Electronic Supplementary Material (ESI) for RSC Advances. This journal is © The Royal Society of Chemistry 2022

Design, synthesis, and biological evaluation of novel bioactive thalidomide analogs as anticancer immunomodulatory agents

Anas Ramadan Kotb^a, Dina A. Bakhotmah^b*, Abdallah E. Abdallah^a, Hazem Elkady^a, Mohammed S. Taghour^a, Ibrahim. H. Eissa^a, Mohamed Ayman El-Zahabi^a*

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

* Corresponding author:

Mohamed Ayman El-Zahabi

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

Email: malzahaby@azhar.edu.eg

Ibrahim. H. Eissa

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

Email: Ibrahimeissa@azhar.edu.eg

Dina A. Bakhotmah

Department of Chemistry, Faculty of Science, King Abdulaziz University, Jeddah, kingdom of Saudi Arabia.

Email: dbakhotmah@kau.edu.sa

^b Department of Chemistry, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia.

Content

1	Biological testing
	✓ <i>In vitro</i> anti-proliferative activity.
	✓ Estimation of TNF-α, CASP8, and VEGF in HepG-2 cells supernatant
	✓ Estimation of nuclear factor kappa-B P65 (NF-κB P65) in HepG-2 cell
	lysate
2	In silico studies
	✓ Pharmacokinetic profiling study
	✓ ADMET studies
	✓ Toxicity studies
3	Chemistry and materials
4	Spectral and elemental analysis
5	Toxicity report

1- Biological testing

a- In vitro anti-proliferative activities

Anti-proliferative activity screening of the synthesized compounds was performed against three human cancer cell lines: hepatocellular carcinoma (HepG-2), prostate cancer (PC3), and mammary gland cancer (MCF-7), using MTT assay as follows: The cells were cultured in RPMI-1640 medium with 10% fetal bovine serum. Penicillin (100 units/ml) and streptomycin (100μg/ml) were added at 37 C in a 5% CO₂ incubator. The cells were seeded in a 96-well plate at a density of 1.0×10^4 cells/well. at 37 C for 48 h under 5% CO₂. After incubation the cells were treated with different concentration of the synthesized compounds and incubated for 24 h. Then, 20 μl of MTT solution of a concentration of 5mg/ml was added and incubated for 4 h. DMSO (100 μl) was added into each well to dissolve the formed purple formazan. The color intensity was measured and recorded at absorbance of 570 nm using a plate reader (EXL 800, USA). The percentage of cell viability in was calculated as (A570 of treated samples/A570 of untreated sample) X 100.

b- Supernatant preparation

Cell culture supernatants were prepared from HepG-2 cell line. Cells were cultured in RPMI 1640 medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) having 100 U/ml of penicillin and 50 μ g/ml of streptomycin at 37°C in a humidified 5% CO₂/air mixture. The candidate compounds and thalidomide were dissolved in DMSO as a stock solution at 100 mmol/L and diluted with an FBS-free medium to achieve the designated concentrations (10 μ M). The same concentration of DMSO without any compounds was used as a control. The HepG-2 cells were cultured onto 6-well plates (1x106 cells/ well), allowed to adhere for 24 h followed by treatment with the tested compounds. After 72 h, the media was collected and centrifuged for 15 min at 5000 rpm and equal volume of cell culture supernatants were collected and utilized for immunoassay using different kits.

c- Estimation of TNF-α, CASP8, and VEGF in HepG-2 supernatant

The levels of TNF-α, CASP8, and VEGF in cell culture supernatants were estimated by ELISA technique using commercially available matched paired antibodies (R&D Systems Inc., Minneapolis, MN) according to reported procedure.

d- Cell lysate preparation

The tested compounds and thalidomide were incubated for 72 h with HepG-2 cells. Then, the cells were treated with trypsin/EDTA solution (0.25 mM trypsin and 1 mM EDTA dissolved in a phosphate buffer). Cell lysate were washed three times with phosphate buffer saline (PBS, Sigma Chemical Company, St. Louis, MO, USA) and lysed by three repetitive freezing/thawing cycles (thawing at 37 °C for 2 min and freezing at –80°C for 15 min), followed by homogenization of the cells by passing through a 20G needle.

e- Estimation of nuclear factor kappa-B P65 (NF-κB P65) in HepG-2 cell lysate

The cell lysate samples were applied to the microtiter plates in a concentration of 50 μ L/well. The plates were incubated at 37°C for 1 h, then kept at 4°C for 12 h in a humidified chamber. Next, the cell lysate was extracted from the wells and the plates were washed three times with a buffer (PBS/0.05%Tween-20). Blocking buffer (PBS/0.05%Tween-20/5% FBS) (200 μ L) was added to each well and incubated at 37 °C for 1.5 h then washed. Anti-rabbit NF-kB P65 polyclonal antibody was dispensed as 50 μ L/well and incubated for 2 h at 37 °C. The plates were washed and incubated for 1 h with 50 μ L/well of diluted polyclonal goat anti-rabbit-peroxidase conjugate (1:1000). After that, the plates were washed four times followed by addition of TMB and H₂O₂ in equal volume (50 μ L/well). After the development of color, 50 μ L of stopping buffer (1 M HCl) were added per well. The absorbance was measured at 450 nm using the ELISA plate reader (FLUOstar OPTIMA)

2- In silico studies

1- Pharmacokinetic profiling study

To identify potential drug candidates, drug-likeness profiles were developed for the preliminary estimation of the physicochemical, drug-like parameters, and pharmacokinetics using Discovery studio 4.0 utilizing Lipinski's and Veber's rules. Lipinski suggested that the absorption of an orally administered compound is more likely to be better if the molecule satisfies at least three out of four of the following rules: (1) H bond donors (OH, NH, and SH) \leq 5; (2) H bond acceptors (N, O, and S atoms) \leq 10; (3) molecular weight < 500; (4) logP < 5. Compounds violating more than one of these rules could not have good bioavailability. Moreover, reduced molecular flexibility, as

measured by the number of rotatable bonds, and low polar surface area is found to be an important predictors of good oral bioavailability. Compounds that meet the criteria of 10 or fewer rotatable bonds and polar surface area equal to or less than 140 A will have a high probability of good oral bioavailability.

2- ADMET studies

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

3- 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.

4- Chemistry and materials

All solvents and reagents were commercially available and used without further purification. Progress of reactions was monitored by TLC using TLC sheets coated with UV fluorescent silica gel (Kieselgel 0.25mm, 60 F254, Merck Germany) with a developing solvent system of DCM/methanol (95:5) and was visualized using UV lamp. The melting points were determined using a Gallen lamp melting point apparatus. Elemental analyses were accomplished using a CHN analyzer. The infrared spectra were recorded on FT/IR-6600typeA spectrophotometer. 1 H NMR spectra were recorded at 400 and 700 MHz, while 13 C NMR spectra were run at 100 and 176 MHz, on a Bruker Avance NEO-600 equipped with a 1.7 mm TCI CryoProbe. Chemical shifts were expressed in δ (ppm) with reference to TMS and coupling constant (J) in Hertz using DMSO-d₆ and CDCl₃-d₆ as solvents. The mass spectra were recorded on an Agilent 6410 triple-quadrupole mass spectrometer equipped with an ESI source.

5- Spectral and elemental analysis

Lab code	Compound No.
AR-7	6
AR-8	7a
AR-9	7b
AR-10	24b
AR-11	24a
AR-12	24c
AR-13	13a
AR-14	15a
AR-15	11a
AR-17	19b
AR-18	19a
H2-1	13b
H2-2	15b
H2-3	11b

Viability assay against normal cell line

MTT protocol

Determination of sample cytotoxicity on cells (MTT protocol)

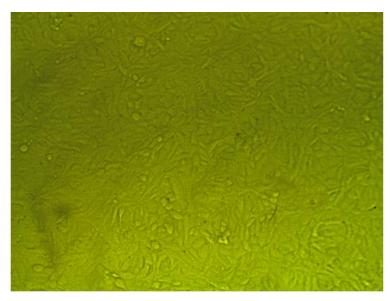
- 1-the 96 well tissue culture plate was inoculated with 1 X 10^5 cells / ml (100 ul / well) and incubated at 37°C for 24 hours to develop a complete monolayer sheet.
- 2- Growth medium was decanted from 96 well micro titer plates after confluent sheet of cells were formed, cell monolayer was washed twice with wash media.
- 3- two-fold dilutions of tested sample was made in RPMI medium with 2% serum (maintenance medium).
- 4- 0.1 ml of each dilution was tested in different wells leaving 3 wells as control, receiving only maintenance medium.
- 5- Plate was incubated at 37°C and examined. Cells were checked for any physical signs of toxicity, e.g. partial or complete loss of the monolayer, rounding, shrinkage, or cell granulation.
- 6- MTT solution was prepared (5mg/ml in PBS) (BIO BASIC CANADA INC).
- 8- 20ul MTT solution were added to each well. Place on a shaking table, 150rpm for 5 minutes, to thoroughly mix the MTT into the media.
- 9) Incubate (37C, 5% CO2) for 4 hours to allow the MTT to be metabolized.
- 10) Dump off the media. (dry plate on paper towels to remove residue if necessary.
- 11) Resuspend formazan (MTT metabolic product) in 200ul DMSO. Place on a shaking table, 150rpm for 5 minutes, to thoroughly mix the formazan into the solvent.
- 12) Read optical density at 560nm and subtract background at 620nm. Optical density should be directly correlated with cell quantity.

Morphological assay

❖ Large-scale, morphological changes that occur at the cell surface, or in the cytoskeleton, can be followed and related to cell viability.

- ❖ Damage can be identified by large decreases in volume secondary to losses in protein and intracellular ions of due to altered permeability to sodium or potassium.
- * Necrotic cells: nuclear swelling, chromatin flocculation, loss of nuclear basophilia
- ❖ Apoptotic cells: cell shrinkage, nuclear condansation, nuclear fragmentation

ID	ug/ml	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
vero		0.84	0.83			0.00360			ug
		2	3	0.83	0.835	6	100	0	
	1000	0.02	0.02	0.01		0.00230	2.63473053	97.3652694	
	1000	2	6	8	0.022	9	9	6	
	500	0.05	0.05	0.06	0.05766	0.00260	6.90618762	93.0938123	163.61
AR 9		3	8	2	7	3	5	8	± 1.12
	250	0.17	0.16	0.17					
		4	8	4	0.172	0.002	20.5988024	79.4011976	
	125	0.43	0.45	0.43	0.44166	0.00480	52.8942115	47.1057884	
	125	5	1	9	7	7	8	2	
	62.5	0.83		0.82		0.00378	99.7604790	0.23952095	
	02.3	2	0.84	7	0.833	6	4	8	
	31.25	0.83	0.84	0.83		0.00305			
	31.23	3	1	1	0.835	5	100	0	
	1000	0.01	0.02	0.01	0.01966	0.00066	2.35528942	97.6447105	
	1000	9	1	9	7	7	1	8	
	500	0.03	0.03	0.05	0.03966	0.00622	4.75049900		161.12
AR 10		2	5	2	7	7	2	95.249501	± 1.42
	250	0.21				0.00916	24.9101796	75.0898203	
		4	0.22	0.19	0.208	5	4	6	
	125	0.43	0.43	0.44	0.43966	0.00366	52.6546906	47.3453093	
	125	6	6	7	7	7	2	8	
	62.5	0.73	0.74		0.73666	0.00352	88.2235528	11.7764471	
	02.3	8	2	0.73	7	8	9	1	
	31.25	0.82	0.84	0.83	0.83366	0.00384	99.8403193	0.15968063	
	31.23	8	1	2	7	4	6	9	

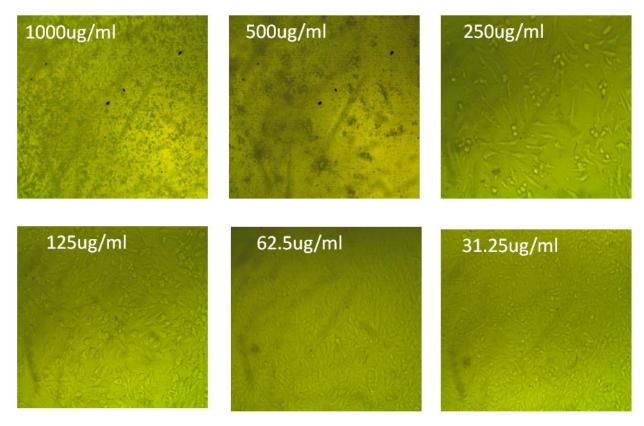


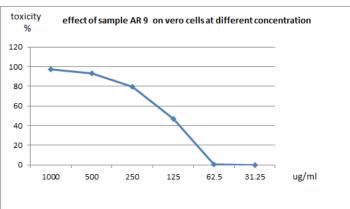
control vero cells

Organism: Cercopithecus aethiops

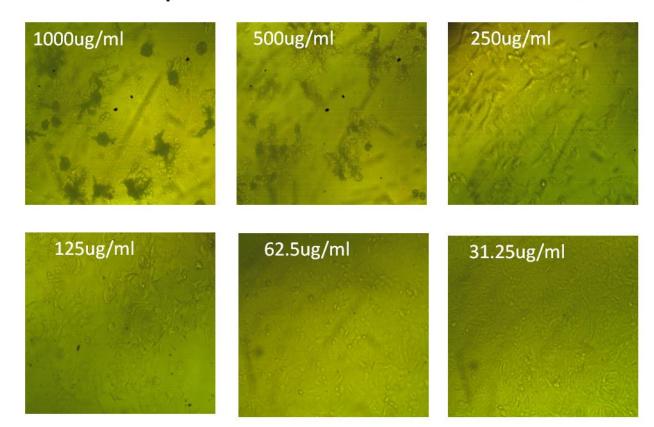
Tissue: kidney
Cell Type: epithelial
Culture Properties: adherent
Disease: normal
ATCC: CCL-81

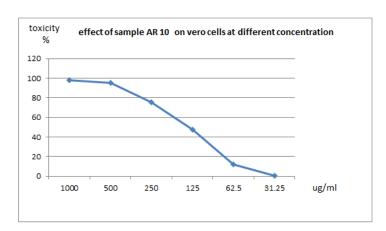
Effect of sample AR 9 on vero cells at different concentration

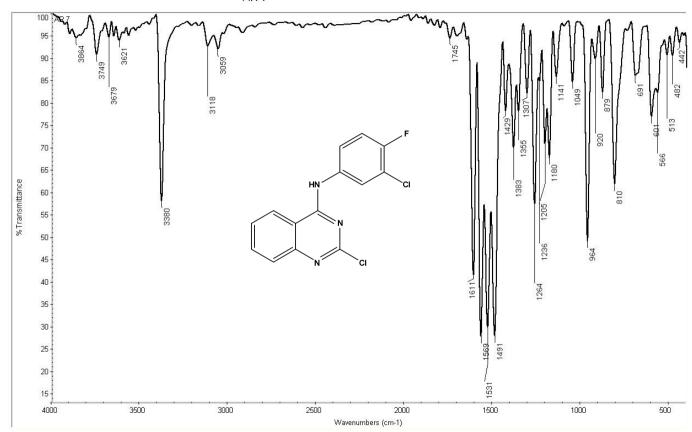


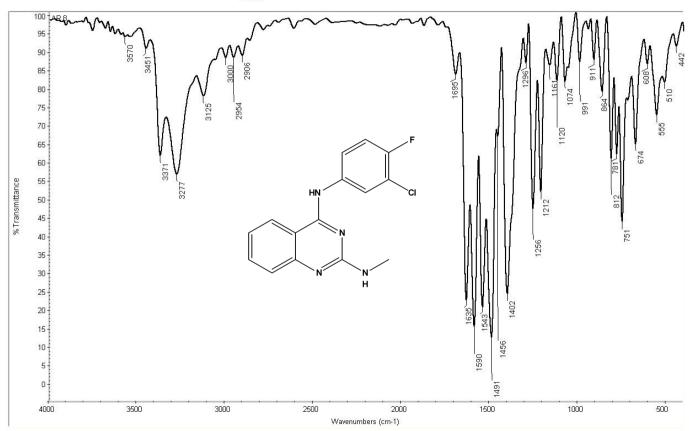


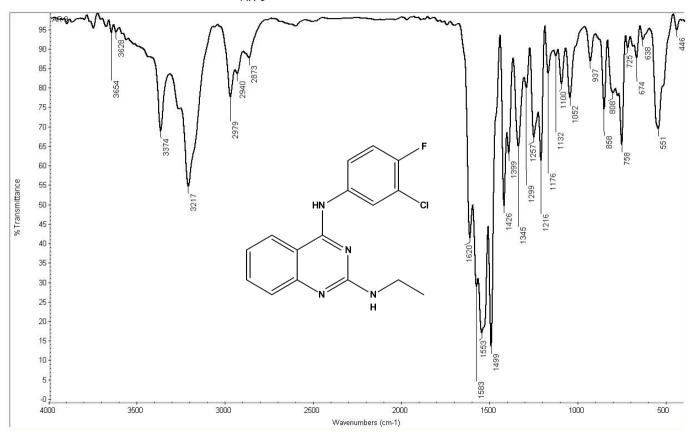
Effect of sample AR 10 on vero cells at different concentration

