]):[*]
- 7. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 8. Unknown ECFP_6 feature: -677309799: [*]:[cH]:n:[c](:[*]):[*]
- 9. Unknown ECFP_6 feature: -710237522: [*]:n:[cH]:n:[*]
- 10. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: -1255706725: [*]CO[c](:[*]):[*]
- 12. Unknown ECFP_6 feature: -1790412586: [*]CCO[*]
- 13. Unknown ECFP_6 feature: -1242906247: [*]:[c](:[*])N[c](:[*]):[*]
- 14. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
- 15. Unknown ECFP 6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 16. Unknown ECFP_6 feature: -182178874: [*]#C[c](:[cH]:[*]):[cH]:[*]
- 17. Unknown ECFP_6 feature: 1139738044: [*]:[c](:[*])C#C
- 18. Unknown ECFP_6 feature: -1545539812: [*]C#C
- 19. Unknown ECFP_6 feature: -1253653003: [*]COC

	Top features	for positive contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	[*]coc	0.13
ECFP_6	1559650422	[*]C[*]	0.129
FCFP_6	3	[*]N[*]	0.0924
	Top Features	for negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134

FCFP_6	1	DE LA	-0.102
FCFP_6	-453677277	[*]O[*] [*]C[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.0906



 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.13

Unit: g/kg_body_weight
Mahalanobis Distance: 11.4

Mahalanobis Distance p-value: 1.36e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

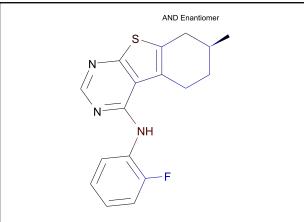
Structural Similar Compounds						
Name	ANILAZINE	D&C YELLOW NO. 11	1-AMINO-2,4- DIBROMOANTHRAQUINO NE			
Structure	HN TO NOT THE COLUMN TO THE CO		H ₂ N ₁ Br			
Actual Endpoint (-log C)	3.78694	4.03869	2.82966			
Predicted Endpoint (-log C)	3.42114	3.54593	3.92444			
Distance	0.563	0.623	0.632			
Reference	NCI/NTP TR-104	NCI/NTP TR-463	NCI/NTP TR-383			

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Feature Contribution					
	Top features	for positive contributio	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	3	AND Enantioner S N H CI [*]N[*]	0.0737		
		ci [*]N[*]		_	

FCFP_2	17	AND Enantiomer S N N N H H CI [*]:S:[*]	0.0441
FCFP_2	590925877	AND Enantiomer N S CI [*]N[c](:[cH]:[*]):[c H]:[*]	0.00762
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	AND Enantiomer N N H CI [*]:[c](:[*])CI	-0.134
FCFP_2	-1272798659	AND Enantiomer S N N N H CI [*]CCC([*])[*]	-0.111
FCFP_2	203677720	AND Enantiomer S CI [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0829



 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.138

Unit: g/kg_body_weight
Mahalanobis Distance: 11.6

Mahalanobis Distance p-value: 6.73e-008

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

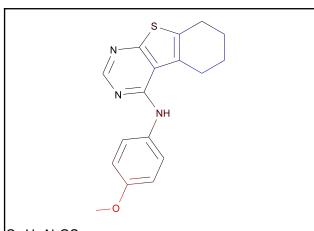
Structural Similar Compounds						
Name	ANILAZINE	D&C YELLOW NO. 11	C.I. SOLVENT YELLOW 14			
Structure	DO NOT THE PROPERTY OF THE PRO		N N N			
Actual Endpoint (-log C)	3.78694	4.03869	4.04277			
Predicted Endpoint (-log C)	3.42114	3.54593	2.8989			
Distance	0.544	0.589	0.612			
Reference	NCI/NTP TR-104	NCI/NTP TR-463	NCI/NTP TR-226			

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Feature Contribution					
Top features	for positive contribution	n			
Bit/Smiles	Feature Structure	Score			
3	AND Enantiomer S NH F [*]N[*]	0.0737			
	Top features	Bit/Smiles Feature Structure 3	Top features for positive contribution Bit/Smiles Feature Structure Score 3 0.0737		

FCFP_2	17	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.0441
FOED 0	500005077	[*]:s:[*]	0.00700
FCFP_2	590925877	AND Enantiomer S N N N H	0.00762
		[*]N[c](:[cH]:[*]):[c H]:[*]	
		or negative contribution	
Fingerprint FCFP_2	Bit/Smiles 71476542	Feature Structure	Score -0.134
		N= NH NH [*]:[c](:[*])CI	
FCFP_2	-1272798659	AND Enantiomer NH NH [*]CCC([[*])[*]	-0.111
FCFP_2	203677720	AND Enantiomer S N N N N H F [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0829



C₁₇H₁₇N₃OS

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: 0.0607

Unit: g/kg_body_weight Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 2.32e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	C.I.PIGMENT RED 3	PHENOLPHTHALEIN	D&C YELLOW NO. 11		
Structure	THE OH O NEO	ОН			
Actual Endpoint (-log C)	2.65635	2.20184	4.03869		
Predicted Endpoint (-log C)	2.97957	2.8857	3.54593		
Distance	0.524	0.579	0.584		
Reference	NCI/NTP TR-407	NCI/NTP TR-465	NCI/NTP TR-463		

Model Applicability

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

- 1. OPS PC9 out of range. Value: 3.507. Training min, max, SD, explained variance: -2.8548, 3.3954, 1.263, 0.0360.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score FCFP_2 136627117 0.173

FCFP_2	1036089772	[*]:[c](:[*])OC	0.0749
FCFP_2	3	S N N N N N N N N N	0.0737
Fig. vo. no. visa t		or negative contributio	
Fingerprint FCFP_2	Bit/Smiles -1272798659	Feature Structure	Score -0.111
		 * CCC([*])[*]	
FCFP_2	203677720	[*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0829
FCFP_2	1	S N N N N N N N N N N N N N N N N N N N	-0.0796

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4 Donors: 2

Model Prediction

Prediction: 0.138

Unit: g/kg_body_weight
Mahalanobis Distance: 9.21

Mahalanobis Distance p-value: 0.000718

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	PHENOLPHTHALEIN	C.I.PIGMENT RED 3	DISPERSE YELLOW 3		
Structure	НО	OH ONEO	NH NH		
Actual Endpoint (-log C)	2.20184	2.65635	2.77703		
Predicted Endpoint (-log C)	2.8857	2.97957	2.80195		
Distance	0.572	0.617	0.650		
Reference	NCI/NTP TR-465	NCI/NTP TR-407	NCI/NTP TR-222		

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Feature Contribution					
Top features	for positive contributio	n			
Bit/Smiles	Feature Structure	Score			
3	AND Enantiomer S N N H H N [*]N[*]	0.0737			
	Top features	Top features for positive contributio Bit/Smiles Feature Structure AND Enantromer H ₂ N	Top features for positive contribution Bit/Smiles Feature Structure Score 3 0.0737		

FCFP_2	17	AND Enantiomer S N N H ₂ N [*]:S:[*]	0.0441
FCFP_2	590925877	AND Enantiomer N N N N N N N N N N N N N	0.00762
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantomer N N N H H P [*]CCC([*])[*]	-0.111
FCFP_2	203677720	AND Enantiomer N N H N H 2 N (N=1) ([*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0829
FCFP_2	16	AND Enantioner N N N N H ₂ N [*][C](:[*]):[*]	-0.0512

 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.1

Unit: g/kg_body_weight
Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 3.79e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	D&C YELLOW NO. 11	C.I. SOLVENT YELLOW 14	PHENOLPHTHALEIN		
Structure		N OH	ОН		
Actual Endpoint (-log C)	4.03869	4.04277	2.20184		
Predicted Endpoint (-log C)	3.54593	2.8989	2.8857		
Distance	0.569	0.610	0.615		
Reference	NCI/NTP TR-463	NCI/NTP TR-226	NCI/NTP TR-465		

Model Applicability

- 1. OPS PC9 out of range. Value: 3.5628. Training min, max, SD, explained variance: -2.8548, 3.3954, 1.263, 0.0360.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Feature Contribution					
	Top features	s for positive contributio	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	3	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.0737		

FCFP_2	17	AND Enantiomer	0.0441
		NH NH	
TOED 0	4004055040	[*]:s:[*]	0.00040
FCFP_2	1294255210	AND Enantiomer	0.00319
		NH NH	
		[*]CN[c](:[*]):[*]	
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantiomer N NH NH	-0.111
		[*]CCC([*])[*]	
FCFP_2	203677720	AND Enantiomer N N N N H H :[*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0829
FCFP_2	16	AND Enantiomer NH NH [*][c](:[*]):[*]	-0.0512

NCI/NTP TR-20

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4 Donors: 3

Model Prediction

Prediction: 0.222

Unit: g/kg_body_weight
Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 3.68e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	FUROSEMIDE	ACETOHEXAMIDE	DAPSONE		
Structure	HO O O NH2 O S O O O O O O O O O O O O O O O O O	NH ON S	H ₂ N O S NH ₂		
Actual Endpoint (-log C)	4.04236	2.55683	3.66258		
Predicted Endpoint (-log C)	2.8614	3.62413	3.26993		
Distance	0.571	0.696	0.700		

Model Applicability

Reference

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

NCI/NTP TR-050

- 1. OPS PC10 out of range. Value: 3.9942. Training min, max, SD, explained variance: -4.0578, 3.2378, 1.25, 0.0353.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

NCI/NTP TR-356

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score FCFP_2 3 0.0737

			-
FCFP_2	17	AND Enantiomer	0.0441
		s s	
		NH	
		HN NH ₂	
		s s	
		[*]:s:[*]	
		egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantiomer	-0.111
		s Y	
		N=	
		NH HN	
		NH ₂	
		[*1000([*1)[*1	
FCFP_2	1872154524	[*]CCC([*])[*]	-0.105
FGFP_2	1872154524	AND Enantiomer	-0.105
		N=	
		HN	
		NH ₂ S	
		[*]C(=S)[*]	
FCFP_2	203677720	AND Enantiomer	-0.0829
		N-ST)	
		\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	
		N NH NA	
		H) —NH ₂ S	
		[*]C[c]1:[c]([*]):[*]	
		:[*]:[c]:1:[*]	

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: 0.109

Unit: g/kg_body_weight
Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 5.29e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	FUROSEMIDE	SALICYLAZOSULFAPYRI DINE	DAPSONE	
Structure	HO O NH2 NH2 S CI	N N N N N N N O H	H ₂ N O S NH ₂	
Actual Endpoint (-log C)	4.04236	3.375	3.66258	
Predicted Endpoint (-log C)	2.8614	2.80292	3.26993	
Distance	0.637	0.660	0.748	
Reference	NCI/NTP TR-356	NCI/NTP TR-457	NCI/NTP TR-20	

Model Applicability

- 1. OPS PC10 out of range. Value: 3.8932. Training min, max, SD, explained variance: -4.0578, 3.2378, 1.25, 0.0353.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	3	AND Enantiomer NH O S NH ₂ [*]N[*]	0.0737	

FCFP_2	17	AND Enantiomer S NH NH S NH S NH ₂	0.0441
FCFP_2	590925877	[*]:s:[*] AND Enantiomer N N N	0.00762
	Top Features for ne	[*]N[c](:[cH]:[*]):[c H]:[*]	
Fingerprint			Score
FCFP_2	-1272798659	AND Enantiomer S NH NH S NH ₂ [*]CCC([*])[*]	-0.111
FCFP_2	1872154524	AND Enantiomer S NH NH C S NH ₂ [*]C(=S)[*]	-0.105
FCFP_2	203677720	AND Enantiomer N N N N N N N N N N N N N 1 1 1 1 1 1	-0.0829

N NH NH O S H₂N O

C₁₇H₁₈N₄O₂S₂ Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: 0.109

Unit: g/kg_body_weight Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 5.29e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

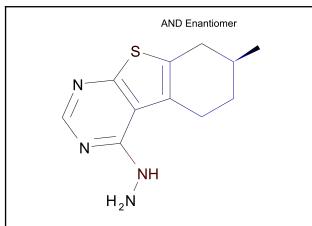
Name	FUROSEMIDE	SALICYLAZOSULFAPYRI DINE	DAPSONE
Structure	HO O NH2 NH2 S O S	N N N N N N N O O O O O O O O O O O O O	H ₂ N O O NH ₂
Actual Endpoint (-log C)	4.04236	3.375	3.66258
Predicted Endpoint (-log C)	2.8614	2.80292	3.26993
Distance	0.637	0.660	0.748
Reference	NCI/NTP TR-356	NCI/NTP TR-457	NCI/NTP TR-20

Model Applicability

- 1. OPS PC10 out of range. Value: 3.8932. Training min, max, SD, explained variance: -4.0578, 3.2378, 1.25, 0.0353.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	3	AND Enantiomer N N N N H N N H 2N [*]N[*]	0.0737	
		<u> </u>		-

FCFP_2	17	AND Enantiomer N N N N H H O18 H ₂ N [*]:S:[*]	0.0441
FCFP_2	590925877	AND Enantiomer N N N N N N N N (*]N[c](:[cH]:[*]):[c H]:[*]	0.00762
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantomer N N N N N H H 2019 [*]CCC([*])[*]	-0.111
FCFP_2	1872154524	AND Enantomer S N N N H H C H 2 C (=S)[*]	-0.105
FCFP_2	203677720	AND Enantiomer S N N N N (*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0829