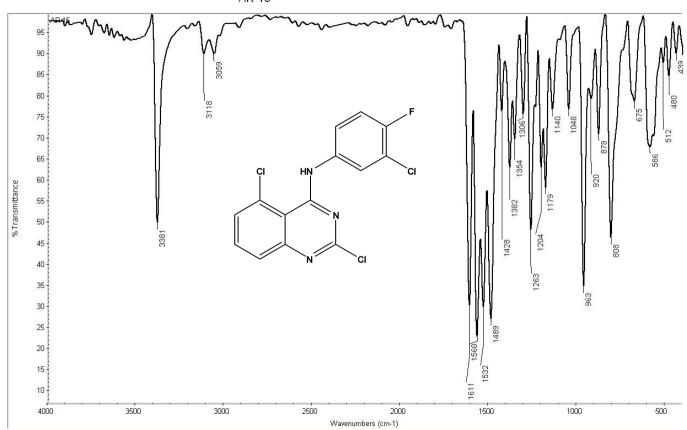


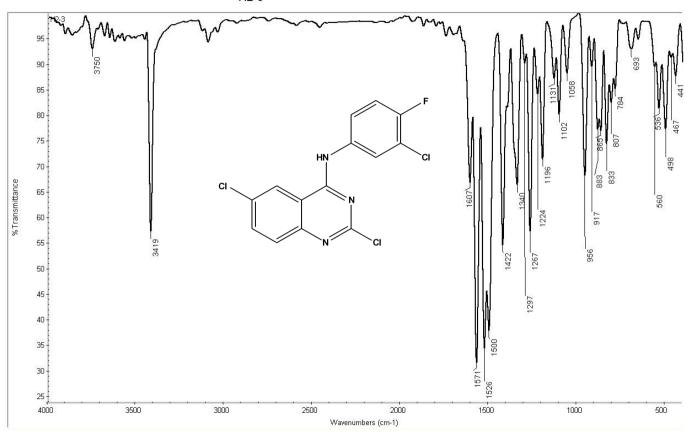


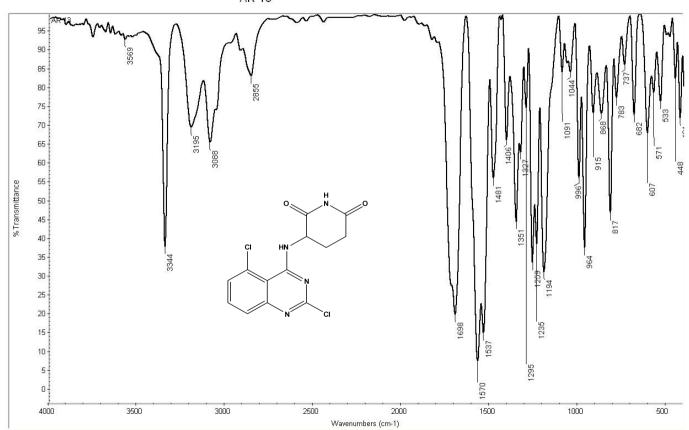


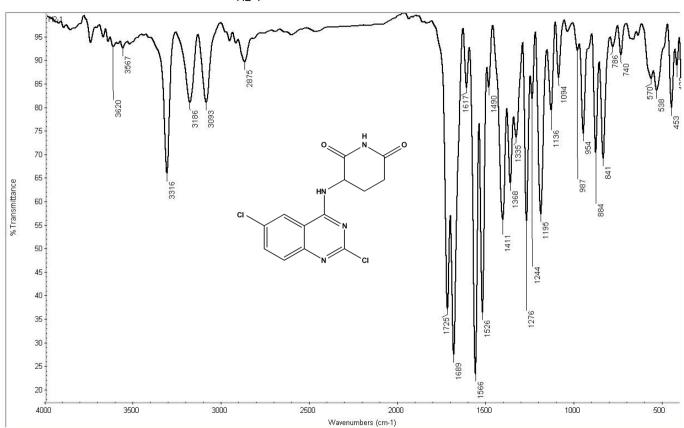


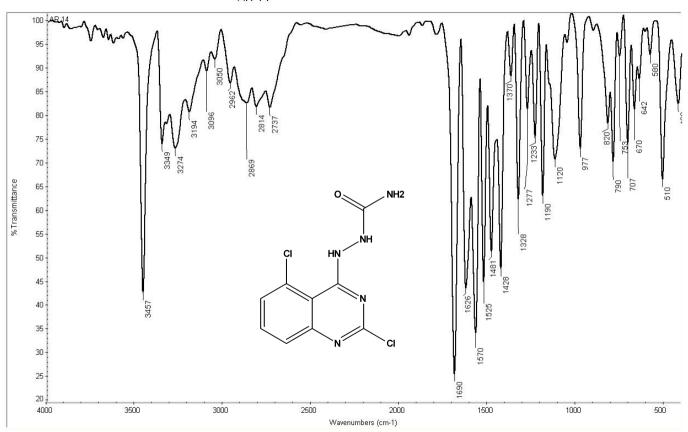


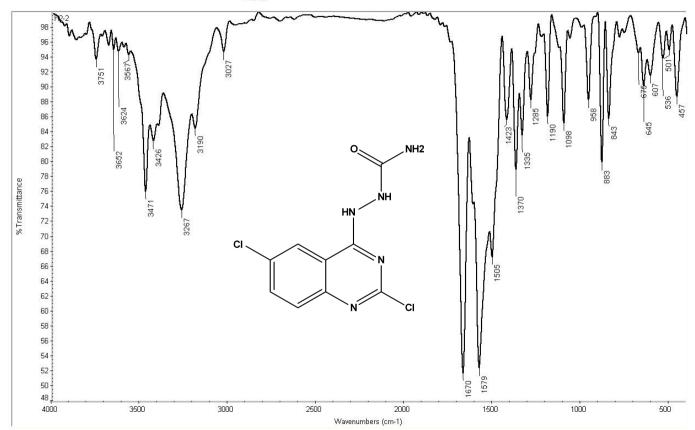


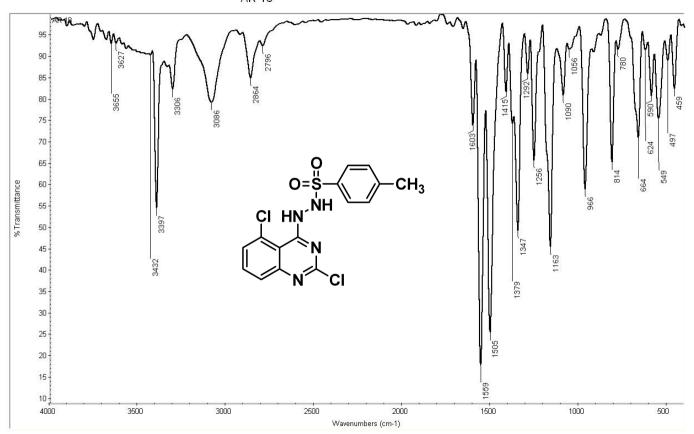


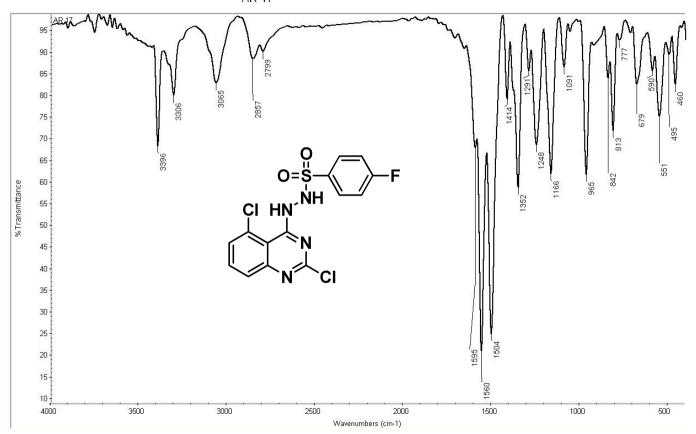


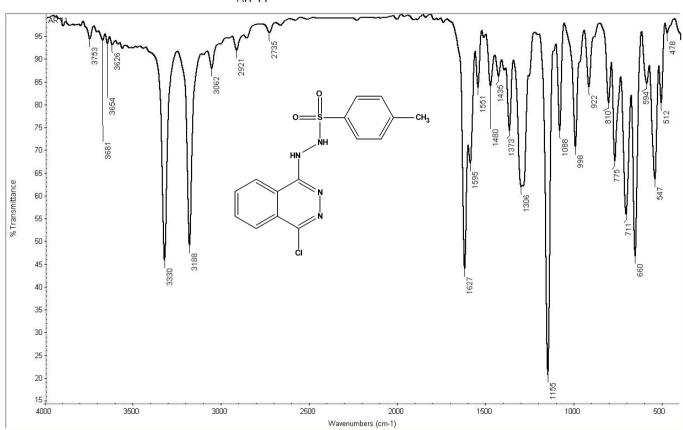


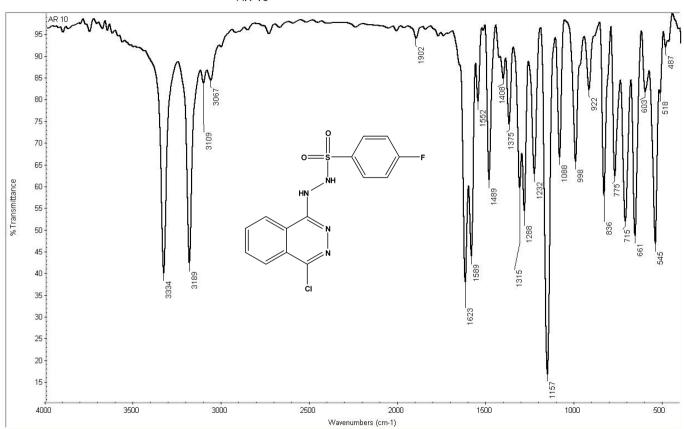


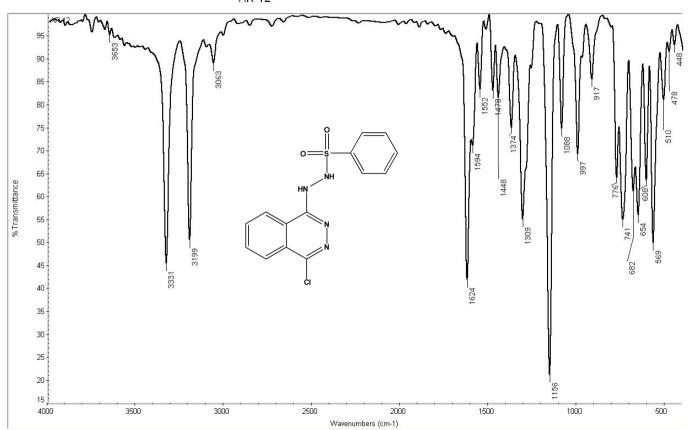


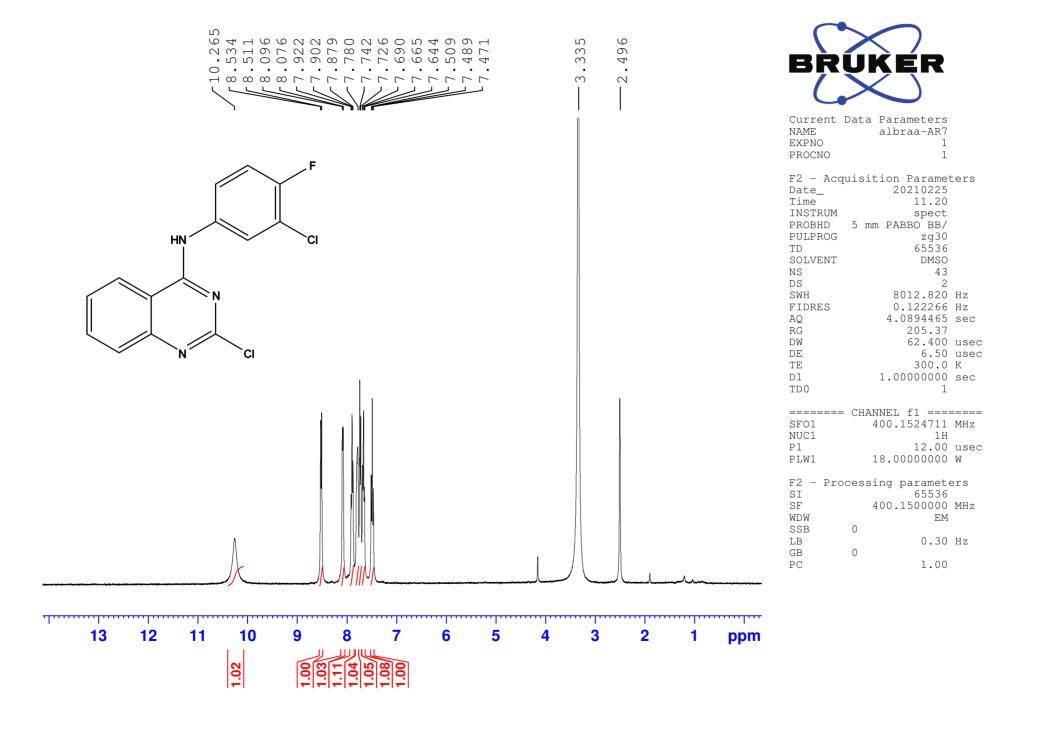


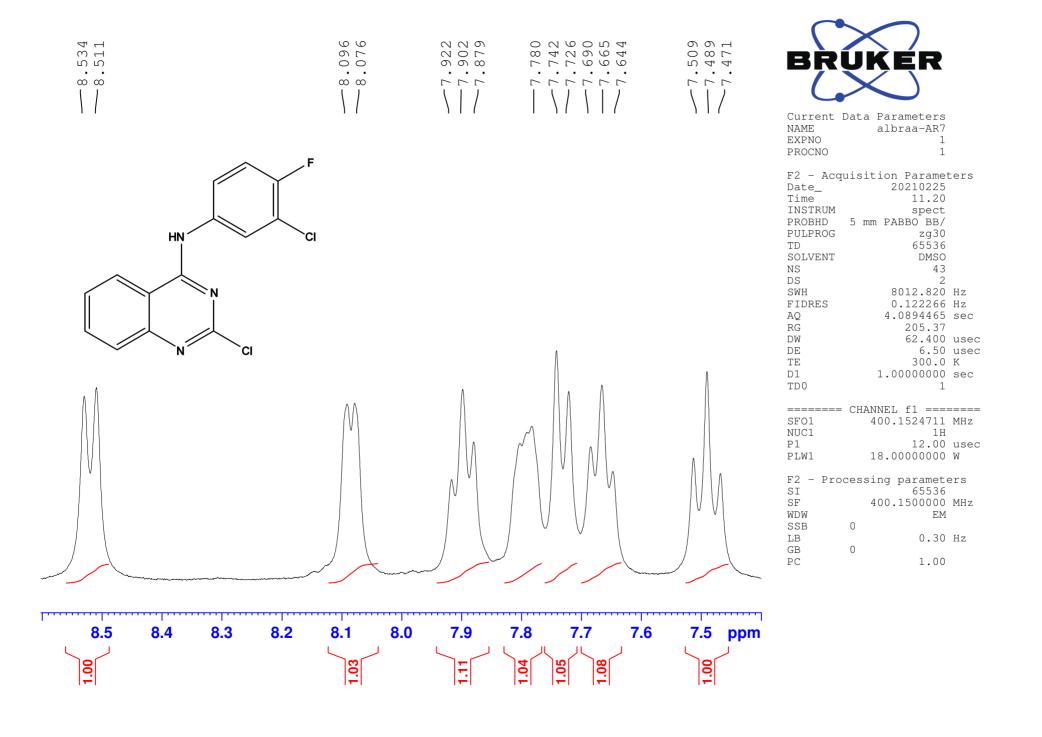


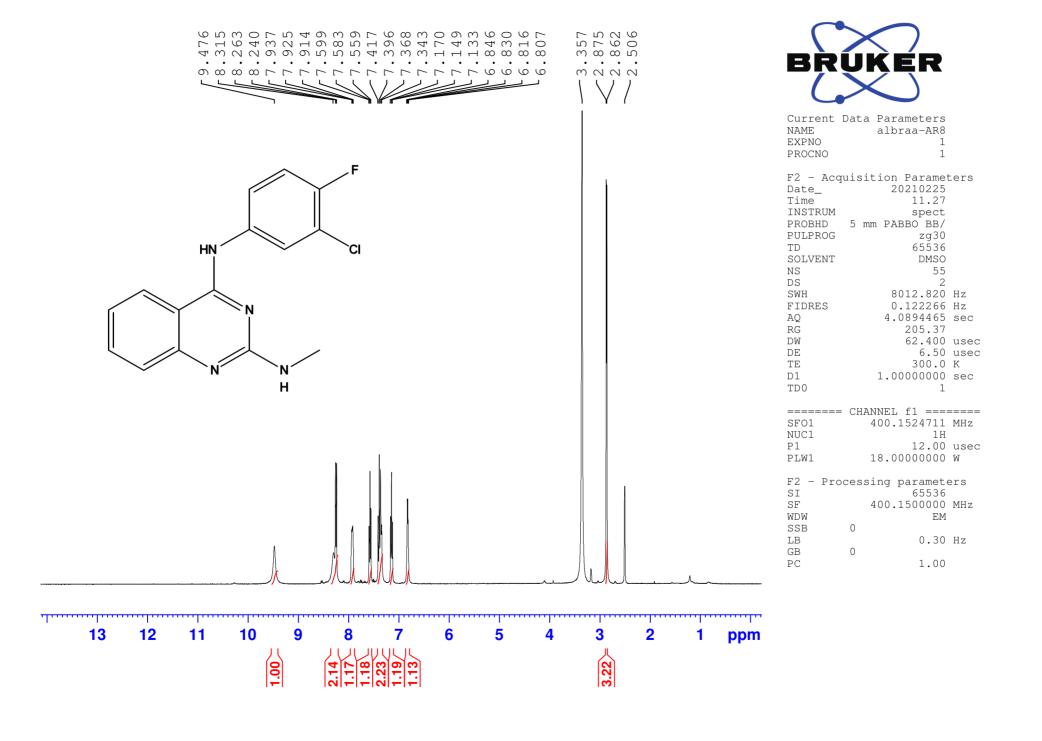


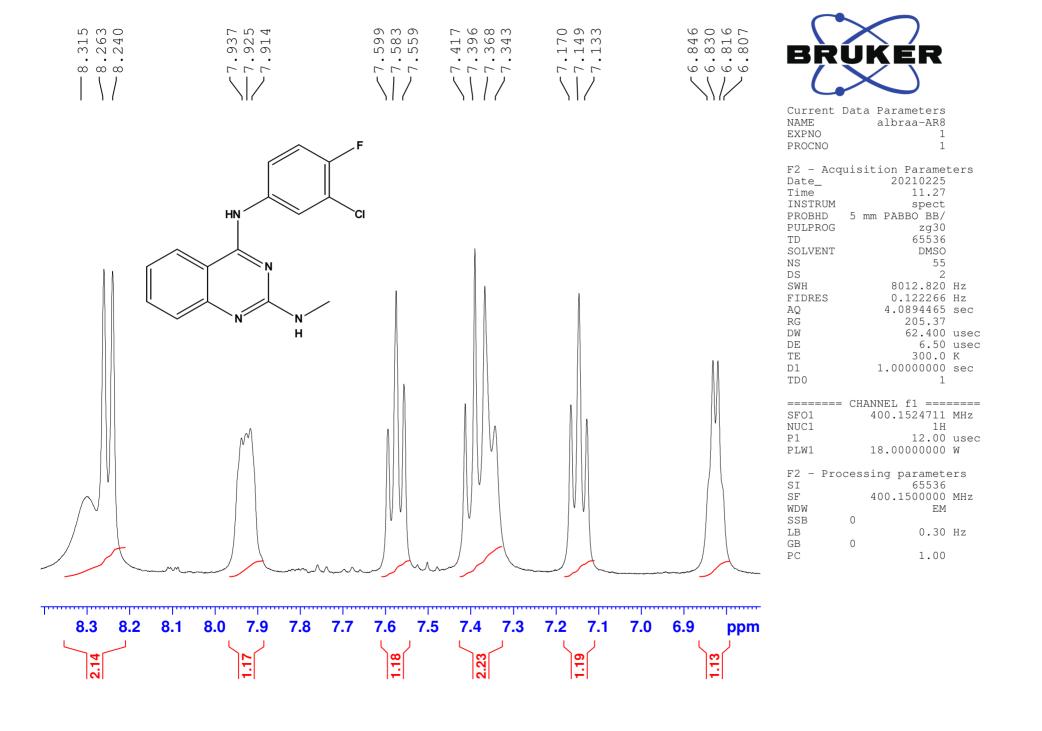


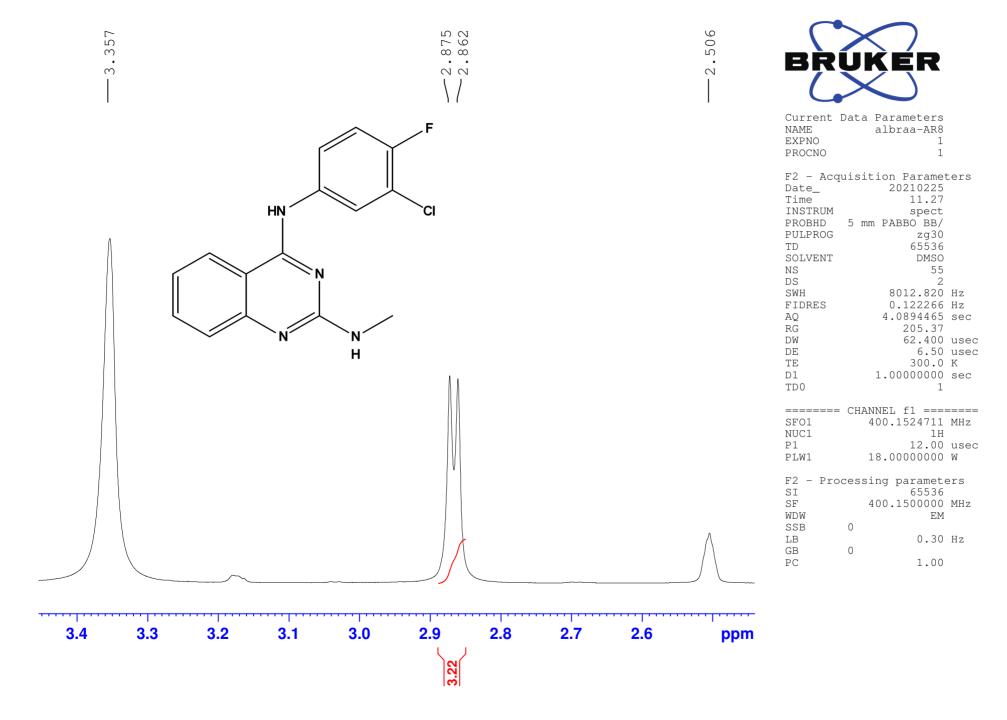


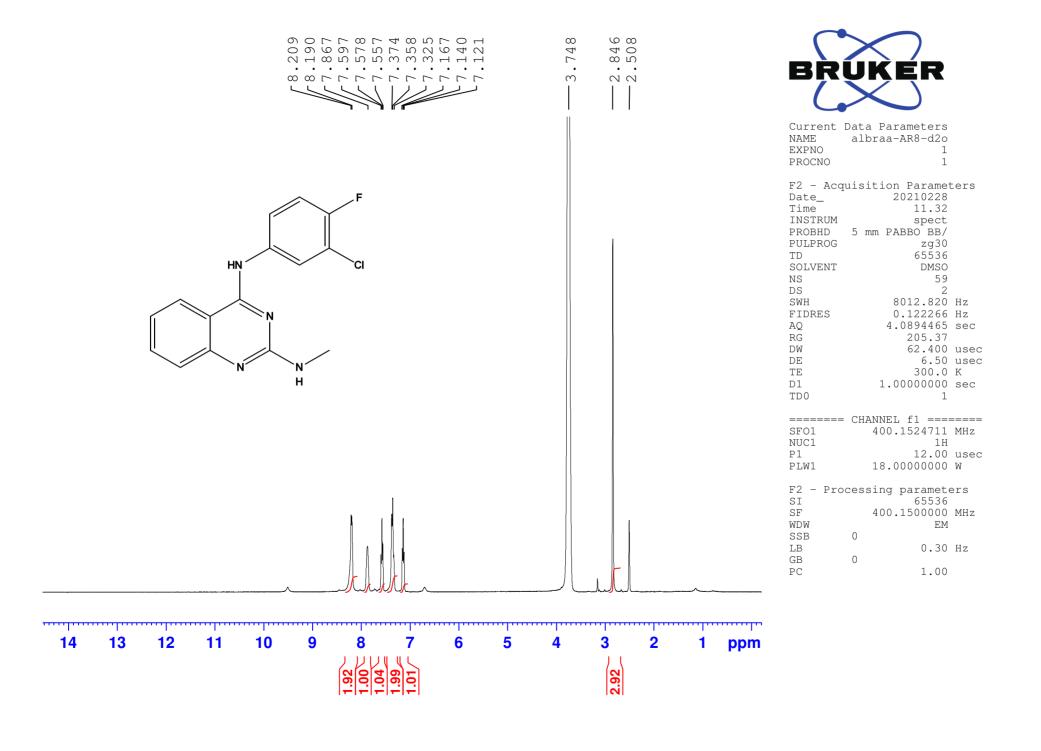


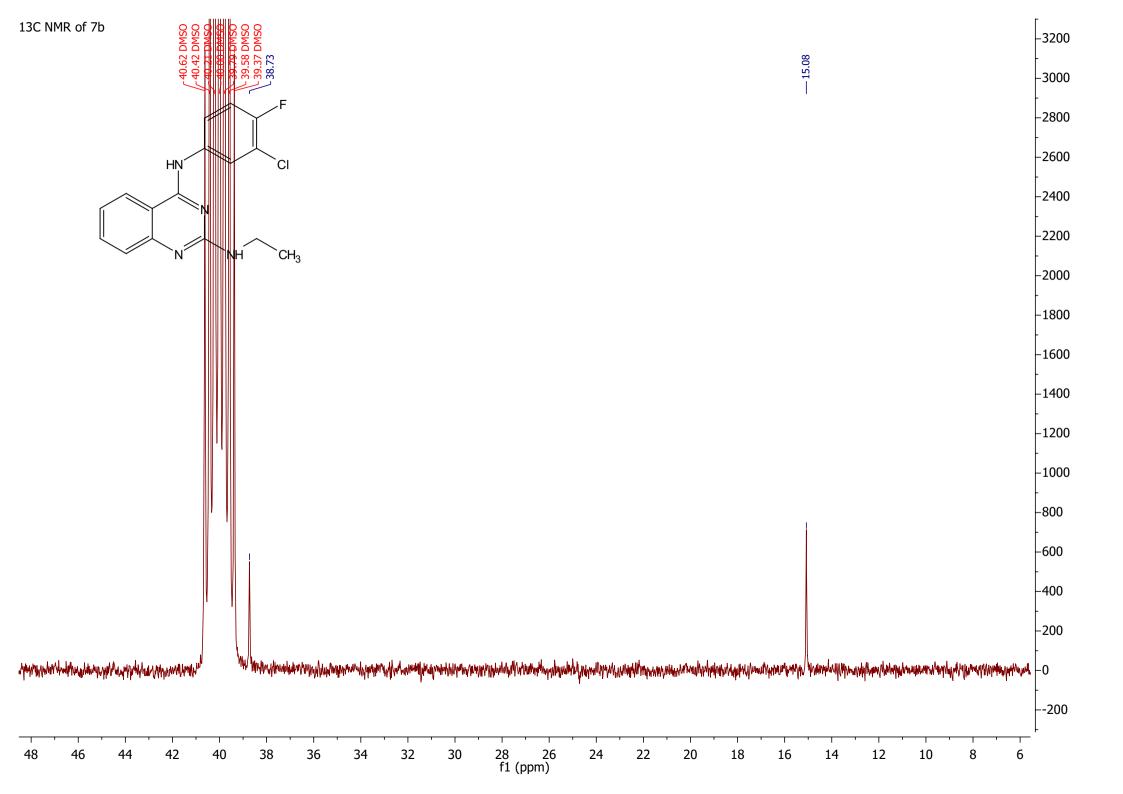