 $C_{11}H_{14}N_{4}S$

Donors: 2

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Model Prediction

Prediction: 0.133

Unit: g/kg_body_weight Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 2.76e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	DAPSONE	4,4'-THIODIANILINE	BENZIDINE,3,3'- DIMETHOXY-		
Structure	H ₂ N OF S	H ₂ N NH ₂	NH 2		
Actual Endpoint (-log C)	3.66258	3.20473	4.06569		
Predicted Endpoint (-log C)	3.26993	3.3362	3.57405		
Distance	0.560	0.578	0.611		
Reference	NCI/NTP TR-20	NCI/NTP TR-47	NCI/NTP Report 10, Nov. 1987		

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 3. Unknown FCFP_2 feature: 1070150408: [*]NN

Feature Contribution						
	Top features	for positive contributio	n			
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_2	3	AND Enantiomer S NH H ₂ N [*]N[*]	0.0737			

FCFP_2	17	AND Enantiomer S NH H ₂ N [*]:s:[*]	0.0441
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantiomer S N NH H ₂ N [*]CCC([*])[*]	-0.111
FCFP_2	203677720	AND Enantiomer S N NH H ₂ N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0829
FCFP_2	16	AND Enantiomer S NH H ₂ N	-0.0512

[*][c](:[*]):[*]

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.0912
Unit: g/kg_body_weight
Mahalanobis Distance: 13.2

Mahalanobis Distance p-value: 6.77e-011

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRI DINE	C.I.PIGMENT RED 3		
Structure	DH NAME OF THE PARTY OF THE PAR	HN HO HO OH	TOH ONEO		
Actual Endpoint (-log C)	2.30052	3.375	2.65635		
Predicted Endpoint (-log C)	3.55333	2.80292	2.97957		
Distance	0.888	1.031	1.059		
Reference	NCI/NTP TR-411	NCI/NTP TR-457	NCI/NTP TR-407		

Model Applicability

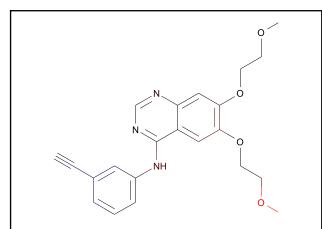
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. ALogP out of range. Value: 7.842. Training min, max, mean, SD: -4.271, 7.574, 2.3494, 1.981.
- 2. Num_AromaticRings out of range. Value: 5. Training min, max, mean, SD: 0, 4, 1.1685, 0.8469.
- 3. OPS PC8 out of range. Value: 4.5344. Training min, max, SD, explained variance: -3.8548, 3.9137, 1.331, 0.0400.
- 4. OPS PC9 out of range. Value: 3.4665. Training min, max, SD, explained variance: -2.8548, 3.3954, 1.263, 0.0360.
- 5. OPS PC10 out of range. Value: 3.814. Training min, max, SD, explained variance: -4.0578, 3.2378, 1.25, 0.0353.
- 6. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

		T	
FCFP_2	3	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.0737
FCFP_2	17	AND Enantiomer N S N N N N N N N N N N	0.0441
FCFP_2	590925877	AND Enantiomer N S O O O O O O O O O O O O O O O O O O	0.00762
	Top Features for no	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.111
FCFP_2	203677720	AND Enantlomer N N N N N N N N N N N N N N N N N N	-0.0829

FCFP_2	16	S N N N N N N N N N N N N N N N N N N N	-0.0512
		s <u> </u>	



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model	Prediction
-------	-------------------

Prediction: 0.0828

Unit: g/kg_body_weight
Mahalanobis Distance: 9.37

Mahalanobis Distance p-value: 0.000417

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	BUTYL BENZYL PHTHALATE	3,3'- DIMETHOXYBENZIDINE- 4,4'-DIISOCYANATE	PYRILAMINE		
Structure		-01m-10-10-10-10-10-10-10-10-10-10-10-10-10-			
Actual Endpoint (-log C)	2.79569	2.17504	3.32511		
Predicted Endpoint (-log C)	3.18498	3.78717	3.65163		
Distance	0.758	0.795	0.820		
Reference	NCI/NTP TR-458	NCI/NTP TR-128	NCI/NTP Report 10, Nov. 1987		

Model Applicability

Unknown features are fingerprint features in 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: -124685461: [*]:n:c:n:[*]
- 3. Unknown FCFP_2 feature: 902193919: [*]:[c](:[*])C#C
- 4. Unknown FCFP_2 feature: 131784192: [*]C#C

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	136627117	[*]oc	0.173	

FCFP_2	-1143715940		0.095
FCFP_2	1036089772	[*]COC	0.0749
		[*]:[c](:[*])OC	
	Ton Footures f		
Fingerprint	Bit/Smiles	or negative contributio Feature Structure	Score
FCFP_2	203677720	[*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.0829
FCFP_2	1	[*]O[*]	-0.0796
FCFP_2	16	[*][c](:[*]):[*]	-0.0512

AND Enantiomer S N NH CI

C₁₇H₁₆CIN₃S

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.00765
Unit: g/kg_body_weight
Mahalanobis Distance: 9.61

Mahalanobis Distance p-value: 1.91e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	PHENYLBUTAZONE	o-BENZYL-p- CHLOROPHENOL	CHLORPHENIRAMINE MALEATE	
Structure	N N N N N N N N N N N N N N N N N N N	HO NO CI	CI	
Actual Endpoint (-log C)	3.48909	3.26063	3.96188	
Predicted Endpoint (-log C)	3.17333	3.64448	3.83117	
Distance	0.862	0.886	0.887	
Reference	NCI/NTP TR-367	NCI/NTP TR-424	NCI/NTP TR-317	

Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- OPS PC7 out of range. Value: -2.8755. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 5. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score FCFP_2 32 AND Enantomer 0.526

FCFP_2	367998008	AND Enantiomer S N N N H C [*]:[cH]:[c](CI):[cH] :[*]	0.413
FCFP_2	3	AND Enantiomer S N N N H CI [*]N[*]	0.104
Fig. vo. no. visa t		for negative contributio	
Fingerprint FCFP_2	Bit/Smiles 136597326	Feature Structure	Score -0.489
		S N=N+ CI [*]C([*])C	
FCFP_2	203677720	AND Enantioner S S C C C [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.406
FCFP_2	0	AND Enantiomer S N N N N N CI [*]C[*]	-0.29

AND Enantiomer S N NH F

 $C_{17}H_{16}FN_3S$

Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.00972 Unit: g/kg_body_weight Mahalanobis Distance: 9.46

Mahalanobis Distance p-value: 3.06e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

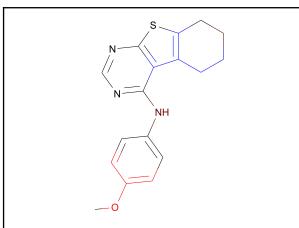
Structural Similar Compounds				
Name	PHENYLBUTAZONE	o-BENZYL-p- CHLOROPHENOL	CHLORPHENIRAMINE MALEATE	
Structure		HO NOT CI	N N CI	
Actual Endpoint (-log C)	3.48909	3.26063	3.96188	
Predicted Endpoint (-log C)	3.17333	3.64448	3.83117	
Distance	0.845	0.861	0.867	
Reference	NCI/NTP TR-367	NCI/NTP TR-424	NCI/NTP TR-317	