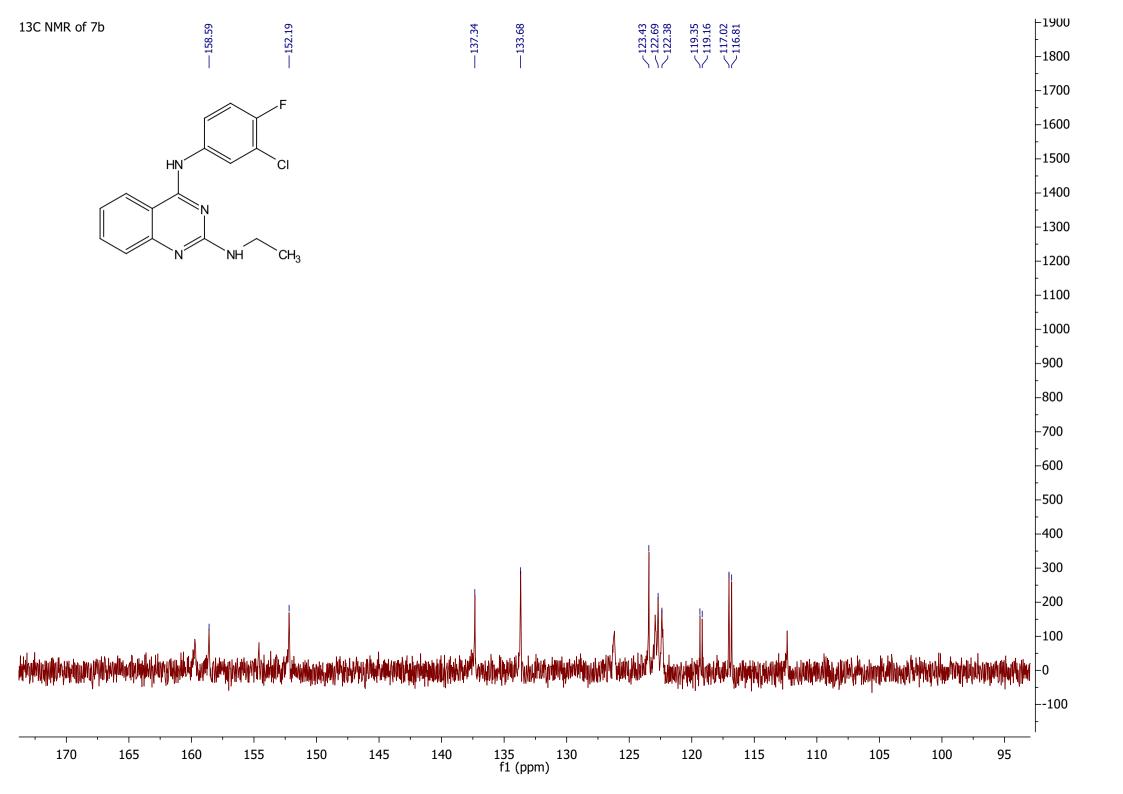


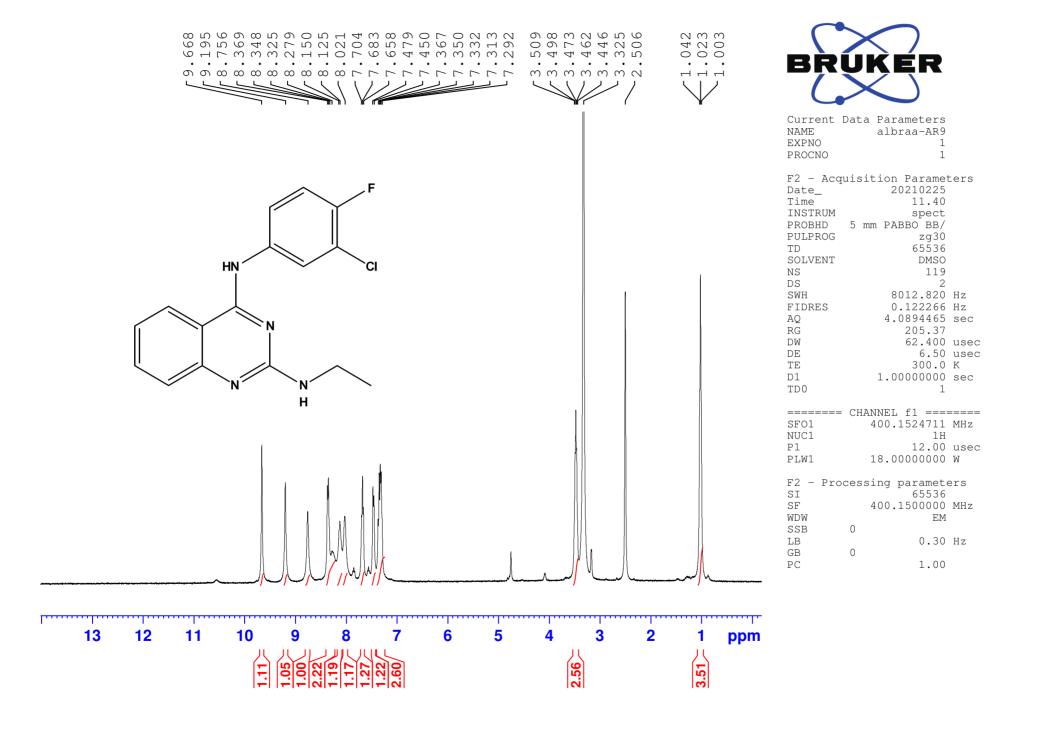


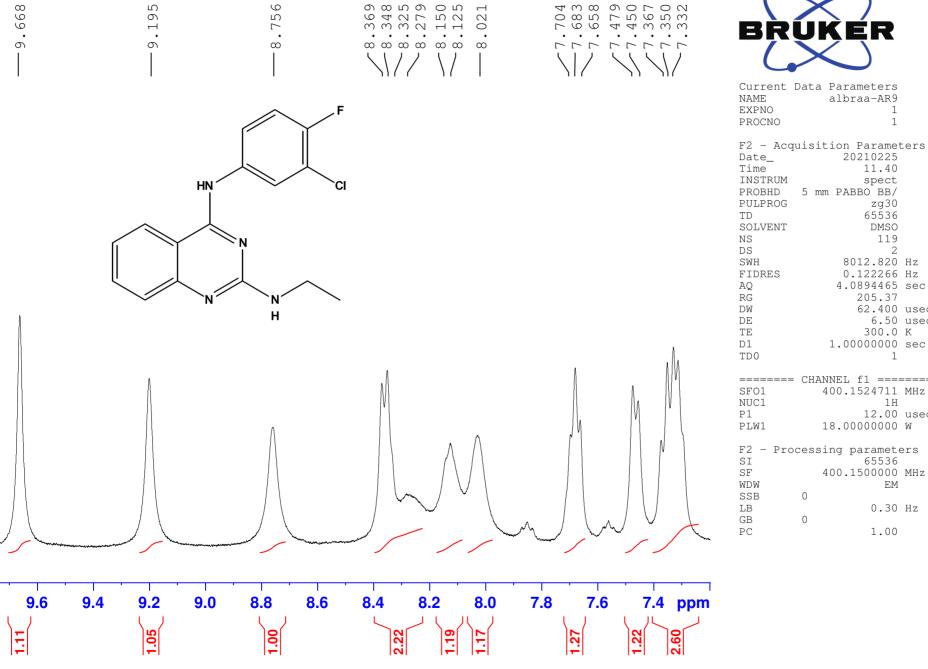










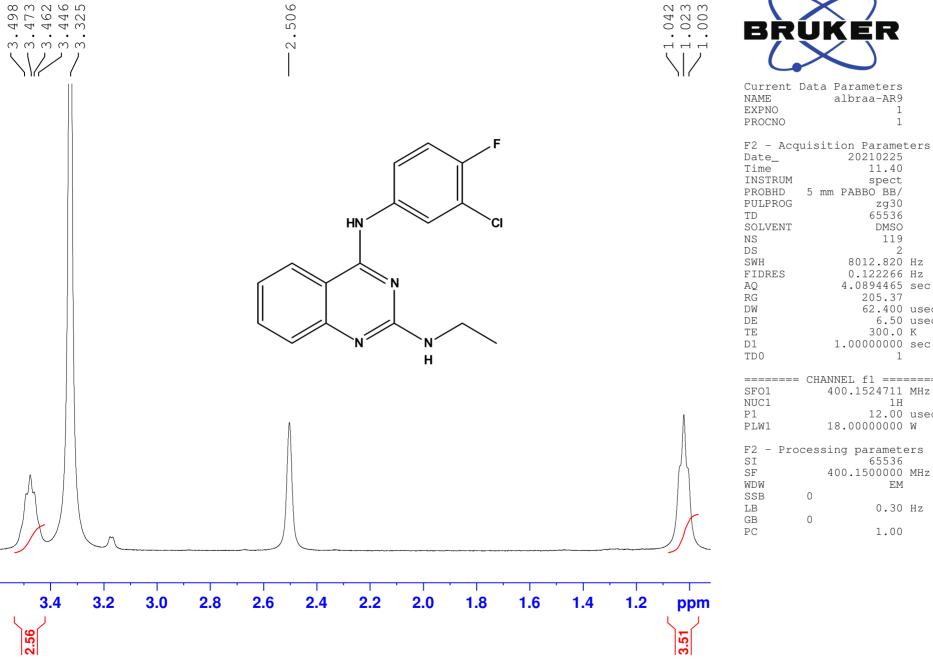




20210225 11.40 spect 5 mm PABBO BB/ zq30 65536 DMSO 119 2 8012.820 Hz 0.122266 Hz 4.0894465 sec 205.37 62.400 usec 6.50 usec 300.0 K 1.00000000 sec 1

====== CHANNEL f1 ====== 400.1524711 MHz 1H 12.00 usec 18.00000000 W

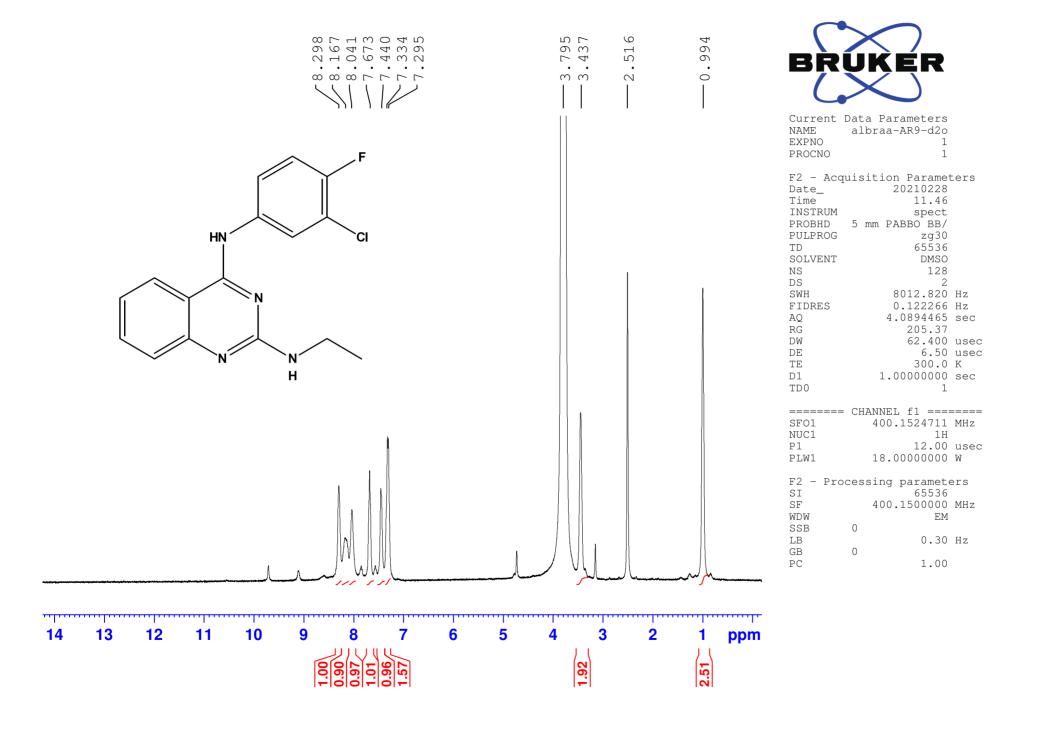
F2 - Processing parameters 65536 400.1500000 MHz ΕM 0.30 Hz

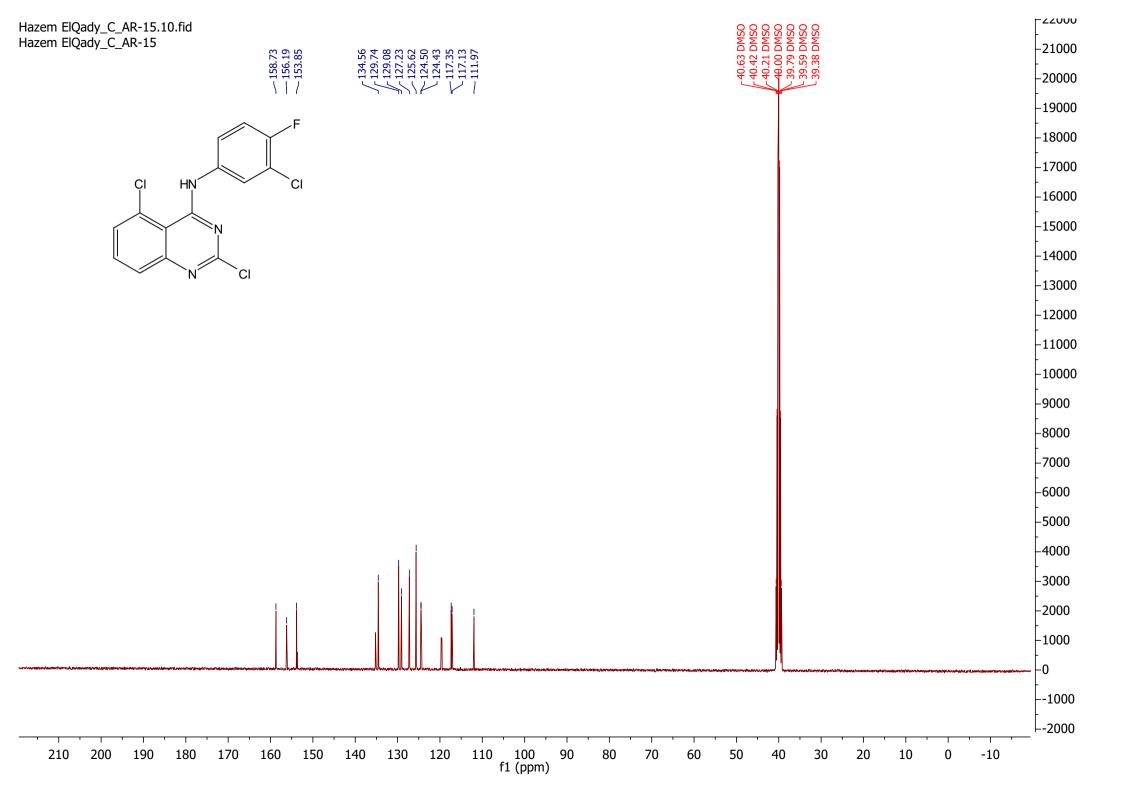


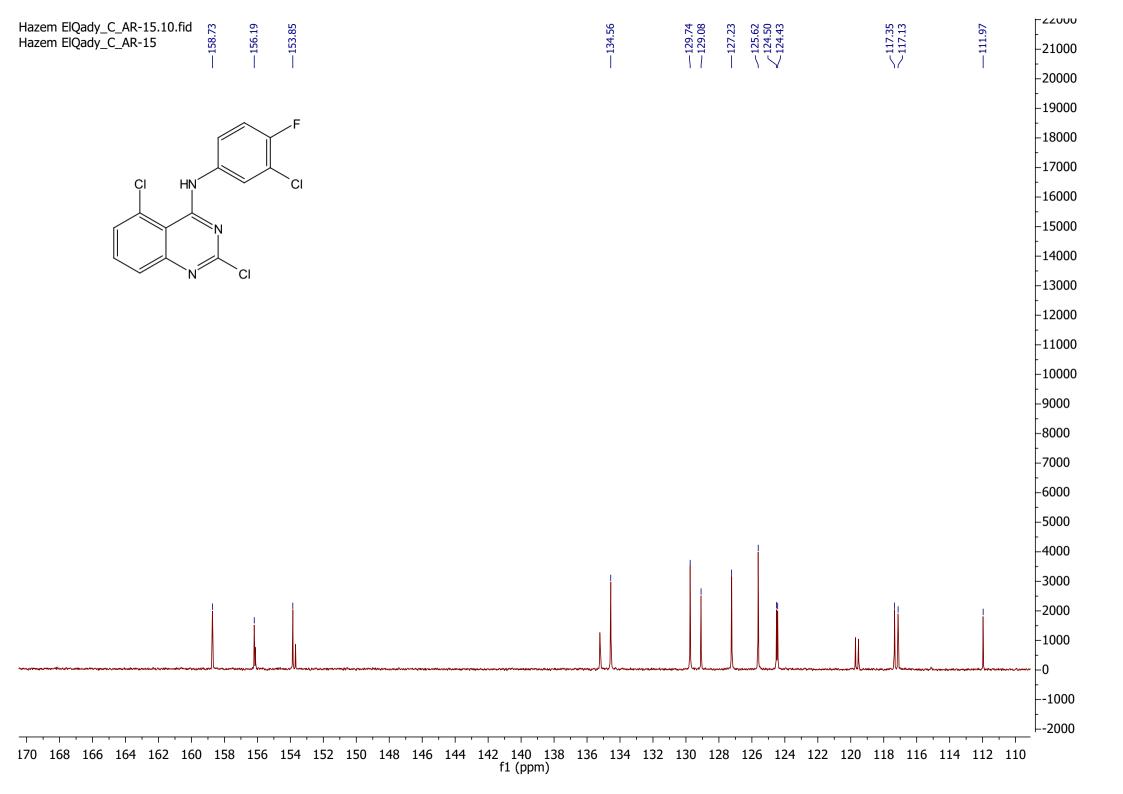


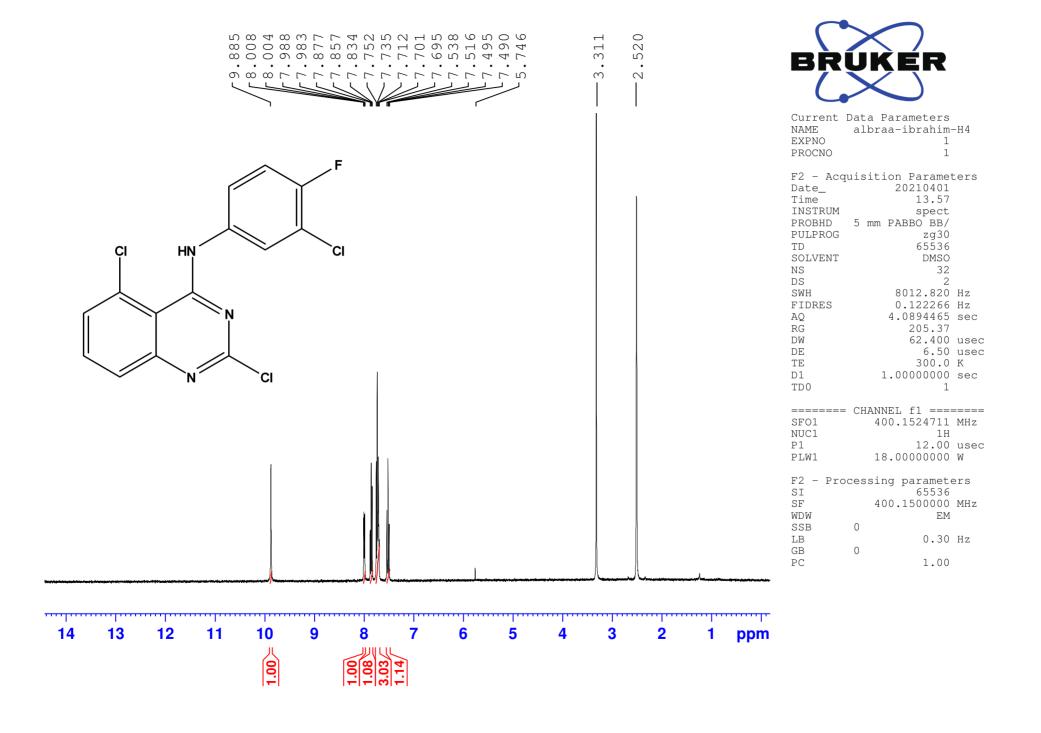
4.0894465 sec 62.400 usec 6.50 usec 1.00000000 sec

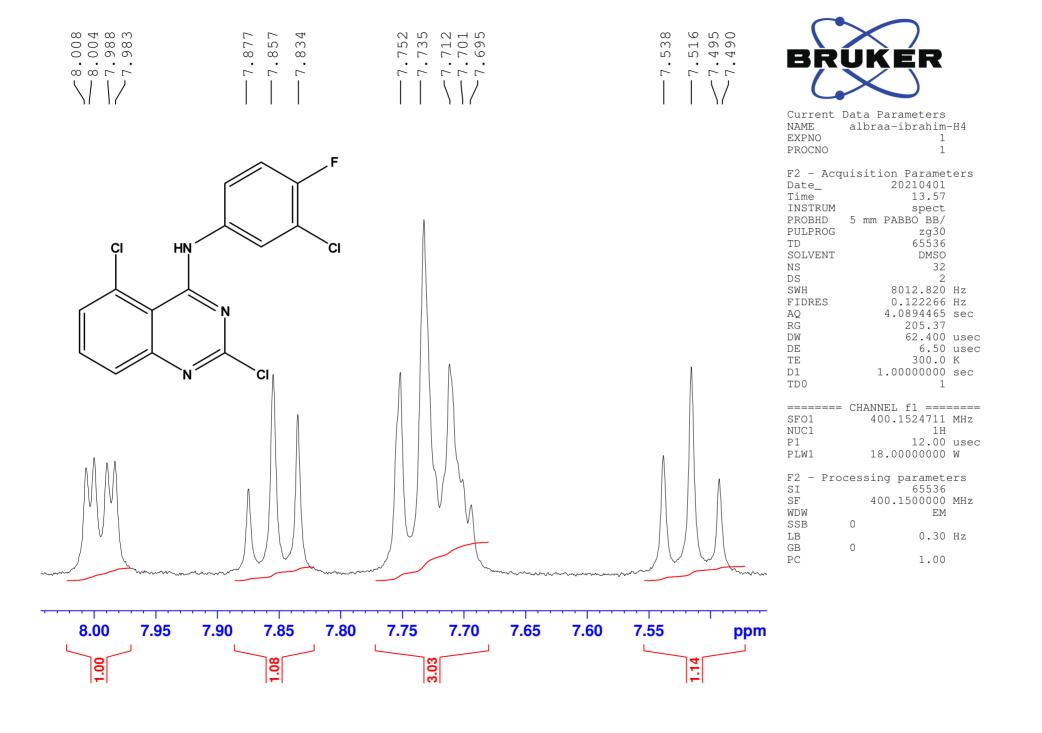
====== CHANNEL f1 ====== 400.1524711 MHz 12.00 usec