Model Applicability

- 1. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- OPS PC7 out of range. Value: -2.8344. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 5. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

-eature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	32	AND Enantiomer N N N N F	0.526		
		[*]Cl			

FCFP_2	367998008	AND Enantiomer S N N H F [*]:[cH]:[c](CI):[cH] :[*]	0.413
FCFP_2	3	AND Enantiomer S NH NH [*]N[*]	0.104
Eingorprint	Top Features Bit/Smiles	for negative contributio Feature Structure	on Score
Fingerprint FCFP_2	136597326	AND Enantiomer	-0.489
		NH NH F [*]C([*])C	
FCFP_2	203677720	AND Enantioner S N H F [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.406
FCFP_2	0	AND Enantiomer S N NH NH [*]C[*]	-0.29



C₁₇H₁₇N₃OS

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: 0.00084
Unit: g/kg_body_weight
Mahalanobis Distance: 9.84

Mahalanobis Distance p-value: 9.32e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	8-METHOXYPSORALEN	PHENYLBUTAZONE	PROMETHAZINE.HCL		
Structure	O NO O		Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z		
Actual Endpoint (-log C)	3.45978	3.48909	3.93152		
Predicted Endpoint (-log C)	4.14745	3.17333	4.72433		
Distance	0.819	0.870	0.902		
Reference	NCI/NTP TR-359	NCI/NTP TR-367	NCI/NTP TR-425		

Model Applicability

- I. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC6 out of range. Value: -3.1044. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown FCFP 2 feature: -124685461: [*]:n:c:n:[*]
- 5. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	332760439	[*]O[c](:[cH]:[*]):[c H]:[*]	0.672	

FCFP_2	1	S N N H N N H N N N H N N N N N N N N N	0.511
FCFP_2	3		0.104
	Top Features for ne		
Fingerprint	D://Cm:los	F1 Ct1	
		Feature Structure	Score
FCFP_2	203677720 0	[*]C[c]1:[c]:([*]):[*]	-0.406 -0.29

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: 0.067

Unit: g/kg_body_weight Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 1.12e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	SULFISOOXAZOLE	OCHRATOXIN	2- MERCAPTOBENZOTHIAZ OLE	
Structure	H ₂ N O I S N H	OH HON CI	N SH	
Actual Endpoint (-log C)	2.82494	6.28396	2.34829	
Predicted Endpoint (-log C)	3.0705	5.12358	3.82125	
Distance	0.758	0.915	0.951	
Reference	NCI/NTP TR-138	NCI/NTP TR-358	NCI/NTP TR-332	

Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC9 out of range. Value: -2.7349. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 5. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
		•	•	

FCFP_2	3	AND Enantiomer S N N N	0.104
		H ₂ N [*]N[*]	
FCFP_2	-1272798659	AND Enantiomer S N N N N H	0.0703
		H ₂ N [*]CCC([*])[*]	
Pin and any size t	Top Features for ne		0
Fingerprint FCFP_2	Bit/Smiles 136597326	Feature Structure	Score -0.489
		s N N H H H H (*)C([*))C	
	203677720	AND Enantiomer S N N N N H H 2N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.406
FCFP_2	0	AND Enantlomer S N N N H H 12N [*]C[*]	-0.29

 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.152

Unit: g/kg_body_weight
Mahalanobis Distance: 10.1

Mahalanobis Distance p-value: 4.04e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	PHENYLBUTAZONE	PHENYLBUTAZONE PROMETHAZINE.HCL			
Structure	N N N O		0		
Actual Endpoint (-log C)	3.48909	3.93152	3.45978		
Predicted Endpoint (-log C)	3.17333	4.72433	4.14745		
Distance	0.799	0.872	0.883		
Reference	NCI/NTP TR-367	NCI/NTP TR-425	NCI/NTP TR-359		

Model Applicability

- 1. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]

Feature Contribution					
Top features	for positive contributio	n			
Bit/Smiles	Feature Structure	Score			
3	AND Enantomer N NH [*]N[*]	0.104			
	Top features	Top features for positive contributio Bit/Smiles Feature Structure AND Enanthomer NH	Top features for positive contribution Bit/Smiles Feature Structure Score 3 0.104		

FCFP_2	-1272798659	AND Enantiomer	0.0703
		N=(NH	
		[*1000/[*1]/[*1	
		[*]CCC([*])[*]	
F'		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantiomer	-0.489
		NH	
		[*]C([*])C	
FCFP_2	203677720	AND Enantiomer S N N N	-0.406
		H	
		[*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	
FCFP_2	0	AND Enantiomer	-0.29
		NH	
		[*]C[*]	

 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: 0.32

Unit: g/kg_body_weight Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 5.71e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	SULFISOOXAZOLE	OCHRATOXIN	HC RED 3		
Structure	H ₂ N O N N N N N N N N N N N N N N N N N N	OH MANUAL	OH HN MH 2		
Actual Endpoint (-log C)	2.82494	6.28396	2.59592		
Predicted Endpoint (-log C)	3.0705	5.12358	3.285		
Distance	0.745	0.779	0.893		
Reference	NCI/NTP TR-138	NCI/NTP TR-358	NCI/NTP TR-281		

Model Applicability

Fingerprint

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

- Molecular_PolarSASA out of range. Value: 232.59. Training min, max, mean, SD: 0, 223.97, 50.816, 55.15.
- 2. OPS PC9 out of range. Value: -2.8528. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.

Feature Structure

Score

- 3. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 5. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 6. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]
- 7. Unknown FCFP_2 feature: 1499521844: [*]NC(=S)N

Bit/Smiles

Feature Contribution Top features for positive contribution

FCFP_2	1	AND Enantiomer	0.511
		HN HN NH ₂	
		[*]0[*]	
FCFP_2	3	AND Enantiomer S N NH HN	0.104
) —NH ₂ S [*]N[*]	
FCFP_2	-1272798659	AND Enantiomer	0.0703
		NH HN NH S	
		[*]CCC([*])[*]	
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantiomer S NH NH HN	-0.489
		ў — ^{NH} , s [*]С([*])С	
FCFP_2	203677720	AND Enantiomer S N N N N N N N N N N N N N N N N N N	-0.406
		[*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	

FCFP_2	1872154524	AND Enantiomer	-0.307
		NH HN NH ₂	
		s [*]C(=S)[*]	

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage

AND Enantiomer S N NH O S NH₂ O

 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5 Donors: 2

Model Prediction

Prediction: 0.0657

Unit: g/kg_body_weight Mahalanobis Distance: 9.78

Mahalanobis Distance p-value: 1.14e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

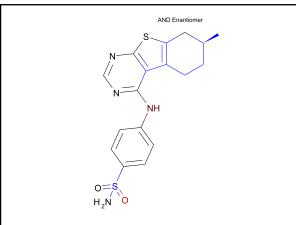
Structural Similar Compounds					
Name	SULFISOOXAZOLE	OCHRATOXIN	PENICILLIN VK		
Structure	H ₂ N O N N H	OH MH OH MA CI	O NH NH S		
Actual Endpoint (-log C)	2.82494	6.28396	2.54455		
Predicted Endpoint (-log C)	3.0705	5.12358	3.9702		
Distance	0.769	0.786	1.038		
Reference	NCI/NTP TR-138	NCI/NTP TR-358	NCI/NTP TR-336		

Model Applicability

- 1. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC9 out of range. Value: -2.8441. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 5. Unknown FCFP 2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution					
	Top features	for positive contributio	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	1	AND Enantiomer S NH S NH S NH [*]O[*]	0.511		

FCFP_2	3	AND Enantiomer S NH O S NH ₂ O [*]N[*]	0.104
FCFP_2	-1272798659	AND Enantiomer NH NH S NH C S NH ₂ [*]CCC([*])[*]	0.0703
	Top Features f	or negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantiomer NH NH NH C NNH ₂ O [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantiomer N N N N N N N N N N N N N N N 1 (*)C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.406
FCFP_2	1872154524	AND Enantiomer S NH O S NH ₂ [*]C(=S)[*]	-0.307



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.0657

Unit: g/kg_body_weight Mahalanobis Distance: 9.78

Mahalanobis Distance p-value: 1.14e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Name	SULFISOOXAZOLE	OCHRATOXIN	PENICILLIN VK
Structure	H ₂ N O N N N N N N N N N N N N N N N N N N	OH O	OH OH
Actual Endpoint (-log C)	2.82494	6.28396	2.54455
Predicted Endpoint (-log C)	3.0705	5.12358	3.9702
Distance	0.769	0.786	1.038
Reference	NCI/NTP TR-138	NCI/NTP TR-358	NCI/NTP TR-336

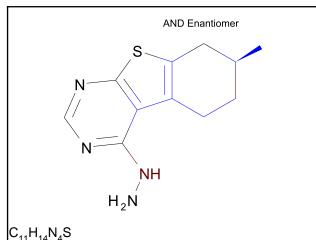
Model Applicability

- I. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC9 out of range. Value: -2.8441. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 4. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 5. Unknown FCFP 2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Cont	ribution			
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	1	AND Enantiomer N N N N N H N N H N N (N) (N) (N) (N) (N) (N) (N) (N) (N)	0.511	

FCFP_2	3	AND Enantiomer N N N N H	0.104
		O: 9 H ₂ N [*]N[*]	
FCFP_2	-1272798659	AND Enantiomer S N N N H	0.0703
		(
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantiomer S N N N N H H 2N [*]C([*])C	-0.489
FCFP_2	203677720	AND Enantiomer S N N N H H 2N 1 (*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.406
FCFP_2	1872154524	AND Enantiomer S N N N N H H 2 N (C S)[*]	-0.307

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



Mala avila a Wainiata 00

Molecular Weight: 234.32065

ALogP: 2.561 Rotatable Bonds: 1 Acceptors: 4

Donors: 2

Model Prediction

Prediction: 0.102

Unit: g/kg_body_weight
Mahalanobis Distance: 8.95

Mahalanobis Distance p-value: 0.00014

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	SULFISOOXAZOLE	2- MERCAPTOBENZOTHIAZ OLE	2-AMINO-5- NITROPHENOL		
Structure	H ₂ N OI OI N H	N S SH	O N NH 2 OH		
Actual Endpoint (-log C)	2.82494	2.34829	2.88684		
Predicted Endpoint (-log C)	3.0705	3.82125	3.28491		
Distance	0.623	0.783	0.795		
Reference	NCI/NTP TR-138	NCI/NTP TR-332	NCI/NTP TR-334		

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[*]:[c]:1:[*]
- 3. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 4. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 1070150408: [*]NN

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	3	AND Enantiomer S N NH H ₂ N [*]N[*]	0.104		

FCFP_2	-1272798659	AND Enantiomer S NH H ₂ N [*]CCC([*])[*]	0.0703
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantiomer S N NH H ₂ N [*]C([*])C	-0.489
FCFP_2	203677720	NH H ₂ N [*]C[c]1:[c]([*]):[*] :[*]:[c]:1:[*]	-0.406
FCFP_2	0	AND Enantiomer S N NH H ₂ N [*]C[*]	-0.29

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4

Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.00163
Unit: g/kg_body_weight
Mahalanobis Distance: 17.8

Mahalanobis Distance p-value: 9.44e-017

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	OCHRATOXIN	SULFISOOXAZOLE	PHENYLBUTAZONE		
Structure	OH NH OH NH CI	H ₂ N O S N T H			
Actual Endpoint (-log C)	6.28396	2.82494	3.48909		
Predicted Endpoint (-log C)	5.12358	3.0705	3.17333		
Distance	1.371	1.541	1.694		
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-367		

Model Applicability

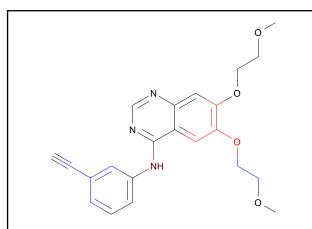
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. Molecular_Weight out of range. Value: 512.69. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
- 2. Num_AromaticRings out of range. Value: 5. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 3. OPS PC6 out of range. Value: -3.2739. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
- OPS PC9 out of range. Value: -2.9987. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 5. Unknown FCFP_2 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 6. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 7. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

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

FCFP_2	3	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.104
FCFP_2	-1272798659	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.0703
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	AND Enantomer N N N N N N N N N N N N N N N N N N N	-0.489
FCFP_2	203677720	AND Enantiomer N N N N N N N N N N N N N N N N N N N	-0.406
FCFP_2	0	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.29

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.000344 Unit: g/kg_body_weight Mahalanobis Distance: 11.7

Mahalanobis Distance p-value: 2.74e-008

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	DIALLYL PHTHALATE	OCHRATOXIN	PROBENECID
Structure		OH MANH CI HO WAY CI HO WAY CI HA CI HO WAY CI HA CI H	OH O
Actual Endpoint (-log C)	3.3914	6.28396	2.85333
Predicted Endpoint (-log C)	3.50093	5.12358	2.4258
Distance	1.113	1.136	1.152
Reference	NCI/NTP TR-284	NCI/NTP TR-358	NCI/NTP TR-395