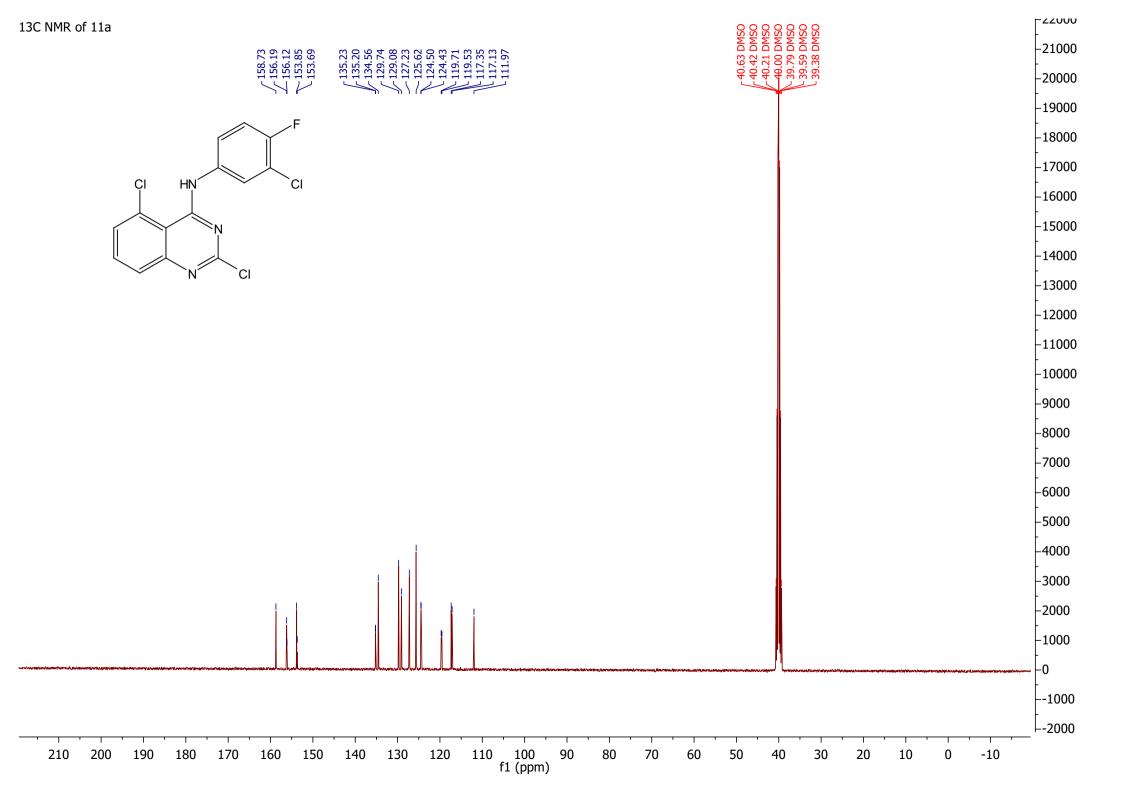


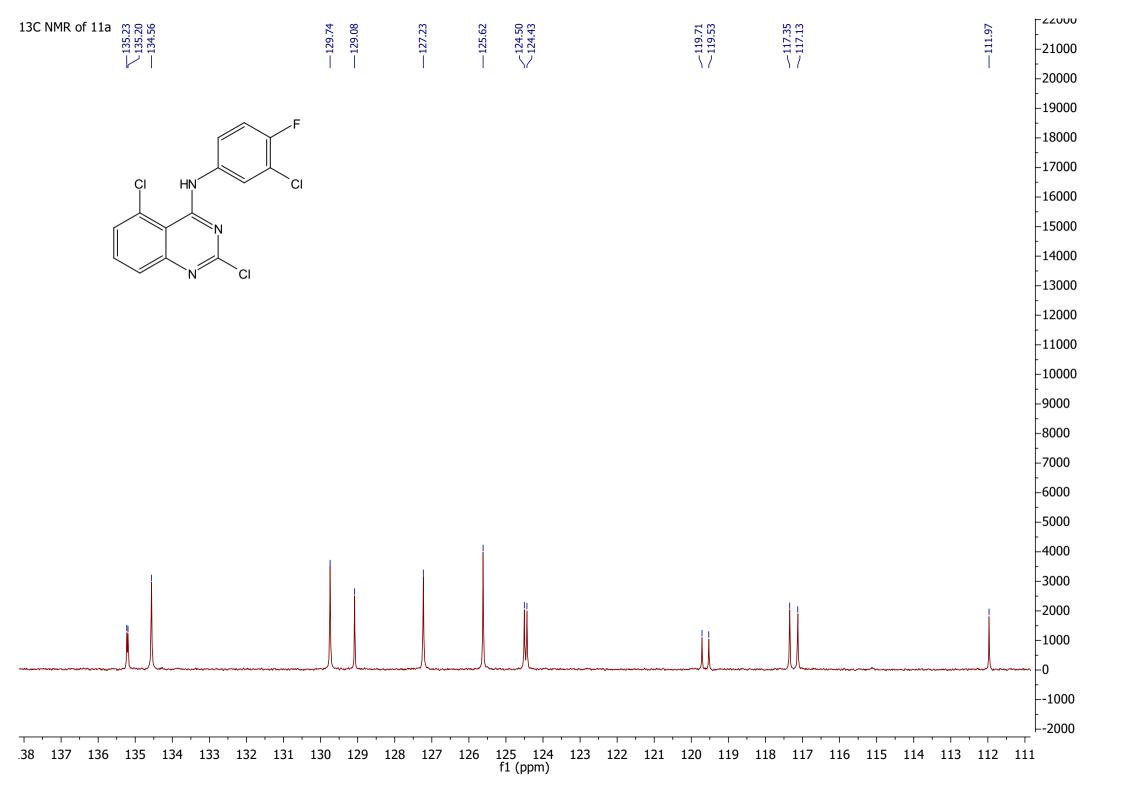


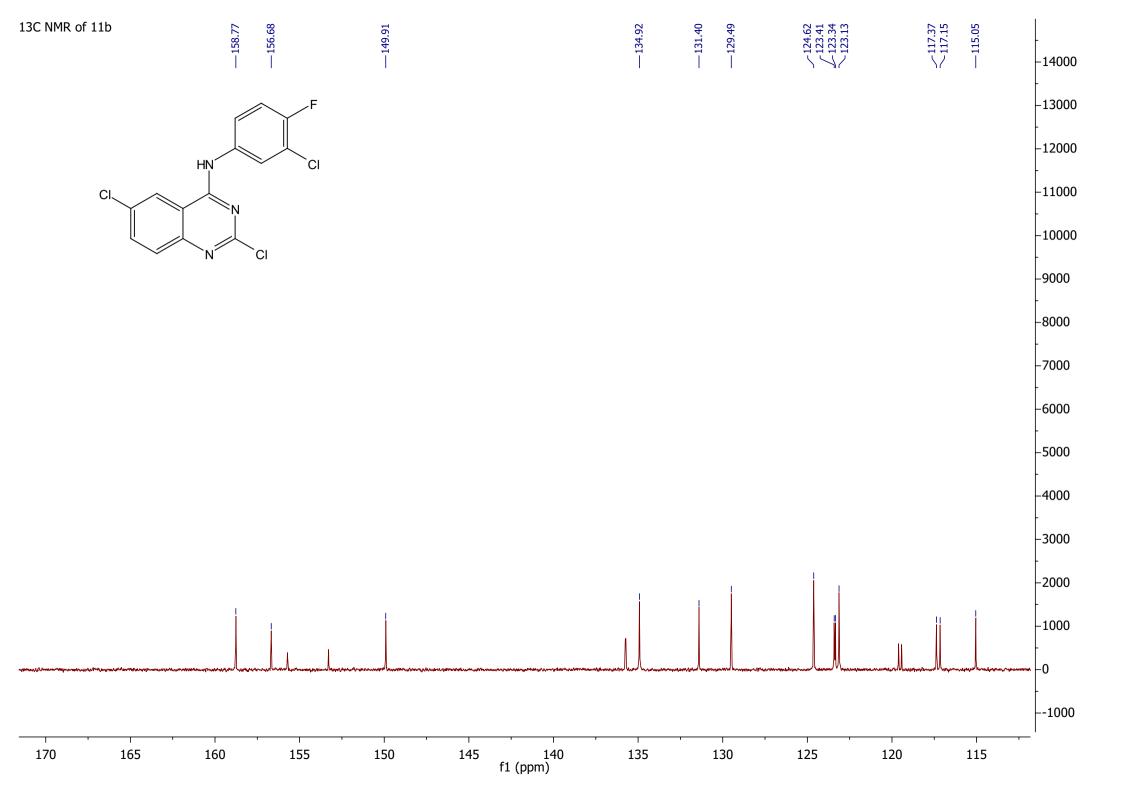


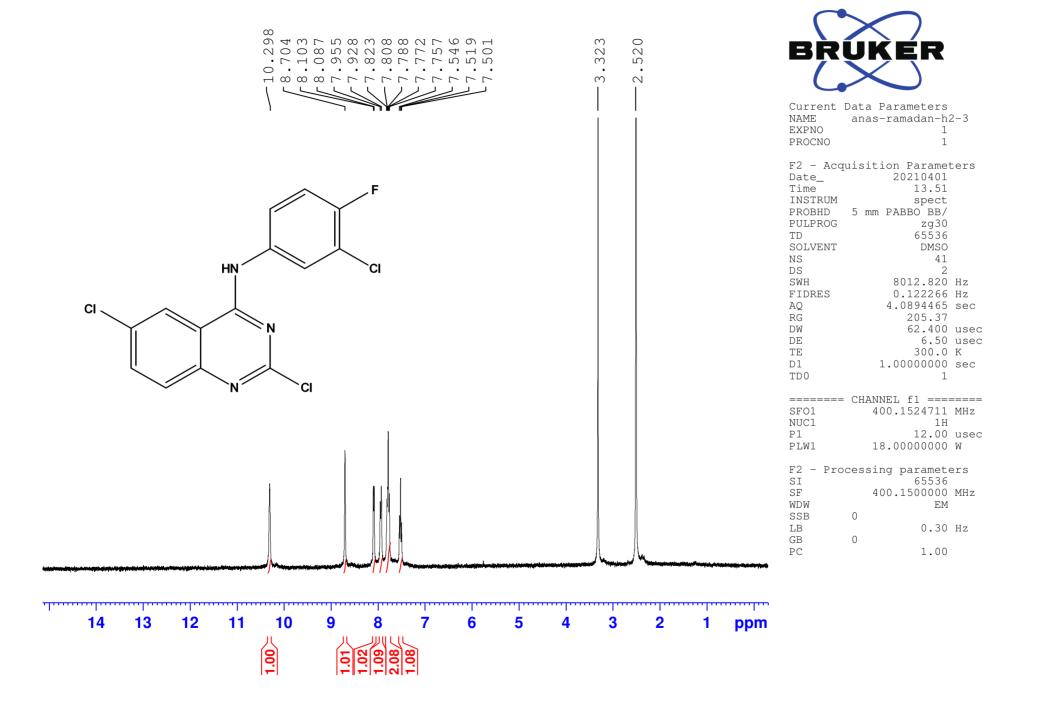


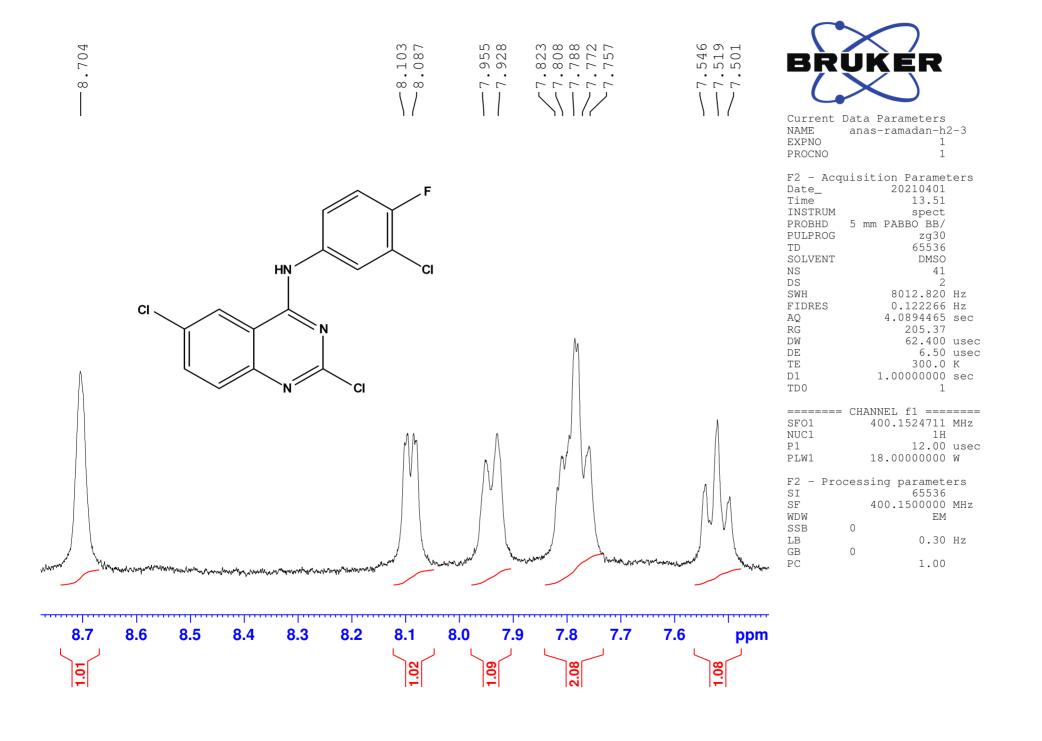


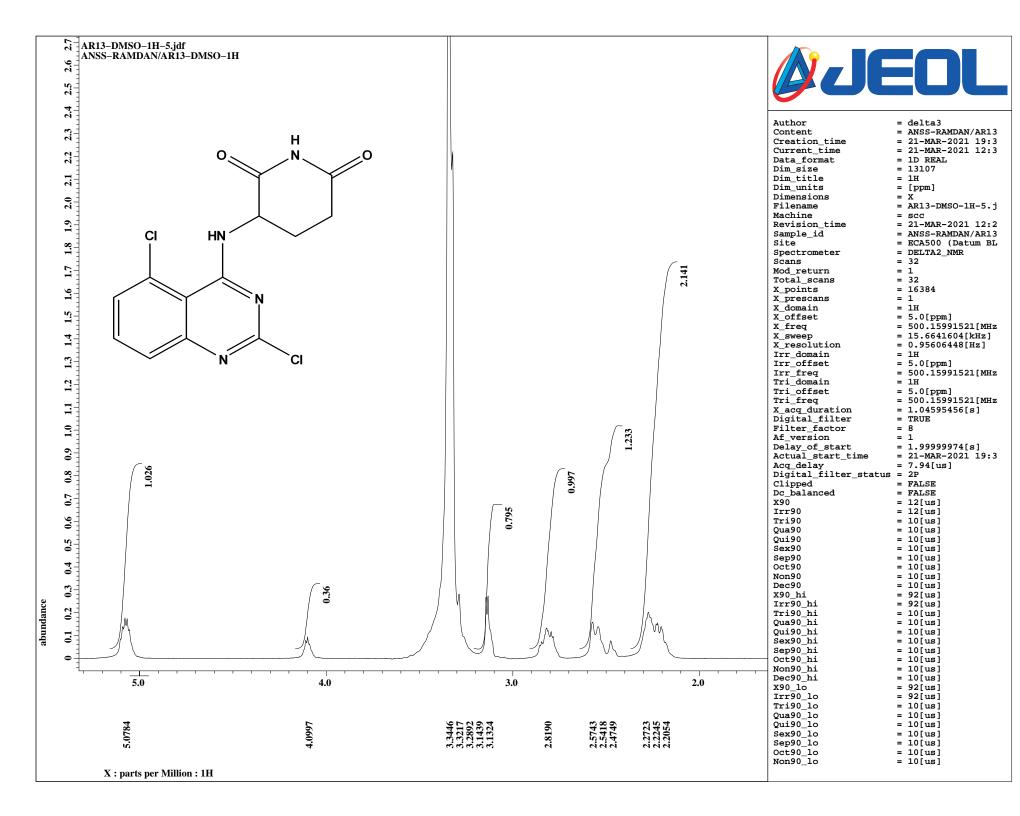


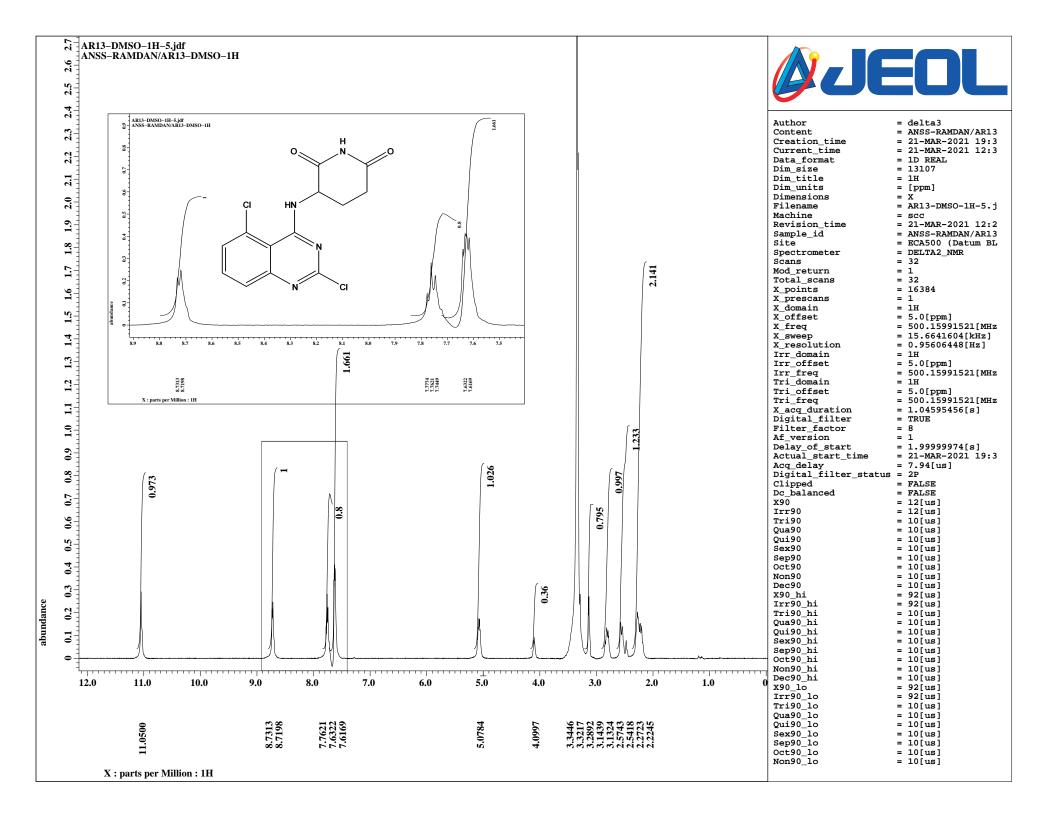


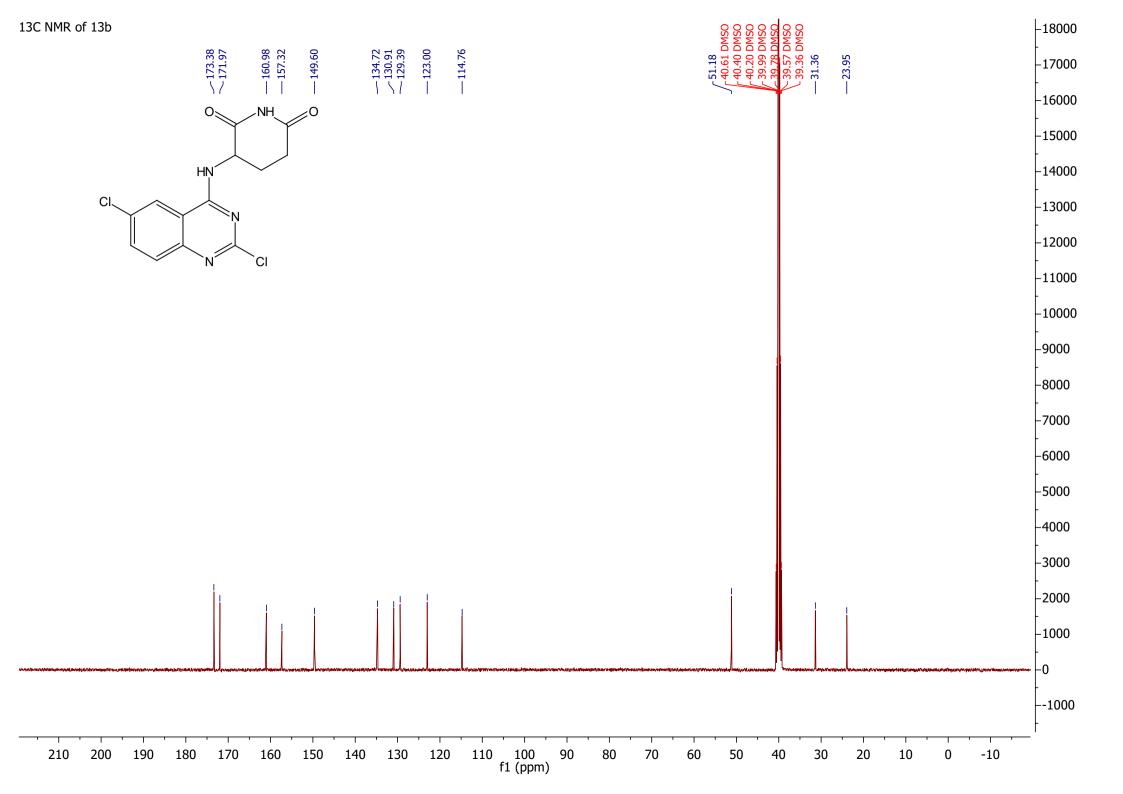


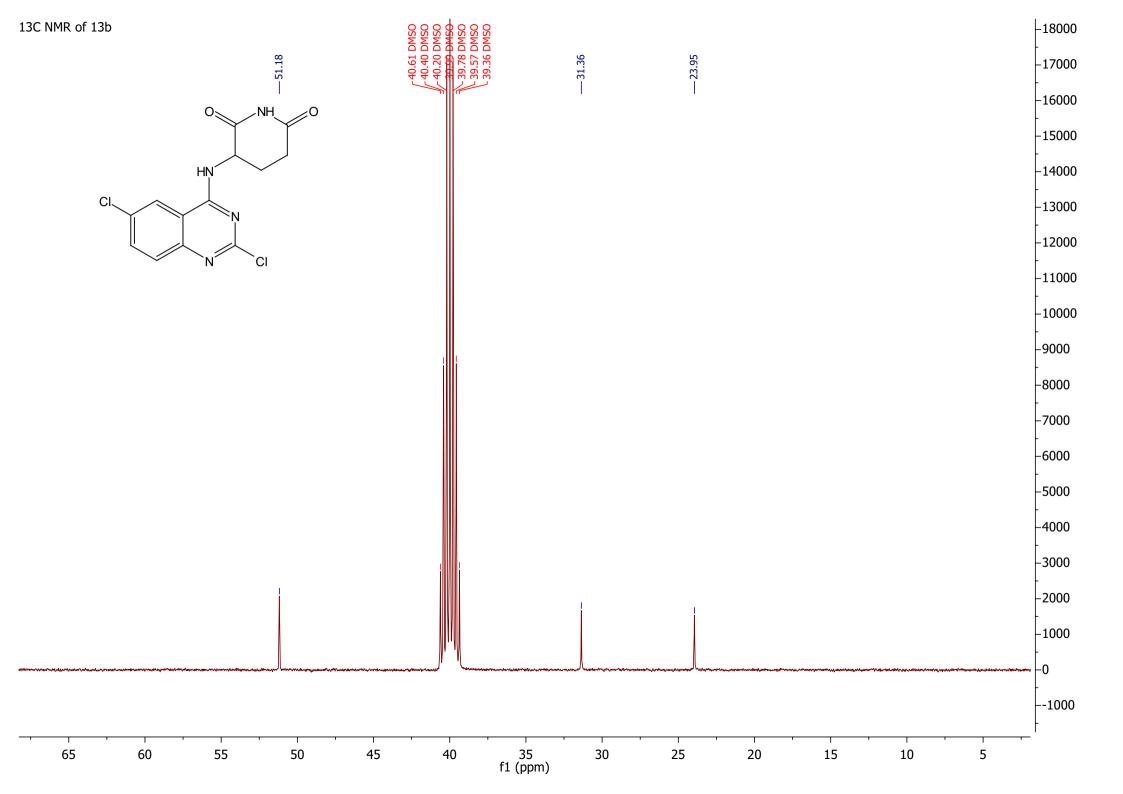


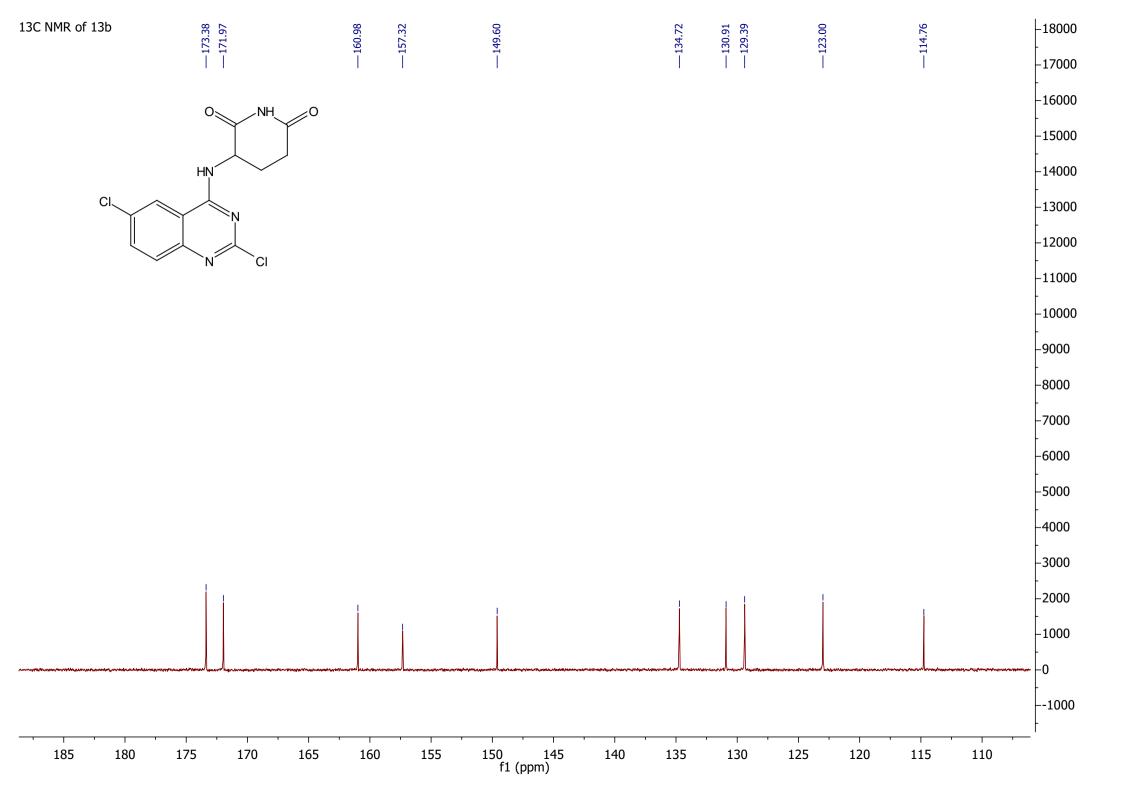


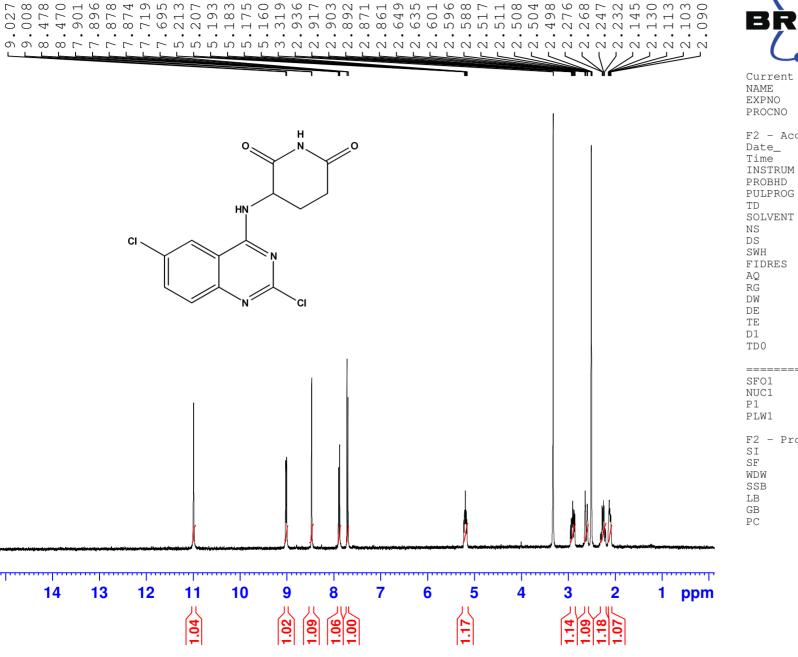












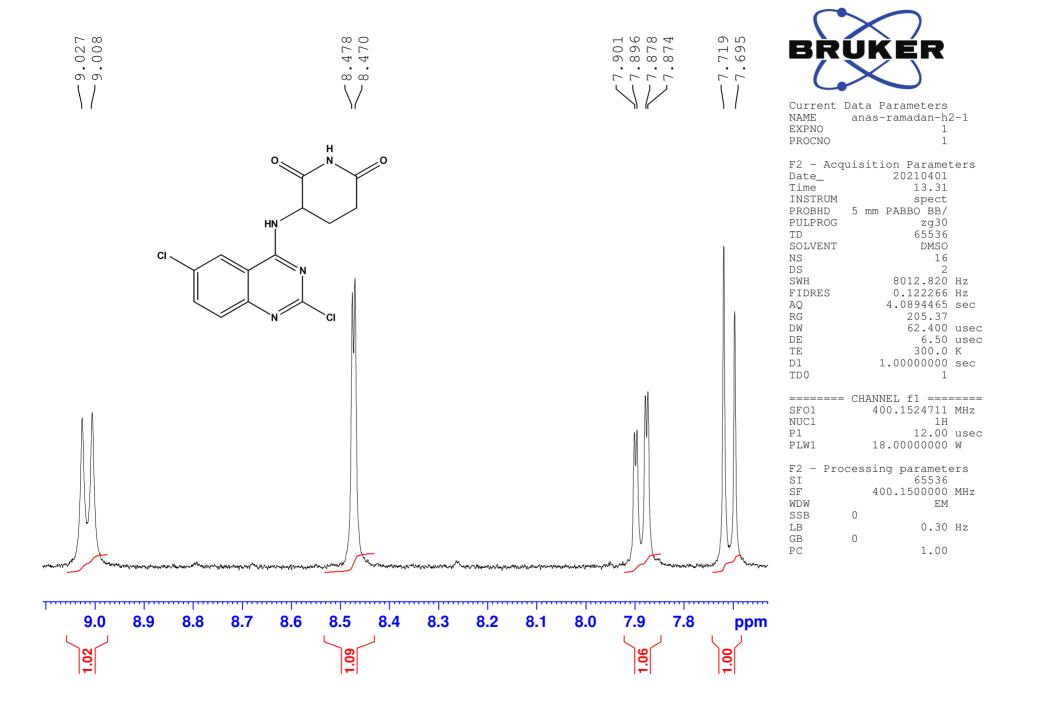


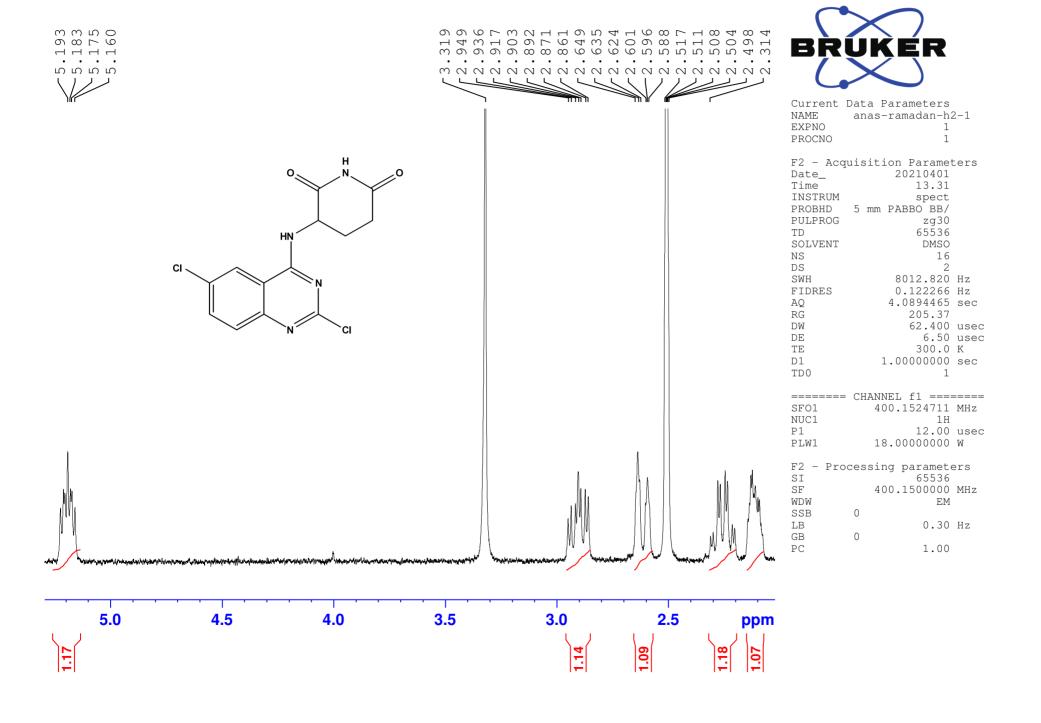
Current Data Parameters
NAME anas-ramadan-h2-1
EXPNO 1
PROCNO 1

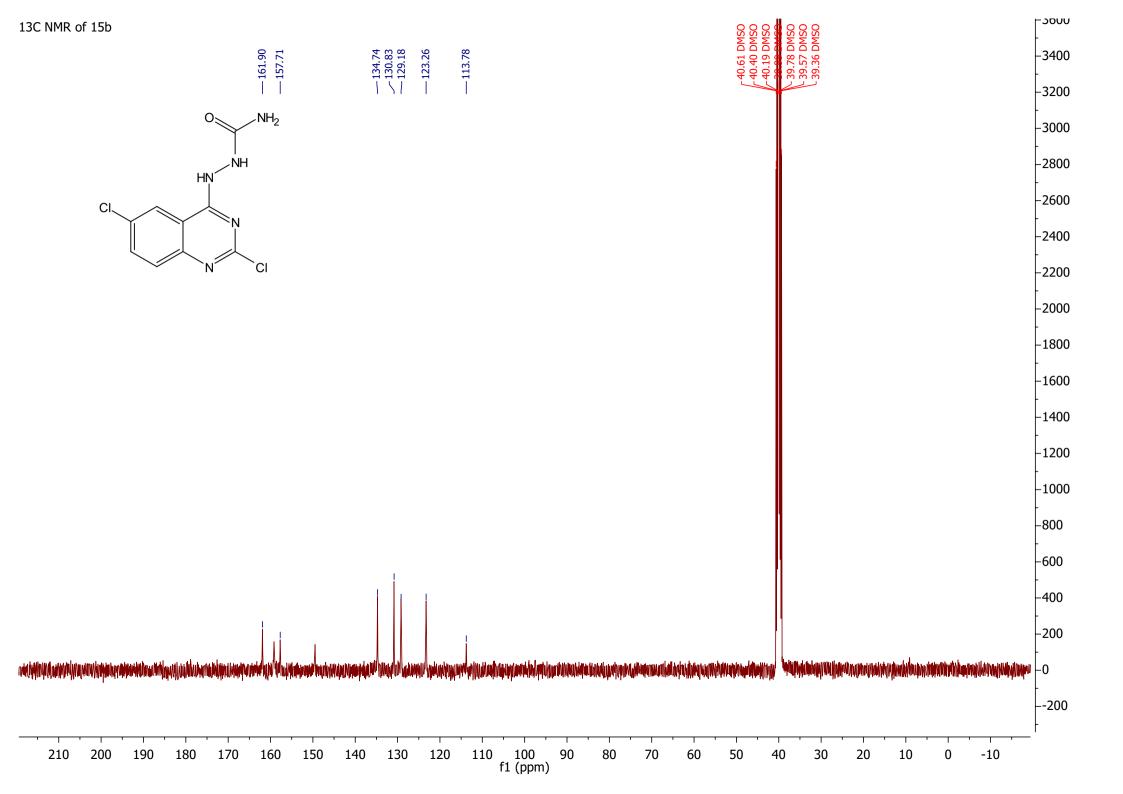