Model Applicability

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

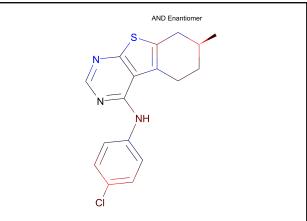
- 1. Num_H_Acceptors out of range. Value: 7. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
- 2. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- OPS PC6 out of range. Value: -3.0997. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
- 4. Unknown FCFP_2 feature: -124685461: [*]:n:c:n:[*]
- 5. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 6. Unknown FCFP_2 feature: 902193919: [*]:[c](:[*])C#C
- 7. Unknown FCFP_2 feature: 131784192: [*]C#C

Feature Contribution

	Top features	s for positive contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	[*]O[c](:[cH]:[*]):[c	0.672
FCFP_2	1	H]:[*]	0.511
FCFP_2	3	[*]O[*]	0.104
	Top Features	for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	[*]C[c]1:[c]:1:[*]	-0.406
FCFP_2	0	[*]C[*]	-0.29

FCFP_2	-1272768868	0′	-0.271
		N 0	
		N H	
		(*)CCO(*)	
		[*]CCO[*]	



 $C_{17}H_{16}CIN_3S$

Molecular Weight: 329.84704

ALogP: 5.5

Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.483

Unit: g/kg_body_weight Mahalanobis Distance: 19.6

Mahalanobis Distance p-value: 1.23e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	FENTIAZAC	BENOXAPROFEN	.alpha(2- CHLOROPHENYL)alpha (4-CHLOROPHENYL)-5- PYRIMIDINE METHANOL
Structure	CION	CI OH	CI NOH
Actual Endpoint (-log C)	2.907	3.408	2.122
Predicted Endpoint (-log C)	2.69613	2.85166	2.72534
Distance	0.554	0.558	0.592
Reference	NIIRDN 6;APP-17;82	YACHDS 9;4445;81	FMCHA2 -;C254;89

Model Applicability

Unknown features are fingerprint features in 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: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[c](:[*]):s:1
 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
- 10. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

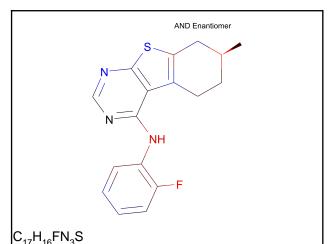
Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

		N N N H		
ECFP_6	-1897341097	[*][C](:[*]):[*] AND Enantiomer N N N H	0.216	
ECFP_6	99947387	CI [*]N[*] AND Enantiomer	0.181	
		[*]:[c](:[*])Cl		
	Top Features for	or negative contribution	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	-817402818	AND Enantlomer S N N N H CI [*]CI	-0.263	
ECFP_6	655739385	AND Enantlomer S N N N H CI [*]:n:[*]	-0.239	

0.281

ECFP_6

ECFP_6	734603939	AND Enantiomer S N N N N	-0.201
		CI [*]C	



Molecular Weight: 313.39244

ALogP: 5.041 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 0.212
Unit: g/kg_body_weight

Mahalanobis Distance: 20.3

Mahalanobis Distance p-value: 6.1e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds			
Name	BENOXAPROFEN	FENTIAZAC	.alpha(2- CHLOROPHENYL)alpha (4-CHLOROPHENYL)-5- PYRIMIDINE METHANOL
Structure	CI	CI OH	CI OH
Actual Endpoint (-log C)	3.408	2.907	2.122
Predicted Endpoint (-log C)	2.85166	2.69613	2.72534
Distance	0.530	0.554	0.570
Reference	YACHDS 9;4445;81	NIIRDN 6;APP-17;82	FMCHA2 -;C254;89

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
- 10. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

		Feature Contribution			
Top features for positive contribution					
Bit/Smiles	Feature Structure	Score			
		<u> </u>	· · · · · · · · · · · · · · · · · · ·		

ECFP_6	-1046436026	AND Enantiomer NH NH F	0.349
ECFP_6	642810091	[*]F AND Enantiomer NH NH [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	AND Enantiomer S NH NH F [*]N[*]	0.216
		or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	AND Enantiomer N NH NH [*]:n:[*]	-0.239
ECFP_6	734603939	AND Enantiomer S N NH NH [*]C	-0.201

FCFP_6	-1539132615	AND Enantiomer S N N N N H F	-0.2
		[*]C[c]1:s:[*]:[c]:1[*]	

N NH NH

C₁₇H₁₇N₃OS

Molecular Weight: 311.40138

ALogP: 4.568 Rotatable Bonds: 3

Acceptors: 4
Donors: 1

Model Prediction

Prediction: 0.45

Unit: g/kg_body_weight Mahalanobis Distance: 18.2

Mahalanobis Distance p-value: 0.000275

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

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

Name -	INDOMETILAZINE	DENOVADBOEEN	EENTLA ZA O
Name	INDOMETHAZINE	BENOXAPROFEN	FENTIAZAC
Structure	HO	CI-CI-CI-CI-CI-CI-CI-CI-CI-CI-CI-CI-CI-C	CI
Actual Endpoint (-log C)	5.17	3.408	2.907
Predicted Endpoint (-log C)	3.33605	2.85166	2.69613
Distance	0.516	0.541	0.552
Reference	ARZNAD 25;1526;75	YACHDS 9;4445;81	NIIRDN 6;APP-17;82

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091		0.281
ECFP_6	-1897341097	[*][c](:[*]):[*] N N N N N N N N N	0.216
FCFP_6	136627117	s N= N= N+ H OC	0.17
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-176455838	[*]O[c](:[cH]:[*]):[c H]:[*]	-0.257
ECFP_6	655739385	N S N N N N N N N N N N N N N N N N N N	-0.239

ECFP_6	734603939	S T N H	-0.201
		-∘′ [*]C	

 $C_{17}H_{18}N_4S$

Molecular Weight: 310.41662

ALogP: 4.089 Rotatable Bonds: 2

Acceptors: 4
Donors: 2

Model Prediction

Prediction: 0.611

Unit: g/kg_body_weight
Mahalanobis Distance: 20.4

Mahalanobis Distance p-value: 2.67e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	FLUBENDAZOLE	CARBAMIC ACID; N-(5- BENZOYLBENZIMIDAZOL -2-YL)-; METHYL ESTER	NAPTALAM	
Structure	NH OH	N H NH	O OH	
Actual Endpoint (-log C)	2.088	2.617	1.551	
Predicted Endpoint (-log C)	2.69288	2.2368	1.89036	
Distance	0.574	0.582	0.644	
Reference	YRTMA6 9;11;78	IYKEDH 19;735;88	FMCHA2 -;C206;89	

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
- 10. Unknown FCFP_6 feature: 1069584379: [*]:[c](:[*])N

tion					
Top features for positive contribution					
Bit/Smiles	Feature Structure	Score			
		· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·		

ECFP_6	642810091	AND Enantiomer S N N H ₂ N [*][c](:[*]):[*]	0.281
ECFP_6	971820502	AND Enantlomer S N N N H ₂ N [*]:[cH]:[c](N):[cH]: [*]	0.276
ECFP_6	-1897341097	AND Enantiomer S N N H H 1 1 1 1 1 1 1 1 1 1 1	0.216
	Top Features f	or negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	AND Enantiomer S N N N H 2 N (*) 1:n:[*]	-0.239
ECFP_6	734603939	AND Enantiomer S N N N H H 2 N [*]C	-0.201