F2 - Acquisition Parameters 20210401 13.31 spect 5 mm PABBO BB/ zq30 65536 DMSO 16 2 8012.820 Hz 0.122266 Hz 4.0894465 sec 205.37 62.400 usec 6.50 usec 300.0 K 1.00000000 sec 1

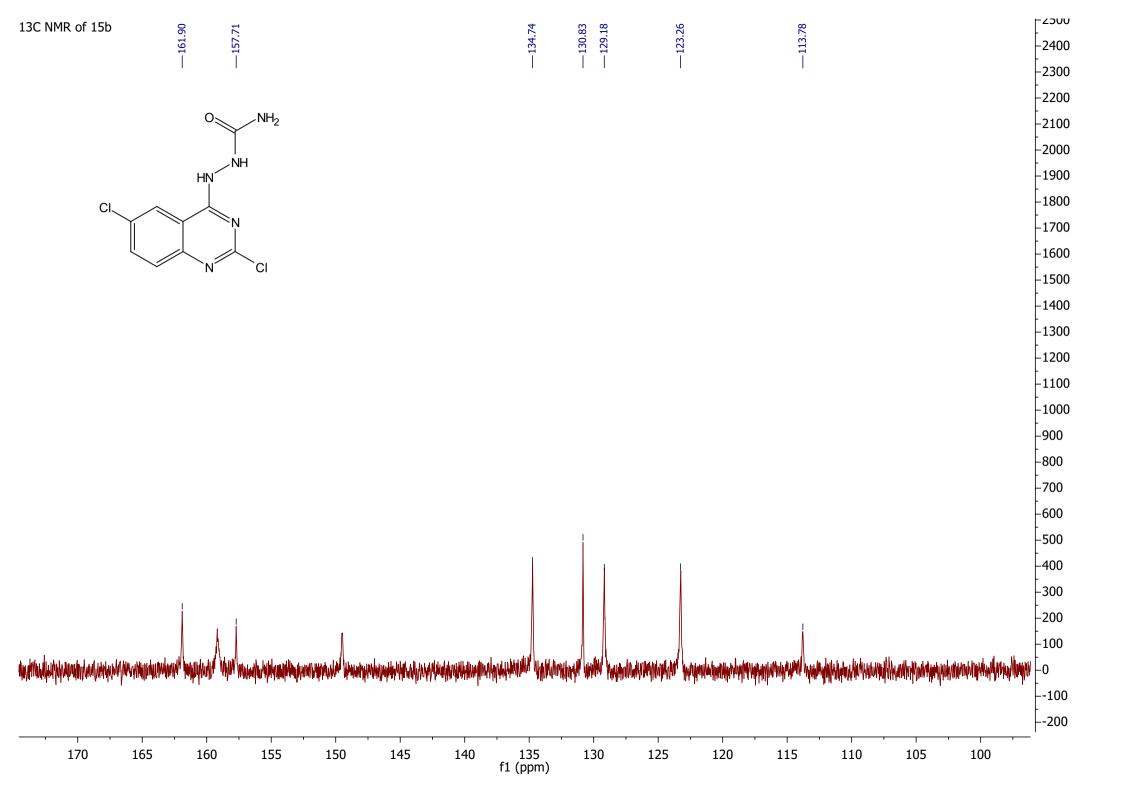
====== CHANNEL f1 ====== SFO1 400.1524711 MHz NUC1 1H P1 12.00 usec PLW1 18.00000000 W

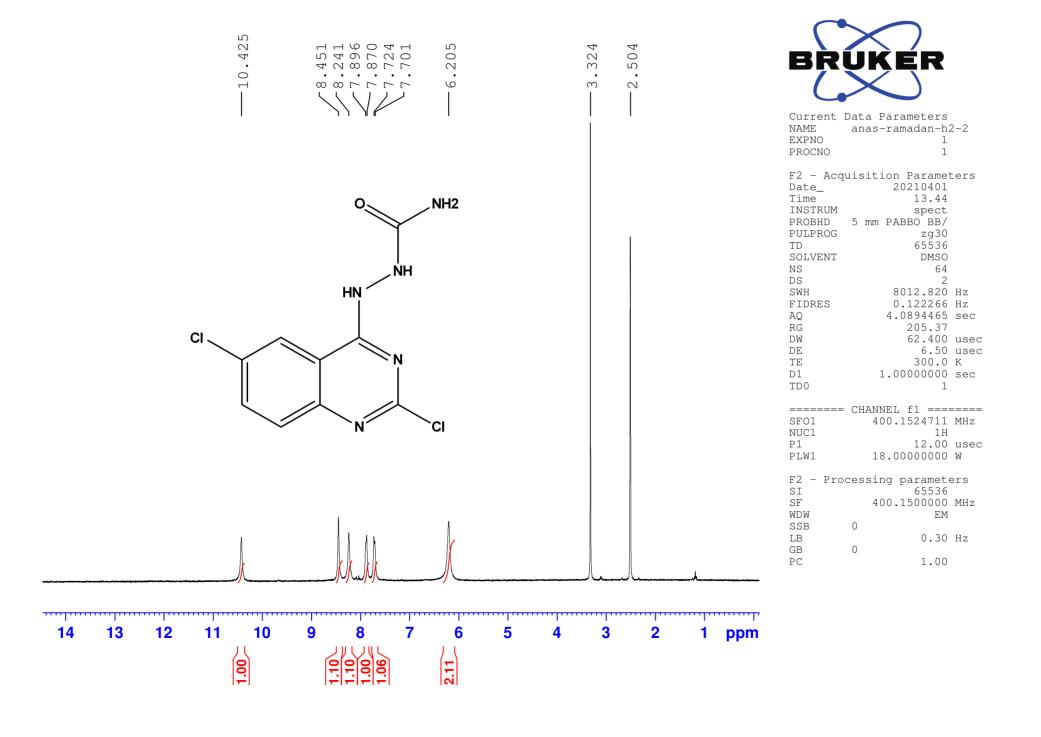
F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

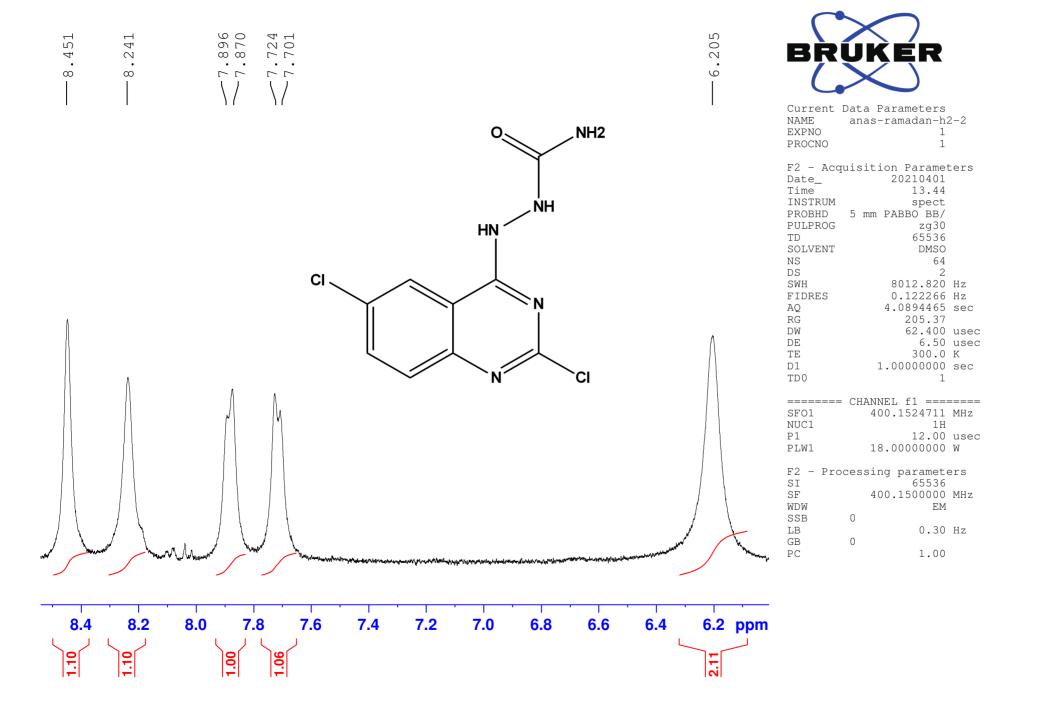


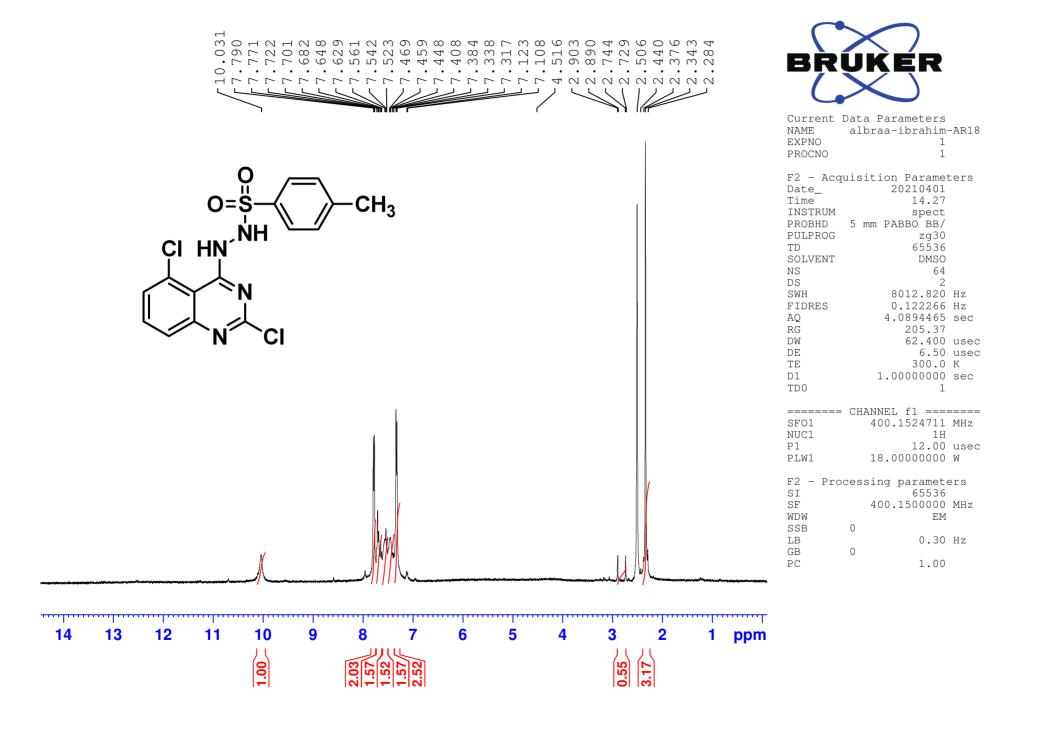








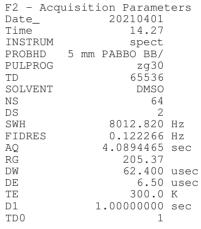




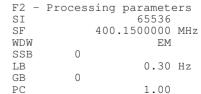


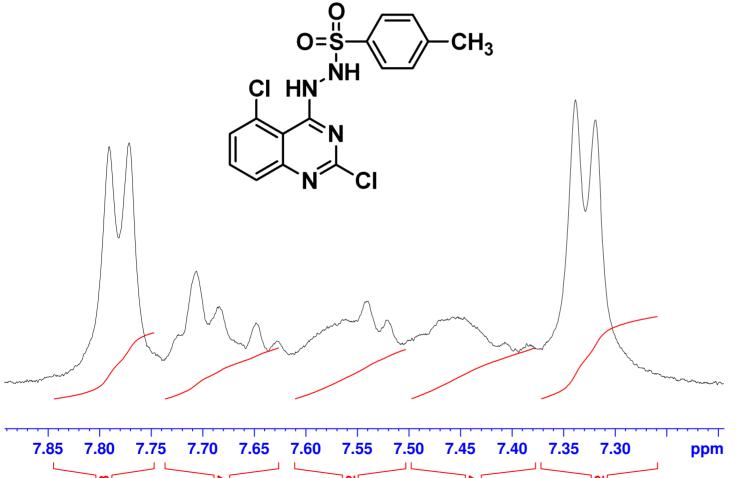
338

Current	Data	Parameters
NAME	albı	raa-ibrahim-AR18
EXPNO		1
PROCNO		1



======	CHANNEL fl ====	
SFO1	400.1524711	MHz
NUC1	1H	
P1	12.00	usec
PLW1	18.00000000	W





61 42 23

2 2 2

C

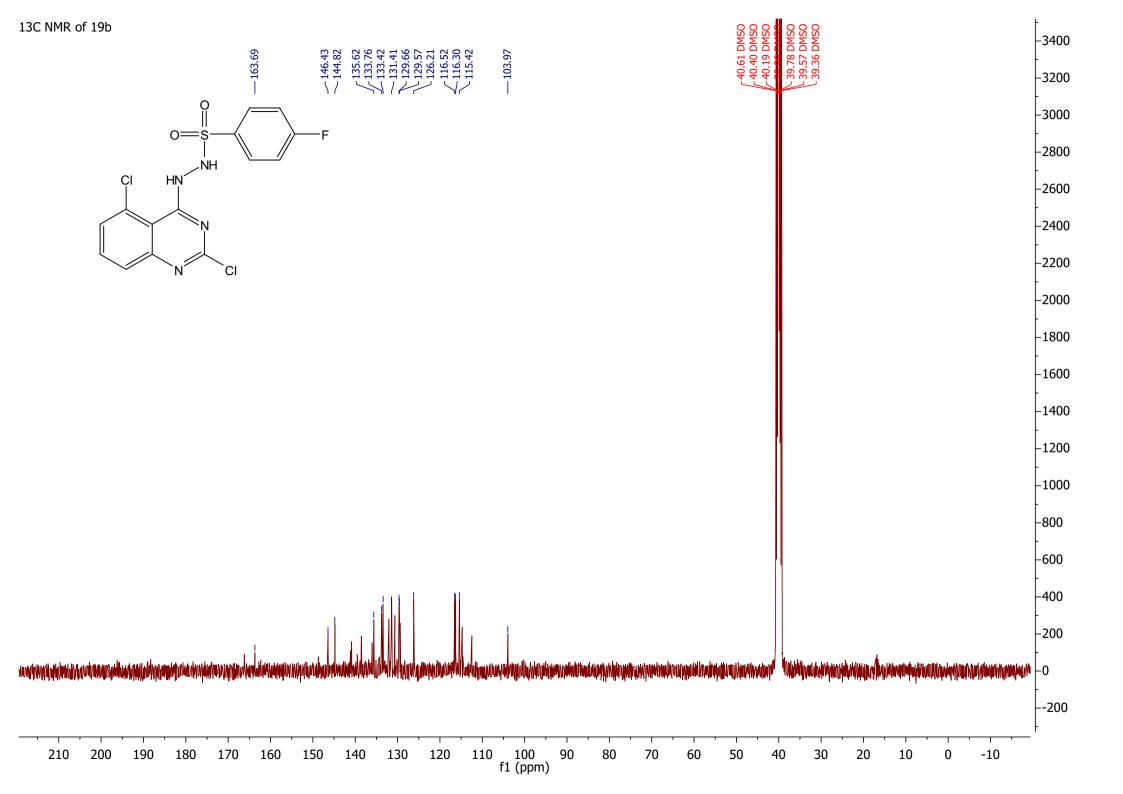
648 629

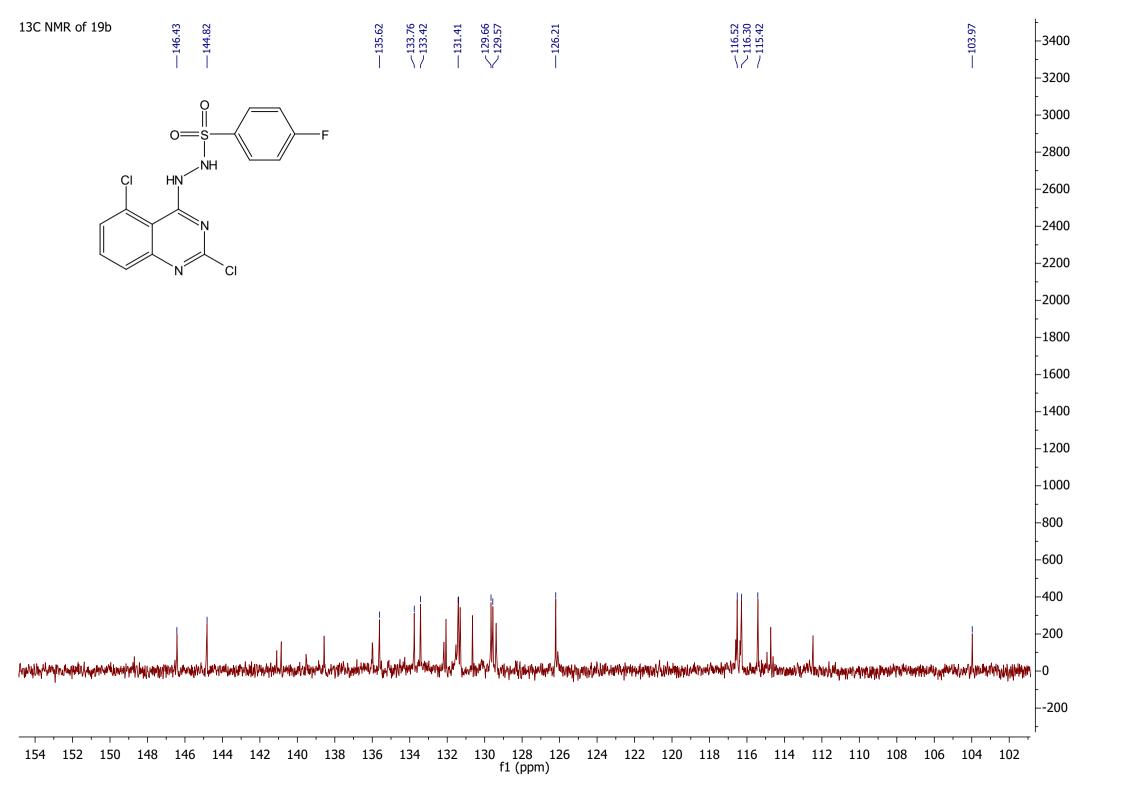
 $\omega \omega \omega$

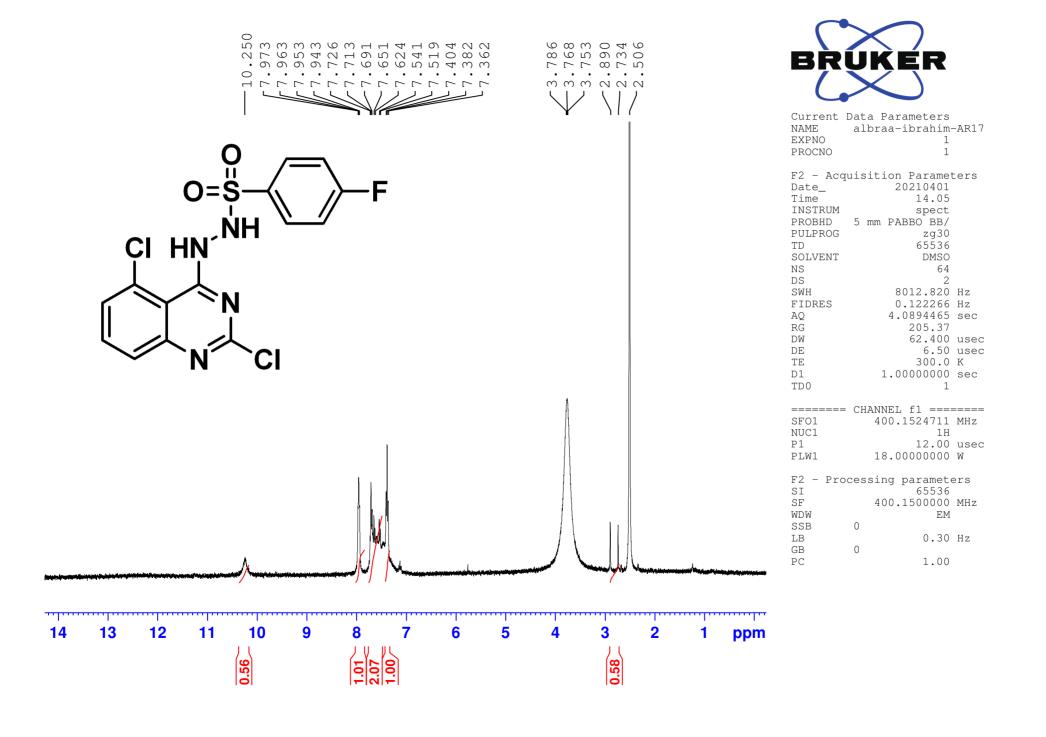
9 2 4

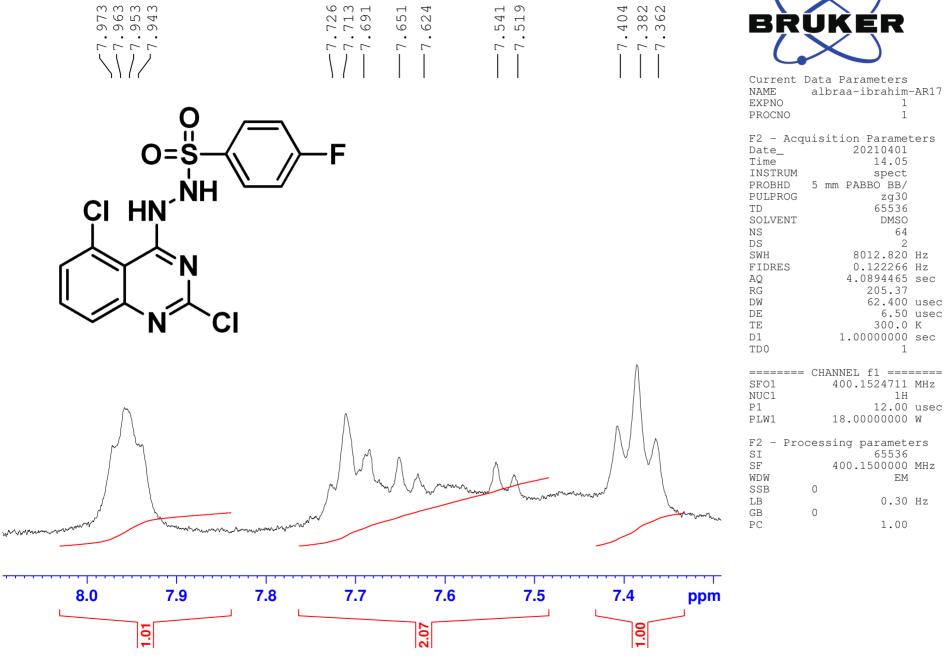
4 4 4

4 8









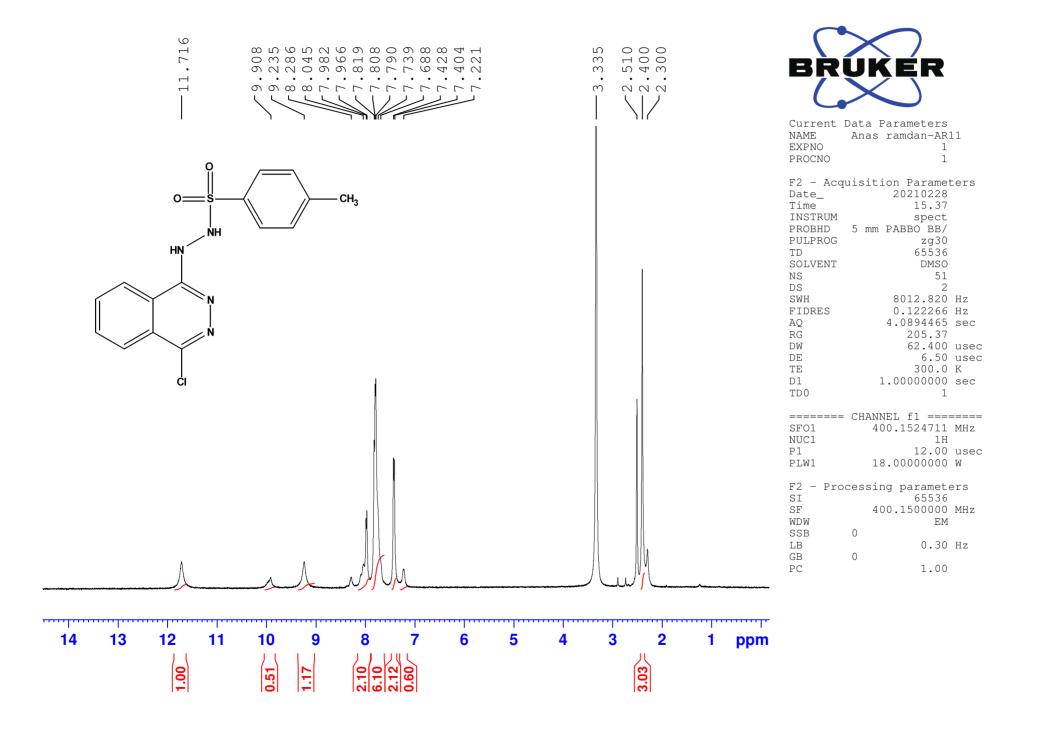


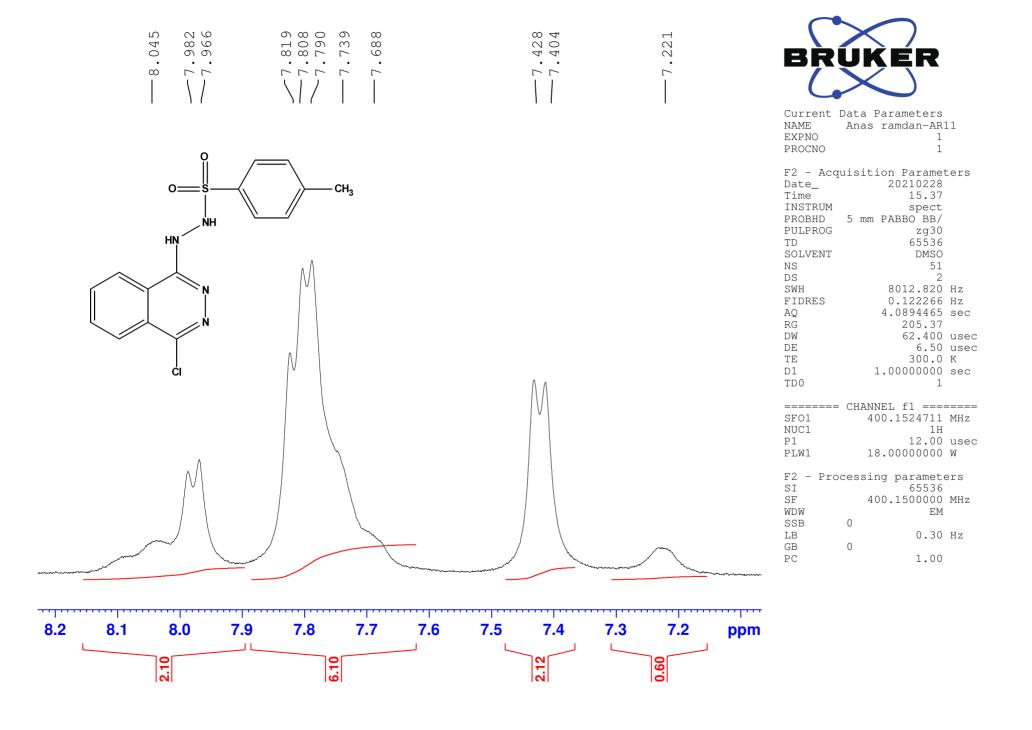
Current Data Parameters albraa-ibrahim-AR17 1 1

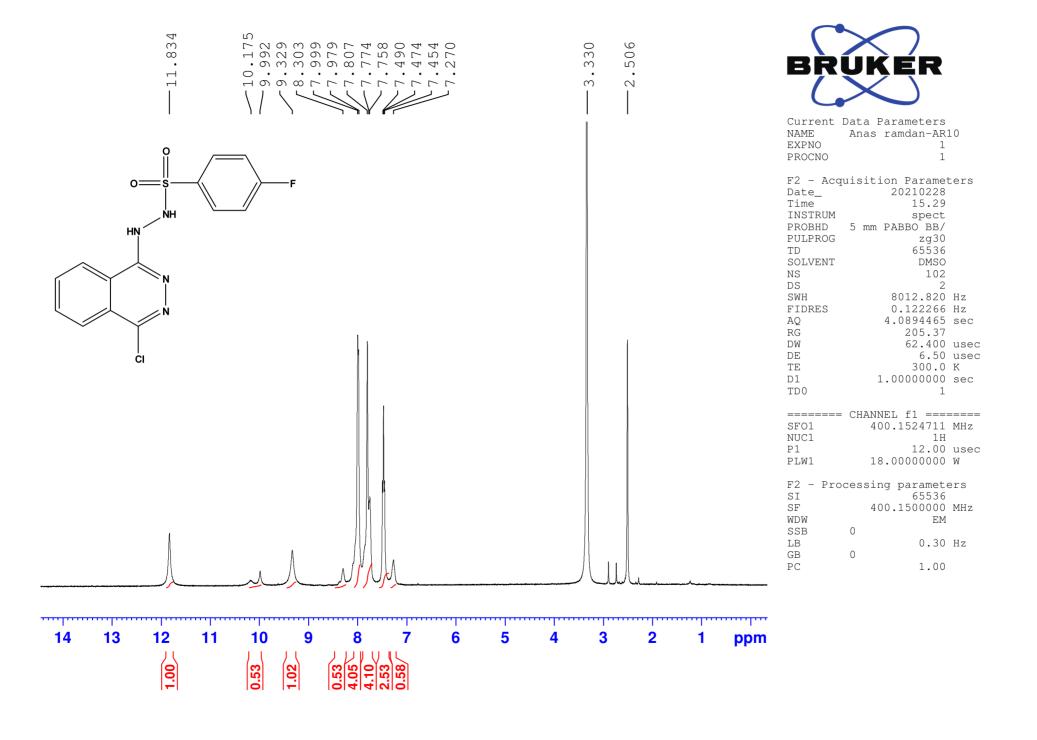
F2 - Acquisition Parameters 20210401 14.05 spect 5 mm PABBO BB/ zq30 65536 DMSO 64 2 8012.820 Hz 0.122266 Hz 4.0894465 sec 205.37 62.400 usec 6.50 usec 300.0 K 1.00000000 sec 1

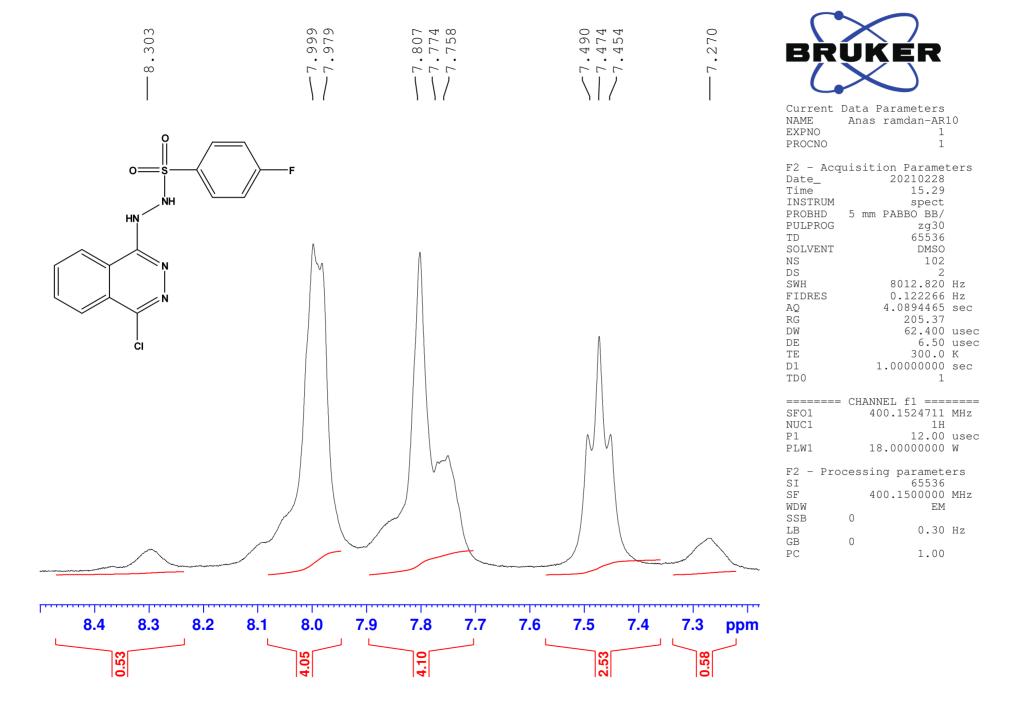
400.1524711 MHz 1H 12.00 usec 18.00000000 W