FCFP_6	-1539132615	AND Enantiomer N N N N N H 2 N H 2 N (*)C[c]1:s:[*]:[c	-0.2
]:1[*]	

 $|C_{18}H_{19}N_3S|$

Molecular Weight: 309.42856

ALogP: 4.678 Rotatable Bonds: 3

Acceptors: 3 Donors: 1

Model Prediction

Prediction: 0.36

Unit: g/kg_body_weight
Mahalanobis Distance: 19.8

Mahalanobis Distance p-value: 2.15e-008

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	BENOXAPROFEN	FENTIAZAC	TRIARIMOL	
Structure	CI OH	CI OH	O H	
Actual Endpoint (-log C)	3.408	2.907	2.742	
Predicted Endpoint (-log C)	2.85166	2.69613	2.87257	
Distance	0.497	0.539	0.557	
Reference	YACHDS 9;4445;81	NIIRDN 6;APP-17;82	FMCHA2 -;C293;89	

Model Applicability

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

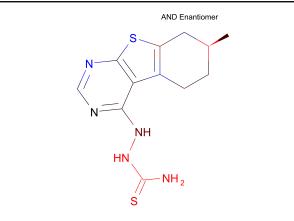
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 907096426: [*]NC[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

Feature Contribution

Top features for positive contribution					
Fingerprint Bit/Smiles Feature Structure Score					
<u> </u>					

ECFP_6	642810091	AND Enantiomer	0.281
		[*][c](:[*]):[*]	
ECFP_6	-1897341097	AND Enantiomer N NH	0.216
		(*)N[*)	
ECFP_6	1571214559	AND Enantiomer S N N N H	0.19
		[*]1:[cH]:[cH]:[cH]:[cH]:1	
		egative contribution	
Fingerprint		Feature Structure	Score
ECFP_6	655739385	AND Enantiomer S N N NH	-0.239
		(*/ [*]:n:[*]	
ECFP_6	734603939	AND Enantiomer S N NH	-0.201
		[*]C	

FCFP_6	-1539132615	AND Enantiomer S N N N N H H H H H H H H H H H H H H	-0.2
		[*]C[c]1:s:[*]:[c]:1[*]	



 $C_{12}H_{15}N_5S_2$

Molecular Weight: 293.411

ALogP: 3.209 Rotatable Bonds: 3

Acceptors: 4
Donors: 3

Model Prediction

Prediction: 0.169

Unit: g/kg_body_weight
Mahalanobis Distance: 23.1

Mahalanobis Distance p-value: 4.71e-021

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	3- CYCLOPENTYLMETHYLH YDROCHLOROTHIAZIDE	2-ANILINO-5- NITROBENZENESULFONI C ACID; SODIUM SALT (Na STRIPPED)	SULFAMETHIZOLE	
Structure	HN M	O N N N N N N N N N N N N N N N N N N N	H ₂ N OIS N H	
Actual Endpoint (-log C)	2.58	2.126	1.687	
Predicted Endpoint (-log C)	1.91133	2.19141	1.85826	
Distance	0.709	0.714	0.735	
Reference	YAKUD5 21;775;79	85JCAE -;1061;86	NIIRDN 6;388;82	

Model Applicability

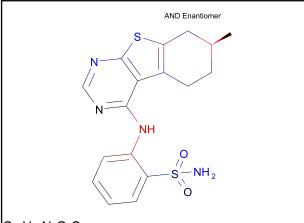
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1294344583: [*]NN[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: -885461129: [*]NNC(=[*])[*]

Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

ECFP_6	642810091	AND Enantiomer S N NH HN NH ₂ NH ₂	0.281
FCFP_6	1499521844	[*][c](:[*]):[*] AND Enantiomer	0.258
		NH NH HN NH ₂ [*]NC(=S)N	
ECFP_6	-1897341097	AND Enantiomer S NH NH NH S NH ₂ [*]N[*]	0.216
	Top Features for n	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	AND Enantiomer NH NH NH S [*]:n:[*]	-0.239
ECFP_6	734603939	AND Enantiomer S N N NH HN NH S NH S [*]C	-0.201

FCFP_6	-1539132615	AND Enantiomer S N N N N N N N N N N N N N N N N N N	-0.2
		[*]C[c]1:s:[*]:[*]:[c]:1[*]	



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 2.95

Unit: g/kg_body_weight
Mahalanobis Distance: 20.8

Mahalanobis Distance p-value: 1.37e-011

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	SULFAQUINOXALINE	2-ANILINO-5- NITROBENZENESULFONI C ACID; SODIUM SALT (Na STRIPPED)	PIRETANIDE	
Structure	N NH O S NH 2	OH NN NH	O OH NH 2 NS O	
Actual Endpoint (-log C)	2.341	2.126	1.811	
Predicted Endpoint (-log C)	2.42674	2.19141	1.83976	
Distance	0.631	0.658	0.662	
Reference	MahWM# 16NOV82	85JCAE -;1061;86	DRFUD4 2;393;77	

Model Applicability

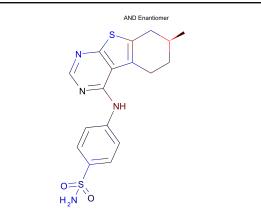
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

ECFP_6	642810091	AND Enantiomer S NH S NH [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	AND Enantiomer S NH O NH O NH O [*]N[*]	0.216
ECFP_6	-1074141656	AND Enantiomer S NH O NH O NH O (*) = O	0.142
		r negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	AND Enantiomer S NH O NH O [*]:n:[*]	-0.239
FCFP_6	-1096219292	AND Enantiomer S N N N N N N N N N N N N	-0.225

ECFP_6	734603939	AND Enantiomer S N NH O NH O S NH ₂	-0.201
		[*]C	



 $C_{17}H_{18}N_4O_2S_2$

Molecular Weight: 374.48042

ALogP: 3.541 Rotatable Bonds: 3

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 2.61

Unit: g/kg_body_weight Mahalanobis Distance: 20.8

Mahalanobis Distance p-value: 9.44e-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 prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	SULFAQUINOXALINE	BENZOTHIAZOLE; 6- NITRO-2-(p- NITROBENZOYLAMINO)-	PIRETANIDE	
Structure	NH NH NH NH NH 2	N NH	OH NH 2 NH	
Actual Endpoint (-log C)	2.341	2.361	1.811	
Predicted Endpoint (-log C)	2.42674	2.96257	1.83976	
Distance	0.626	0.667	0.672	
Reference	MahWM# 16NOV82	JPETAB 90;260;47	DRFUD4 2;393;77	