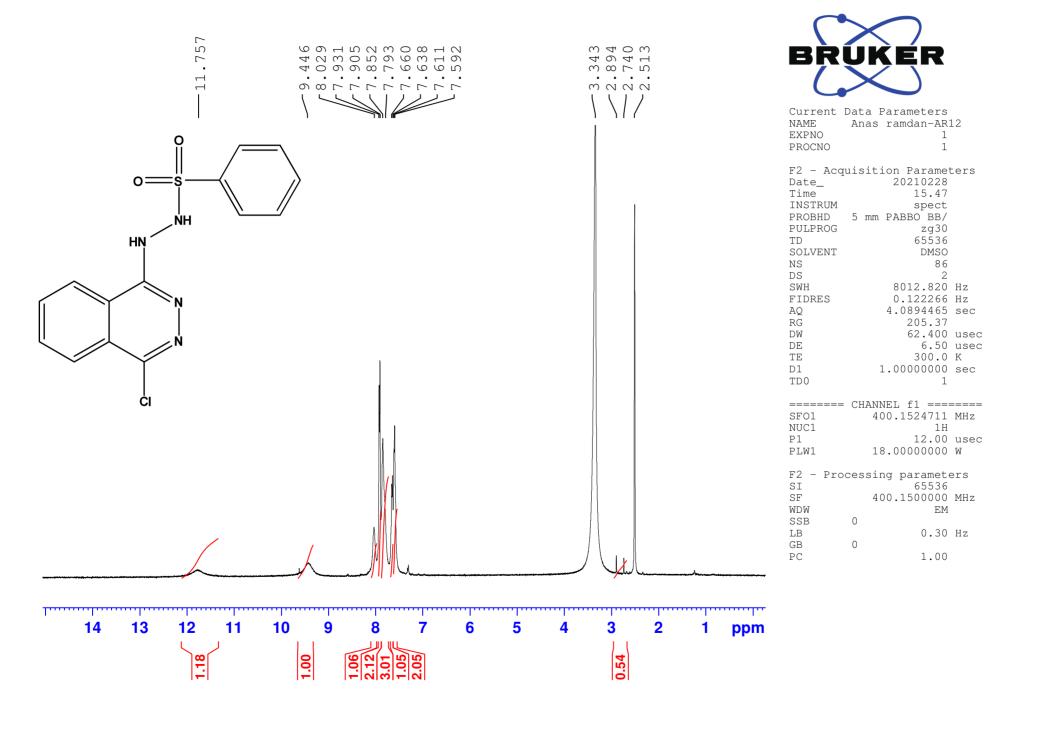
F2 - Processing parameters 65536 400.1500000 MHz ΕM 0.30 Hz 1.00

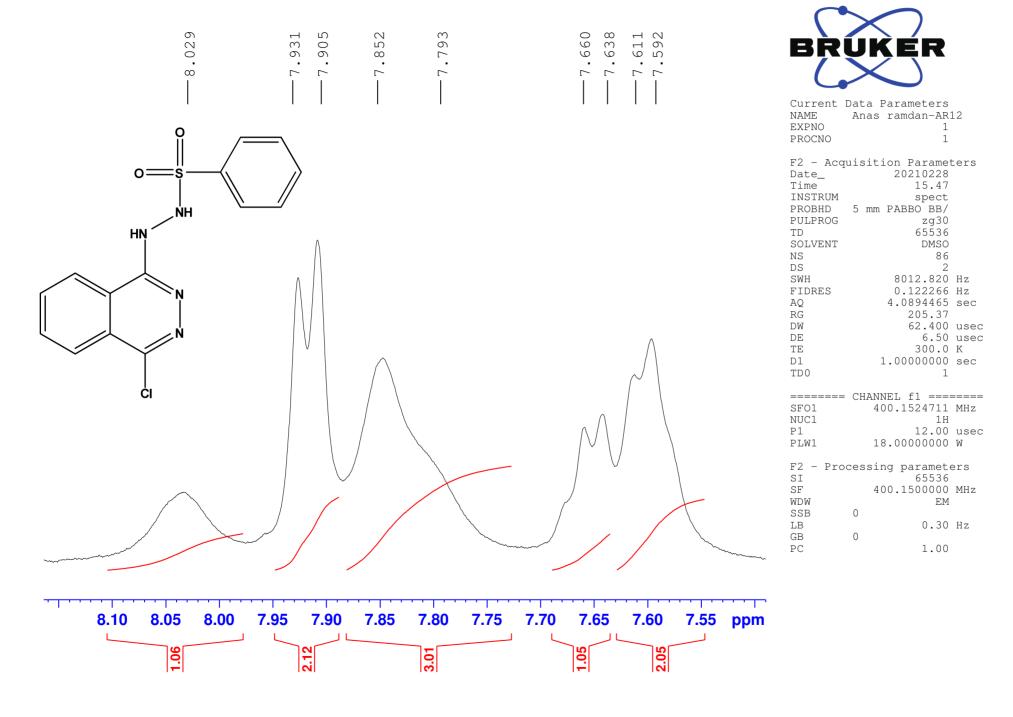






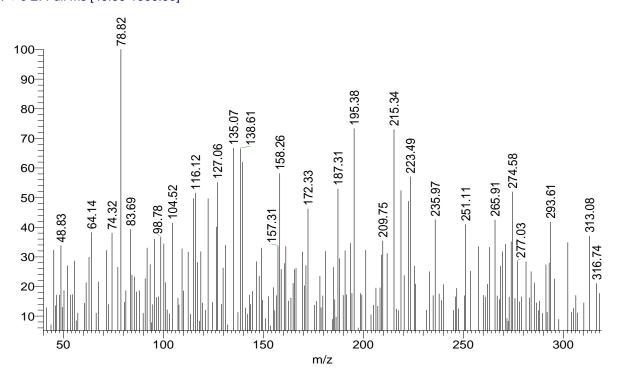




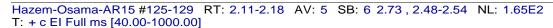


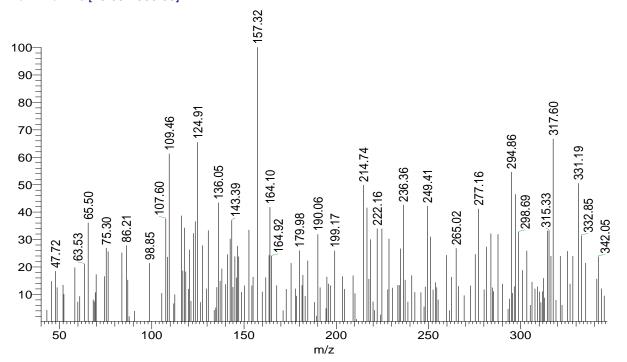
Mass spec. of 7b

Hazem-Osama-AR9 #54-63 RT: 0.92-1.07 AV: 10 SB: 2 2.44 , 2.44 NL: 8.08E1 T: + c El Full ms [40.00-1000.00]



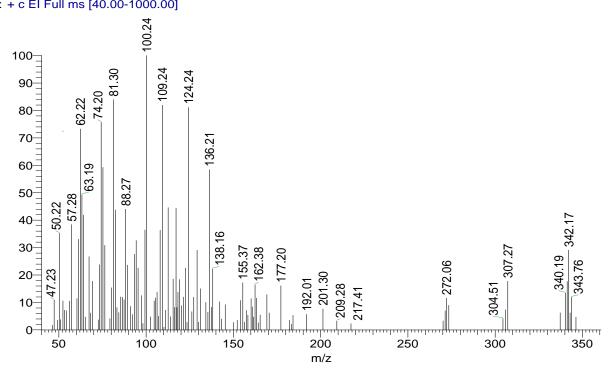
Mass spec. of 11a





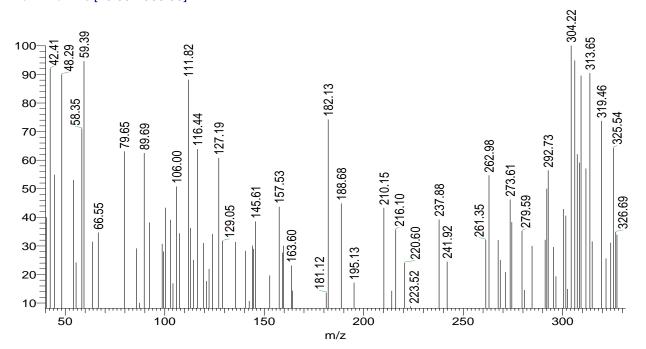
Mass spec. of 11b

Hazem-Osama-H2-3 #107 RT: 1.81 AV: 1 SB: 6 2.79 , 2.48-2.54 NL: 3.21E3 T: + c El Full ms [40.00-1000.00]



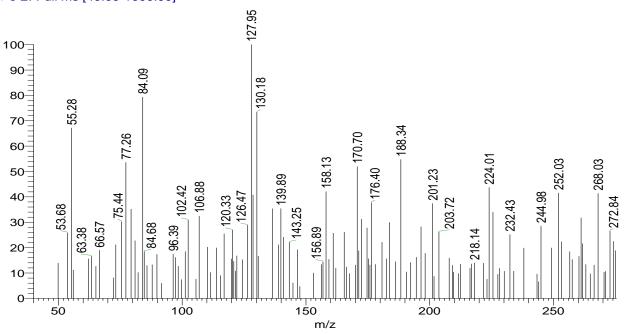
Mass spec. of 13b

Hazem-Osama-H2-1 #126 RT: 2.13 AV: 1 SB: 6 2.56 , 2.48-2.54 NL: 3.17E2 T: + c El Full ms [40.00-1000.00]

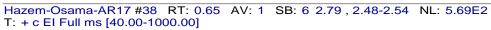


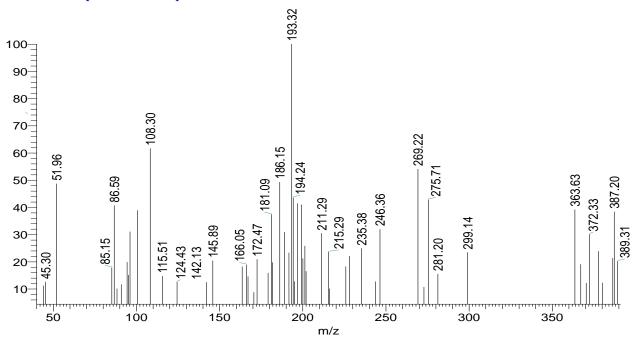
Mass spec. of 15b

Hazem-Osama-H2-2_20220731103618 #48-55 RT: 0.82-0.94 AV: 8 SB: 2 2.38 , 2.38 NL: 1.13E2 T: + c EI Full ms [40.00-1000.00]



Mass spec. of 19b





The Regional Center for Mycology and Biotechnology



Requester Data:

Dr. Anas Ramadan Abdel-Sattar Qotb Name: Authority: Faculty of Pharmacy, Al-Azhar University

Sample Data:

Sixteen samples had been submitted for elemental analysis.

Analysis Report:

Sample Code	C%	Н%	N%
AR1	58.41	5.40	14.61
AR2	53.72	5.49	16.98
AR3	55.35	5.98	16.30
AR5	59.63	5.81	13.97
AR6	65.18	4.96	12.19
AR7	54.73	2.85	13.81
AR8	59.38	4.12	18.32
AR9	60.91	4.42	17.87
AR10	47.85	2.97	15.62
AR11	51.89	3.85	16.32
AR12	50.34	3.45	16.90
AR13	48.28	3.34	17.41
AR14	39.88	2.75	25.53
AR15 (H4)	49.29	2.18	12.49
AR17	43.54	2.50	14.63
AR18	47.27	3.38	14.89

INVESTIGATOR

DIRECTOR

Al-Azhar University Campus - Nasr City, Cairo, Egypt. Tel: 0202 22620373 Fax: 0202 22620373

Website: http://www.azhar.edu.eg.htm * http://www.azhar.edu.eg/pages/fungi_center.htm Facebook: RCMB AZHAR

P.O. box mail: 11751 Nasr City Cairo, Egypt.

The Regional Center for Mycology and Biotechnology



Requester Data:

Name: Dr. Anas Ramadan Abdel-Sattar Qotb Authority: Faculty of Pharmacy, Al-Azhar University

Sample Data:

Thirteen samples had been submitted for elemental analysis.

Analysis Report:

Sample Code	C%	Н%	N%
H2-1	48.25	3.31	17.59
H2-2	39.90	2.78	25.91
H2-3	49.24	2.18	12.44
M1	49.71	4.27	13.96
M2	49.86	4.31	13.98
M3	52.29	4.50	14.47
M4	50.97	3.45	9.98
M5	50.99	3.48	9.93
M6	52.91	3.63	10.12
M7	52.79	3.68	10.04
Y1	54.60	4.41	10.28
Y2	54.56	4.49	10.02
Y3	56.78	4.62	10.68

INVESTIGATOR

DIRECTOR

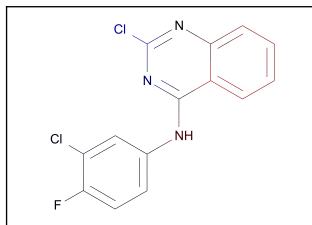
Al-Azhar University Campus - Nasr City, Cairo, Egypt.

E.mail:rcmb@azhar.edu.eg

Website: http://www.azhar.edu.eg.htm * http://www.azhar.edu.eg/pages/fungi_center.htm

Facebook: RCMB AZHAR P.O. box mail: 11751 Nasr City Cairo, Egypt.

Toxicity report



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mutagen
Probability: 0.75

Enrichment: 1.34
Bayesian Score: 0.611

Mahalanobis Distance: 7.68

Mahalanobis Distance p-value: 0.998

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	61421-82-7	4947-27-7	61421-83-8
Structure	N N N N N N N N N N N N N N N N N N N	NH O	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	0.472	0.501	0.527
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

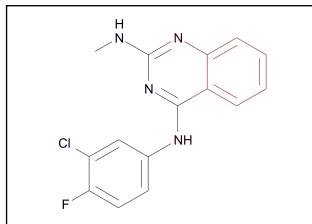
Model Applicability

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.

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set	
SCFP_12	1651620003	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690	

SCFP_12	-1379673609	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.375	934 out of 1123
SCFP_12	112346096	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
	Top Feat	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	[*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	CI NH NH STATE STA	-0.887	2 out of 11
SCFP_12	1328855840	[*]:[cH]:[*]	-0.496	1 out of 4



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Mutagen
Probability: 0.781
Enrichment: 1.4
Bayesian Score: 2.66
Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.236

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

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

Enrichment: An 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 Simila	r Compounds		
Name	106063-42-7	75776-00-0	61421-83-8
Structure	N NH	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	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.433	0.477	0.493
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.

Feature Co	ntribution			
	Top fea	ntures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1551584806	[*]N[c]1:n:[*]:[c](:[*]):[c](:[cH]:[*]):n :1	0.464	46 out of 50

SCFP_12	1651620003	CI NH H *][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690
SCFP_12	-1379673609	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*];[c]: 1:[*]	0.375	934 out of 1123
	Top Featur	es for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-760164567	CI NH [*][c]1:[*]:[c](:[*]) :n:[c](NC):n:1	-1.35	0 out of 5
SCFP_12	1328855840	CI NH H [*]:[cH]:[c](N[c](:n: [*]):[c](:[*]):[*]): [cH]:[*]	-0.496	1 out of 4
SCFP_12	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.465	154 out of 430

H N N NH

C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.717
Enrichment: 1.28
Bayesian Score: -1.04
Mahalanobis Distance: 9.34

Mahalanobis Distance p-value: 0.673

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

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

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

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

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

Structural Similar Compounds

Name	106063-42-7	75776-00-0	DICLOFENAC
Structure	NH NH NH	N N N N N N N N N N N N N N N N N N N	OH O
Actual Endpoint	Mutagen	Mutagen	Non-Mutagen
Predicted Endpoint	Mutagen	Mutagen	Non-Mutagen
Distance	0.482	0.505	0.543
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	PDR 1994

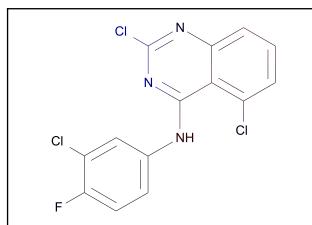
Model Applicability

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

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	-1551584806	[*]N[c]1:n:[*]:[c](:[*]):[c](:[cH]:[*]):n :1	0.464	46 out of 50	

SCFP_12	1651620003	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690
SCFP_12	-1379673609	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[e]: 1:[*]	0.375	934 out of 1123
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-760164567	[*][c]1:[*]:[c](:[*]) :n:[c](NC):n:1	-1.35	0 out of 5
SCFP_12	-150051645	[*]:n:[c](NCC):n:[*]	-0.998	0 out of 3
SCFP_12	-1301088181	[*][c]1:[*]:[c](:[*]) :n:[c](NCC):n:1	-0.998	0 out of 3



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.693
Enrichment: 1.24
Bayesian Score: -2.02
Mahalanobis Distance: 7.76

Mahalanobis Distance p-value: 0.997

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	61421-82-7	21232-47-3	4947-27-7
Structure	N. The NH	CI N N N X.	NH NH

Actual Endpoint Mutagen Non-Mutagen Mutagen Predicted Endpoint Mutagen Non-Mutagen Mutagen 0.570 0.545 0.547 Distance Reference Kazius et. al., J. Med. Kazius et. al., J. Med. Kazius et. al., J. Med. Chem. (2005) 48, 312-320 Chem. (2005) 48, 312-320 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.

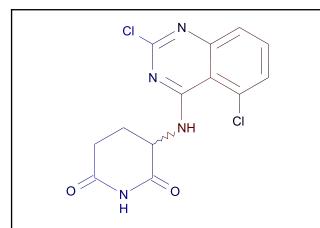
1. All properties and OPS components are within expected ranges.

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

SCFP_12	-1381862798	CI NH CI NH CI (1):[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]	0.314	339 out of 433
SCFP_12	403834996	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.314	138 out of 176
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	[*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	CI NH CI NH CI [*]:n:[c](CI):n:[*]	-0.887	2 out of 11
SCFP_12	1328855840	[*]:[cH]:[*]	-0.496	1 out of 4

0.534

EMIC



 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.637
Enrichment: 1.14
Bayesian Score: -3.96
Mahalanobis Distance: 8.67

Mahalanobis Distance p-value: 0.923

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

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

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

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

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

Structural Similar Compounds				
Name	97919-22-7	491-80-5	2-ETHOXYMETHYL-1;3- DIHYDROXYANTHRAQUI NONE	
Structure	NH 2	HOWNOH	O OH 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.

0.534

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

0.527

Feature Contribution					
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set	
SCFP_12	-1231050790	[*]C([*])N[c]1:n:[c]([*]):[*]:[c](:[*]):[c]:1:[c]([*]):[*]	0.388	3 out of 3	

SCFP_12	112346096	CI NH CI NH CI (I CH]:[*][c](:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
SCFP_12	-1817806383	[]CC(N[1])C(=[1])[1]	0.337	2 out of 2
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	CI NH CI NH CI (I CI	-1.82	0 out of 9
SCFP_12	395945879	CI NH CI NH CI H CI	-1.19	0 out of 4
SCFP_12	-1286592310	CI NH	-0.887	2 out of 11

CI NH NH HN O NH 2

C₀H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Mutagen

Probability: 0.717 Enrichment: 1.28 Bayesian Score: -1.04 Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.0482

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

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

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

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

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

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

Name	6027-98-1	Picloram	520-36-5
Structure	HCI H ₂ O H ₂ O	O H CI	HO to OH
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.560	0.561	0.564
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Mut. Res. 204: 17-115; 1988	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	-1977951815	CI NH	0.362	7 out of 8	

SCFP_12	112346096	CI NH NH ONH ₂ [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
SCFP_12	-1381862798	CI HN NH O NH ₂ [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.314	339 out of 433
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	CI NH CI NH CI NH2 [*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	CI NH CI NH ON NH 2 [*]:n:[c](CI):n:[*]	-0.887	2 out of 11
SCFP_12	-116109291	CI NH CI NH CI NH. NH. ST. [c] (CI):[cH]:[cH]:[cH]:[c](CI):[cH]:1	-0.502	76 out of 221

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

CI NH NH CI F

C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.677
Enrichment: 1.21
Bayesian Score: -2.64
Mahalanobis Distance: 7.76

Mahalanobis Distance p-value: 0.997

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

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

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

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

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

Structural Similar Compounds				
Name	21232-47-3	61421-82-7	4947-27-7	
Structure	CI N N N N N N N N N N N N N N N N N N N	N. T. NH	O NH	
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen	
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen	
Distance	0.548	0.557	0.575	

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.

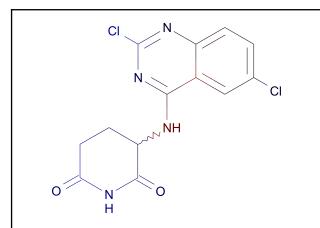
Chem. (2005) 48, 312-320

Feature Co	Feature Contribution					
	Top fe	atures for positive of	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set		
SCFP_12	112346096	CI NH CI NH [*][c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263		

SCFP_12	-1381862798	CI NH CI [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.314	339 out of 433
SCFP_12	10	CI NH CI NH [*]N[*]	0.306	1774 out of 2287
	Top Feat	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	[*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	CI NH CI NH [*]:n:[c](CI):n:[*]	-0.887	2 out of 11
SCFP_12	-116109291	CI NH CI NH CI NH [*]:[c]1:[*]:[cH]:[cH]:[cH]:1	-0.502	76 out of 221

0.539

EMIC



 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.616 Enrichment: 1.1

Bayesian Score: -4.58
Mahalanobis Distance: 8.67

Mahalanobis Distance p-value: 0.923

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

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

Enrichment: An 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	491-80-5	97919-22-7	2-ETHOXYMETHYL-1;3- DIHYDROXYANTHRAQUI NONE	
Structure	HO W OH	ON NH 2	O OH 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.