Model Applicability

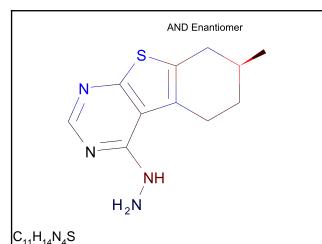
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

ECFP_6	642810091	AND Enantiomer N N N H H (1) (1) (2) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	0.281
ECFP_6	-1897341097	AND Enantiomer N N N H H H (*) N (*	0.216
ECFP_6	-1074141656	AND Enantiomer S N N N H H H E T T T T T T T T T T T T	0.142
		r negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	AND Enantiomer N N N N H H 2 N (*):n:[*]	-0.239
FCFP_6	-1096219292	AND Enantiomer S N N N N H H N (*):[c](:[*])S(=0)(=0) N	-0.225

ECFP_6	734603939	AND Enantiomer S N N H	-0.201
); H ₂ N [*]C	



Molecular Weight: 234.32065

ALogP: 2.561
Rotatable Bonds: 1
Acceptors: 4

Donors: 2

Model Prediction

Prediction: 0.467

Unit: g/kg_body_weight Mahalanobis Distance: 20.5

Mahalanobis Distance p-value: 1.49e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	PHENAZOPYRIDINE .HCI (HCI STRIPPED)	m-PHENYLENEDIAMINE; 4-(PHENYLAZO)-	4;4'-SULFONYLDIPHENOL	
Structure	N N N N N N N N N N N N N N N N N N N	N N N N N N N N N N N N N N N N N N N	HO OI OH	
Actual Endpoint (-log C)	2.724	2.109	1.74	
Predicted Endpoint (-log C)	2.95034	2.69628	2.31209	
Distance	0.497	0.527	0.530	
Reference	TXAPA9 1;42;59	85JCAE -;1308;86	BIOFX* 601-05501;74	

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1294344583: [*]NN[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1070150408: [*]NN

Feature Contribution

i catale contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score		
	•	•	1		

		NH H ₂ N	
ECFP_6	-1897341097	[*][C](:[*]):[*] AND Enantiomer S NH H ₂ N	0.216
ECFP_6	865482986	[*]N[*] AND Enantiomer S N NH H ₂ N [*]C([*])C	0.132
	Top Features f	or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	AND Enantiomer S NH H ₂ N [*]:n:[*]	-0.239
ECFP_6	734603939	AND Enantiomer S N NH H ₂ N [*]C	-0.201

0.281

AND Enantiomer

ECFP_6

FCFP_6	-1539132615	AND Enantiomer	-0.2
		NH H ₂ N	
		[*]C[c]1:s:[*]:[c]:1[*]	

 $C_{28}H_{28}N_6S_2$

Molecular Weight: 512.69212

ALogP: 7.842 Rotatable Bonds: 4 Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.705

Unit: g/kg_body_weight Mahalanobis Distance: 25.3

Mahalanobis Distance p-value: 7.16e-033

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	BENZENESULFONIC ACID; 2;2'-(4;4'- BIPHENYLYLENEDIVINYL ENE)DI-; DISODIUM SALT (Na STRIPPED)	ANTHRAQUINONE; 1;4- bis-(p-TOLYLAMINO)-	FENDOSAL	
Structure	OH DE CONTRACTOR	HN eta.	OH OH	
Actual Endpoint (-log C)	1.968	2.058	2.928	
Predicted Endpoint (-log C)	1.72109	1.57464	2.59	
Distance	0.793	0.960	1.025	
Reference	MVCRB3 2;193;73	85JCAE -;1330;86	AGACBH 8;209;78	

Model Applicability

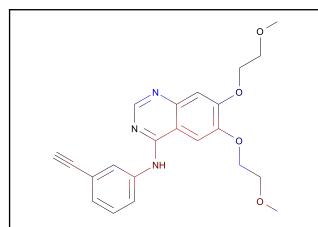
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1597579789: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 5. Unknown FCFP_6 feature: -1564473960: [*]:n:[c]1:s:[*]:[c]:1:[*]
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]

	Top features	for positive contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
	Bidonnics	i catare of actare	

ECFP_6	642810091	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.281
ECFP_6	-1897341097	AND Enantiomer N N N N N N N N N N N N N N N N N N N	0.216
ECFP_6	865482986	AND Enantiomer N N N N N N N N N N N N N N N N N N	0.132
	Top Features for	or negative contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.239
ECFP_6	734603939	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.201

FCFP_6	-1539132615	AND Enantiomer N N N N N N N N N N N N N N N N N N	-0.2
		[*]C[c]1:s:[*]:[c]:1[*]	

Erlotinib



 $C_{22}H_{23}N_3O_4$

Molecular Weight: 393.43572

ALogP: 4.309

Rotatable Bonds: 10

Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.662

Unit: g/kg_body_weight Mahalanobis Distance: 20.8

Mahalanobis Distance p-value: 9.57e-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 prediciton. For highly nonnormal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds				
Name	TALNIFLUMATE	3- QUINOLINECARBOXYLIC ACID; 6;7-bis- (CYCLOPROPYLMETHOX Y)-4-HYDROXY-; ETHYL ESTER	DIXYRAZINE .HCI (HCI STRIPPED)	
Structure	P P P P P P P P P P P P P P P P P P P	OH OH	N N N N N N N N N N N N N N N N N N N	
Actual Endpoint (-log C)	1.538	2.076	3.029	
Predicted Endpoint (-log C)	2.82541	2.50101	2.47585	
Distance	0.615	0.615	0.654	
Reference	FRPSAX 36;372;81	TXAPA9 18;185;71	ANPBAZ 61;669;61	

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1139738044: [*]:[c](:[*])C#C
- 3. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
- 4. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
 5. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[c](:[*]):s:1
- 6. Unknown FCFP_6 feature: -124685461: [*]:n:[cH]:n:[*]
- 7. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: 902193919: [*]:[c](:[*])C#C
- 10. Unknown FCFP_6 feature: 131784192: [*]C#C

Feature Contribution					
Top features for positive contribution					
Fingerprint Bit/Smiles Feature Structure Score					
	Top features	Top features for positive contribution	Top features for positive contribution		

ECFP_6	642810091	0,	0.281
		N N N N N N N N N N N N N N N N N N N	
		[*][c](:[*]):[*]	
ECFP_6	-1897341097		0.216
		N H	
		[*]N[*]	
FCFP_6	136627117	, N	0.17
		İ Ş	
		[*]OC	
	Top Features for no	egative contribution	
Fingerprint	Bit/Smiles		Score
ECFP_6	655739385	, N, Q, 10	-0.239
		N N N N N N N N N N N N N N N N N N N	
		[*]:n:[*]	
ECFP_6	734603939		-0.201
		N N N N N N N N N N N N N N N N N N N	
		[*]C	

FCFP_6	1036089772	oʻ -0.1	36
		N T T	
		N N N	
		• •	
		[*]:[c](:[*])OC	