0.537

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

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

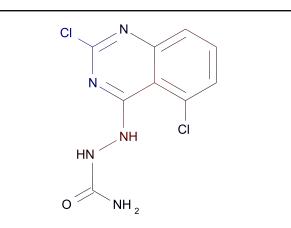
Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

0.534

Feature Contribution							
	Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set			
SCFP_12	-1231050790	CI N CI CI N CI N CI N CI N CI N CI N C	0.388	3 out of 3			

SCFP_12	112346096	CI NH ON NH (*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
SCFP_12	-1817806383	[,]cc(N[,])c(=[,])[,]	0.337	2 out of 2
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	CI NH OH [*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-1.82	0 out of 9
SCFP_12	395945879	CI NH	-1.19	0 out of 4
SCFP_12	-1286592310	CI NH CI NH (CI NH) (C	-0.887	2 out of 11



C₀H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Mutagen

Probability: 0.731 Enrichment: 1.31

Bayesian Score: -0.416
Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.0482

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

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

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

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

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

Structural Similar Compounds				
Name	2-Benzyliden-1-(3- aminoindol)-2- carbohydrazide	6027-98-1	Picloram	
Structure	NH H ₂ N ^t N	HCI H ₂ O H ₂ O	CI N NH 2 OH CI	
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen	
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen	
Distance	0.554	0.559	0.561	
Reference	Mutagenesis 7(1):37-39; 1992	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Mut. Res. 204: 17-115; 1988	

Model Applicability

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

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

Feature Contribution					
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set	
SCFP_12	-1977951815	CI NH	0.362	7 out of 8	

SCFP_12	112346096	CI NH ₂ [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
SCFP_12	-1381862798	[*]:n:[c]1:[cH]: [cH]:[*]:[c]:1:[*]	0.314	339 out of 433
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	CI NH CI NH ₂ [*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	CI NH	-0.887	2 out of 11
SCFP_12	-1378360678	CI NH CI HNN NH CI NH ₂ [*]:[o]1:[*]:[cH]:[cH]:[cH]:[cH]:[o]:1CI	-0.474	101 out of 285

0.562

Kazius et. al., J. Med.

Chem. (2005) 48, 312-320

CI N NH CI HN O S O

 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.623
Enrichment: 1.12
Bayesian Score: -4.38
Mahalanobis Distance: 9.26

Mahalanobis Distance p-value: 0.715

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	N/A	110021-94-8
Structure	N. H.	NNH NH	NH NH
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	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.

0.538

1992

Mut. Res. 280:233-244;

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

Mut. Res. 280:233-244;

0.531

1992

Feature Co	ntribution			
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	112346096	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263

SCFP_12	2054891299	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](C):[cH]:[cH]:1	0.322	13 out of 16
SCFP_12	-1381862798	[*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[o]:1:[*]	0.314	339 out of 433
	Top Feat	tures for negative of	ontribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	[*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-1.82	0 out of 9
SCFP_12	182991870	[*]NS(=O)(=O)[c](:[*]):[*]	-1.15	8 out of 48
SCFP_12	-1247518081	CI N CI [']NS(=0)(=0)(c)1:[cH]:[cH]:[']:[cH]:[cH] ::1	-1.01	6 out of 32

Mutagen

Mutagen

Mutagenesis 7(1):37-39;

0.557

1992

C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Prediction: Non-Mutagen

Probability: 0.577
Enrichment: 1.03
Bayesian Score: -5.64
Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.0424

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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/A	N/A	3-(4'- Acetylaminobenzylidenam ino)-5H-1;2;3-triazin- [5;4b]indol-4-one
Structure	N NH	NH NH	N N N N N N N N N N N N N N N N N N N

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.

Mutagen

Mutagen

Mut. Res. 280:233-244;

0.538

1992

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

Mut. Res. 280:233-244;

Mutagen

Mutagen

0.526

1992

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

SCFP_12	-1381862798	[*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.314	339 out of 433
SCFP_12	403834996	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.314	138 out of 176
	Top Feat	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	[*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-1.82	0 out of 9
SCFP_12	182991870	[*]NS(=0)(=0)(c](:[*]	-1.15	8 out of 48
SCFP_12	-1247518081	CI	-1.01	6 out of 32

 $C_{15}H_{13}CIN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.61
Enrichment: 1.09
Bayesian Score: -4.74
Mahalanobis Distance: 9.19

Mahalanobis Distance p-value: 0.749

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

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

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

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

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

Structural Similar Compounds				
Name	2425-85-6	N/A	3-(4'- Acetylaminobenzylidenam ino)-5H-1;2;3-triazin- [5;4b]indol-4-one	
Structure	OH NTM	N.H.	N Z = 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	0.537	0.548	0.549	
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Mut. Res. 280:233-244; 1992	Mutagenesis 7(1):37-39; 1992	

Model Applicability

Unknown features are fingerprint features in 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	Mutagen in training set		
SCFP_12	-1038366601	[*]N[c]1:n:[*] ^Q [c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4		
		j.[or i].[o]. 1.2				

SCFP_12	2097816882	[*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[cH]:[c H]:[cH]:[c]:1:2	0.42	4 out of 4
SCFP_12	1651620003	[*][c](:[*]):[c] ^{ft} [cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690
	Top Fea	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	182991870	["]NS(=O)(=O)(c](:[") NH NH N	-1.15	8 out of 48
SCFP_12	-1247518081	(*)NS(=0)(=0)(=1:1cH : cH :[*]:(cH :1:1	-1.01	6 out of 32
SCFP_12	-2090779860	[*][c]1:[*]:[c](:[*]) :[c](CI):n:n:1	-0.762	0 out of 2

Chem. (2005) 48, 312-320

C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248
Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model	Prediction
-------	-------------------

Prediction: Non-Mutagen

Probability: 0.56 Enrichment: 1

Bayesian Score: -6.09 Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.0495

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

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

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

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

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

Structural Similar Compounds				
Name	31430-15-6	3-(4'- Acetylaminobenzylidena mino)-5H-1;2;3-triazin- [5;4b]indol-4-one	97919-22-7	
Structure		Z=Z	C T T T T T T T T T T T T T T T T T T T	
Actual Endpoint	Non-Mutagen	Mutagen	Non-Mutagen	
Predicted Endpoint	Non-Mutagen	Mutagen	Non-Mutagen	
Distance	0.550	0.560	0.561	
Reference	Kazius et. al., J. Med.	Mutagenesis 7(1):37-39;	Kazius et. al., J. Med.	

Model Applicability

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

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

Chem. (2005) 48, 312-320 | 1992

Feature Co	ntribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set		
SCFP_12	-1038366601	[*]N[c]1:n:[*] ^{f]} [c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4		

SCFP_12	2097816882	[*]N[c]1:n:[*]:f[c]([*]):[c]2:[cH]:[cH]:[c H]:[cH]:[c]:1:2	0.42	4 out of 4
SCFP_12	1651620003	[*][c](:[*]):[c]f ² [cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	182991870	[*]NS(=O)(=O)[c](:[*]):[*]	-1.15	8 out of 48
SCFP_12	-1247518081	[*]NS(=0)(=0)[c]1:[cH]:[cH]:[*]:[cH]	-1.01	6 out of 32
SCFP_12	-2090779860	[*][c]1:[*]:[c](:[*]) ::[c](Cl):n:n:1	-0.762	0 out of 2

C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.587
Enrichment: 1.05
Bayesian Score: -5.39
Mahalanobis Distance: 9.04

Mahalanobis Distance p-value: 0.812

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	97919-22-7	3778-76-5	31431-39-7
Structure	C 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	HCI HCI

Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Distance	0.534	0.536	0.562
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

Churchinal Cimilar Campainale

Unknown features are fingerprint features in 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	2097816882	[*]N[c]1:n:[*]·[c]([*]):[c]2:[cH]:[cH]:[c H]:[cH]:[c]:1:2	0.42	4 out of 4		

SCFP_12	-1038366601	[*]N[c]1:n:[*f]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4
SCFP_12	1651620003	[*][c](:[*]):[cf]:[cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	182991870	[*]NS(=O)(=O)[c](:[*]	-1.15	8 out of 48
SCFP_12	-1247518081	[*]NS(=O)(=O)[c]1:[cH]]:[cH]:[1*][cH]:[1*]	-1.01	6 out of 32
SCFP_12	-1703874785	[*][c]1:[*]:n:[c](CI) :[c]2:[cH]:[cH]:[*]: [cH]:[c]:1:2	-0.762	0 out of 2

$N \sim$

 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.188 Enrichment: 0.338 Bayesian Score: -14.5 Mahalanobis Distance: 7.7

Mahalanobis Distance p-value: 0.998

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

Probability: The esimated probability that the 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	50-35-1	25057-89-0	81-63-0
Structure	N H	O N H O N H	H ₂ N O
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Distance	0.000	0.509	0.539
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

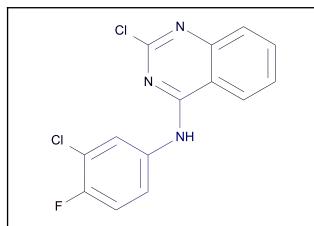
Model Applicability

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

All properties and OPS components are within expected ranges.

Feature Contribution							
Top features for positive contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set			
SCFP_12	280592283	[*]C([*])N1C(=0)[c]2: [cH]:[cH]:[cH]: [c]:2C1=0	0.197	3 out of 4			

SCFP_12	-1379591900	[*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0.108	1480 out of 2326	
SCFP_12	-496409612	[*]:[cH]:[cH]:[*	0.0771	2616 out of 4239	
Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set	
SCFP_12	1947881227		-1.9	0 out of 10	
]C1=[]			
SCFP_12	395945879	["]C1["]CC(=0)NC1=0	-1.19	0 out of 4	
SCFP_12	-1946889102	[,]cc(\(([,])[_,])c(=[,	-0.69	25 out of 89	



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.425
Enrichment: 0.809

Bayesian Score: -3.55

Mahalanobis Distance: 8.91

Mahalanobis Distance p-value: 0.204

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

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

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

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

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

Structural Similar Compounds					
Name	Triclabendazole	Meclofenamate Sodium (Free acid form)	Mequitazine		
Structure	CI NO THE SECOND	OH HN CI			
Actual Endpoint	Toxic	Non-Toxic	Toxic		
Predicted Endpoint	Toxic	Non-Toxic	Toxic		
Distance	0.587	0.595	0.621		
Reference	Toxicology 43(3):283-287;	Fundam Appl Toxicol	Oyo Yakuri 21:881-892;		

Model Applicability

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

5:665-671: 1985

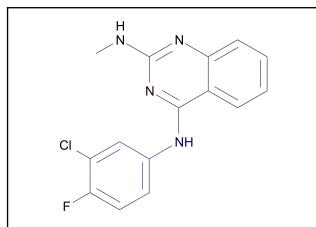
1981

Feature Co	ntribution				
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	129215346	[*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2	

SCFP_6	-1794884847	CI N	0.202	6 out of 9
		CI NH		
SCFP_6	-730654023	[*]:[c](:[*])F	0.202	6 out of 9
		[*][c](:[*]):[c](F):[
	Ton Foot	cH]:[*]		
Fingerprint	Bit/Smiles	tures for negative of Feature Structure	Score	Toxic in training
				set
SCFP_6	-300914917	[*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	[*]N[c]1:[cH]:[*]:[c] ([*]):[c](CI):[cH]:1	-0.594	1 out of 5
SCFP_6	-1380909229	CI NH CI NH [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.449	6 out of 19

Toxicol Appl Pharmacol

56(3):376-82: 1980



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Toxic Probability: 0.443

Enrichment: 0.842 Bayesian Score: -2.99

Mahalanobis Distance: 9.37 Mahalanobis Distance p-value: 0.0899

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	Meclofenamate Sodium (Free acid form)	Flufenamic Acid	Carprofen		
Structure	OH HN TO CI	F NNH OH	CI TO OH OH		
Actual Endpoint	Non-Toxic	Non-Toxic	Non-Toxic		
Predicted Endpoint	Non-Toxic	Non-Toxic	Non-Toxic		
Distance	0.505	0.557	0.589		

Model Applicability

Reference

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

Kiso to Rinsho 13:3302-

3313; 1979

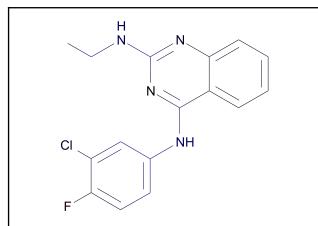
All properties and OPS components are within expected ranges.

Fundam Appl Toxicol

5:665-671: 1985

Feature Contribution				
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	136418580	CI NH	0.381	2 out of 2

SCFP_6	-1551584806	[*]N[c]1:n:[*]:[c](:[*]):[c](:[cH]:[*]):n :1	0.381	2 out of 2
SCFP_6	-1545804258	[*]N[c]1:n:[*]:[c](:[*]):[c](N[*]):n:1	0.381	2 out of 2
	-	ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	CI NH NH [*]:[c](:[*])NC	-0.945	0 out of 3
SCFP_6	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	[*]N[c]1:[cH]:[*]:[c] ([*]):[c](CI):[cH]:1	-0.594	1 out of 5



C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.395

Enrichment: 0.751
Bayesian Score: -4.59

Mahalanobis Distance: 10.1

Mahalanobis Distance p-value: 0.0147

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

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

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

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

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

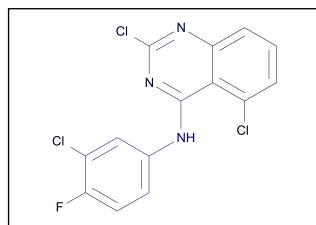
Name	Meclofenamate Sodium (Free acid form)		
Structure	OH HN CCI	F NH O	N N N N N N N N N N N N N N N N N N N
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Toxic
Distance	0.534	0.575	0.602
Reference	Fundam Appl Toxicol 5:665-671: 1985	Kiso to Rinsho 13:3302- 3313: 1979	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 Contribution				
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	129215346	[*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2

SCFP_6	-1551584806	[*]N[c]1:n:[*]:[c](:[*]):[c](:[cH]:[*]):n :1	0.381	2 out of 2
SCFP_6	-1545804258	[*]N[c]1:n:[*]:[c](:[*]):[c](N[*]):n:1	0.381	2 out of 2
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	CI NH [*]:[c](:[*])NC	-0.945	0 out of 3
SCFP_6	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	[*]):[c](CI):[cH]:1	-0.594	1 out of 5



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.411
Enrichment: 0.781
Bayesian Score: -4.05
Mahalanobis Distance: 9.09

Mahalanobis Distance p-value: 0.152

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Benzbromarone	Meclofenamate Sodium (Free acid form)
Structure	CI CI CI N N N N N N N N N N N N N N N N	HO WATER OF THE STATE OF THE ST	OH HN CI
Actual Endpoint	Toxic	Toxic	Non-Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.555	0.624	0.626
Reference	Toxicology 43(3):283-287; 1987	Shinryo to Shinaku 16:1521-1545; 1979	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.

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1794884847	CI NH CI NH CI [*]:[c](:[*])F	0.202	6 out of 9

SCFP_6	-730654023	[*][c](:[*]):[c](F):[cH]:[*]	0.202	6 out of 9
SCFP_6	-1378360678	[*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:[cH]:[c]:1CI	0.137	17 out of 28
	Top Feat	tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	CI NH CI [*]N[c]1:[cH]:[*]:[c] ([*]):[c](CI):[cH]:1	-0.594	1 out of 5
SCFP_6	-1380909229	[*]N[c]1:[cH]:[c] ([*]):[cH]:[cH]:1	-0.449	6 out of 19

Non-Toxic

CI N NH CI NH O H

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.452

Enrichment: 0.859 Bayesian Score: -2.71

Mahalanobis Distance: 10.3

Mahalanobis Distance p-value: 0.0103

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	D&C Yellow 8	Sulfonylurea Gliclazide	Cyclopiazonic Acid
Structure	OH OH	NH ON NH	N O HO

Predicted Endpoint Non-Toxic Toxic Non-Toxic Distance 0.528 0.562 0.611 Reference Food Chem Toxicol 24:819-823; 1986 Yakuri to Chiryo 9:3551-3571; 1981 J Toxicol Environ Health 14:585-594; 1984

Toxic

Model Applicability

Actual Endpoint

Structural Similar Compounds

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

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

Non-Toxic

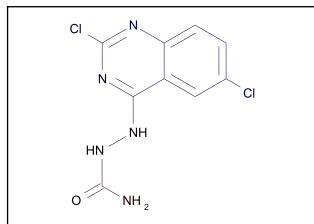
Feature Co	Feature Contribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	1631785938	[']C(=['])NC(=['])[']	0.255	3 out of 4		

SCFP_6	-1378360678	CI NH	0.137	17 out of 28
SCFP_6	622342378	CI NHO [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.114	6 out of 10
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	CI N N CI N N N N CI N N N N N N N N N N	-0.945	0 out of 3
SCFP_6	60071962	CI NH CI OH O [*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-0.422	0 out of 1
SCFP_6	-1286592310	CI NH	-0.422	0 out of 1

0.666

3571; 1981

Yakuri to Chiryo 9:3551-



C₀H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Toxic
Probability: 0.465

Enrichment: 0.885
Bayesian Score: -2.3

Mahalanobis Distance: 8.39

Mahalanobis Distance p-value: 0.42

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Guanabenz	Caffeic Acid	Sulfonylurea Gliclazide
Structure	CI NH H ₂ N NH		DH O=S
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic

Model Applicability

Distance

Reference

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

0.624

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

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

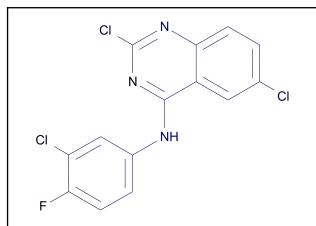
Journal of Toxic Sciences

11:107-119; 1982

0.609

Feature Co	ntribution			
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1357949052	CI NH	0.453	8 out of 9

SCFP_6	129215346	CI N H O NH2 [*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2
SCFP_6	-1378360678	CI NH NH NH ₂ [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	0.137	17 out of 28
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	60071962	[*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-0.422	0 out of 1
SCFP_6	1808332909	CI NH2 CI NH2 [*][c](:[*]):[c]1:[cH]:[c](CI):[cH]:[*]:[c]:1:[*]	-0.422	0 out of 1
SCFP_6	-182622233	CI NH; [*][c]1:[*]:[c]([*]): [c]2:[cH]:[c](CI):[c H]:[cH]:[c]:2:n:1	-0.422	0 out of 1



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.387
Enrichment: 0.736
Bayesian Score: -4.88
Mahalanobis Distance: 8.98

Mahalanobis Distance p-value: 0.184

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Benzbromarone	Meclofenamate Sodium (Free acid form)
Structure	CI TO THE CITY OF	HO WATER OF THE PARTY OF THE PA	OH HNV
Actual Endpoint	Toxic	Toxic	Non-Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.561	0.630	0.633
Reference	Toxicology 43(3):283-287; 1987	Shinryo to Shinaku 16:1521-1545: 1979	Fundam Appl Toxicol 5:665-671: 1985

Model Applicability

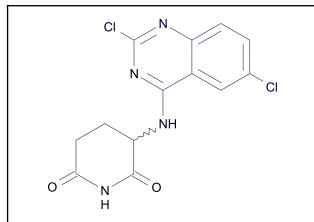
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	129215346	[*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2		

SCFP_6	-730654023	CI NH CI CI [*][c](:[*]):[c](F):[cH]:[*]	0.202	6 out of 9
SCFP_6	-1794884847	CI NH CI NH (*):[c](:[*])F	0.202	6 out of 9
	-	es for negative c		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-300914917	CI N CI CI N CI CI N CI N CI N CI N CI	-0.718	0 out of 2
SCFP_6	1807097289	CI NH CI CI (**)N(c)1:[c]+([*]):[c](CI):[cH]:1	-0.594	1 out of 5
SCFP_6	-1380909229	CI NH CI	-0.449	6 out of 19

0.618

Fundam Appl Toxicol 4:872-882; 1984



C₁₃H₁₀Cl₂N₄O₂

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.425

Enrichment: 0.809 Bayesian Score: -3.55

Mahalanobis Distance: 9.55

Mahalanobis Distance p-value: 0.0613

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.

Structural Similar Compounds					
Name	D&C Yellow 8	Sulfonylurea Gliclazide	Triclopyr		
Structure	OH OH	the NH ONH ONH	CI NO OH		
Actual Endpoint	Non-Toxic	Toxic	Non-Toxic		
Predicted Endpoint	Non-Toxic	Toxic	Non-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.

0.562

3571; 1981

Yakuri to Chiryo 9:3551-

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

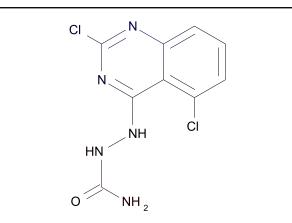
Food Chem Toxicol

24:819-823; 1986

0.533

Feature Co	Feature Contribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	129215346	CI NH CI NH CI NH CI NH CI NH CI NH CI NH CI NH CI NH CI NH CI NH CI NH NH NH NH NH NH NH NH NH NH NH NH NH	0.381	2 out of 2		

SCFP_6	1631785938	CI NH CI	0.255	3 out of 4
SCFP_6	-1378360678	CI NH ON H ON H [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	0.137	17 out of 28
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	CI NH CI NH (CI NH C) NH (CI NH	-0.945	0 out of 3
SCFP_6	1808332909	[*][c](CI):[cH];[*][c]:1:[*]	-0.422	0 out of 1
SCFP_6	60071962	CI NH	-0.422	0 out of 1



C₀H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Toxic
Probability: 0.494

Enrichment: 0.94
Bayesian Score: -1.46

Mahalanobis Distance: 9.07

Mahalanobis Distance p-value: 0.158

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	Guanabenz	Caffeic Acid	Bropirimine
Structure	NH H ₂ N NH	HO THE OH	Br NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Actual Endpoint	Toxic	Toxic	Non-Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.600	0.624	0.661
Reference	Journal of Toxic Sciences 11:107-119; 1982	Toxicol Appl Pharmacol 36(2):227-37; 1976	Teratology 38(1):7-14; 1988

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	-1357949052	CI N N CI N N CI N N N CI N N N N N CI N N N N	0.453	8 out of 9		

SCFP_6	-1378360678 1256786467	CI NH CI NH CI NH2 (*):[c]1:[*]:[cH]:[cH]:[c]:1CI	0.137	17 out of 28
SCI F_0	1230700407	CI NH CI NH CI NH ₂ [*]NC(=O)N	0.12	9 out of 13
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1286592310	CI N NH CI HN NH CI NH ₂ [*]:n:[c](CI):n:[*]	-0.422	0 out of 1
SCFP_6	60071962	CI NH CI NH ₂ [*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	-0.422	0 out of 1
SCFP_6	-1379591900	CI NH CI NH CI NH ₂ [*]:[c]1:[*]:[cH]:[cH]]:[cH]:[cH]:1	-0.282	33 out of 84

C₁₅H₁₂Cl₂N₄O₂S

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Non-Toxic Probability: 0.525 Enrichment: 0.998 Bayesian Score: -0.61 Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 6.09e-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

Enrichment: An 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.

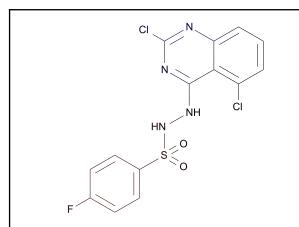
Structural Similar Compounds					
Name	Amsacrine	Ochratoxin a	Benomyl		
Structure	THE THE PART OF TH	OH MAN CI	NH H		
Actual Endpoint	Toxic	Toxic	Toxic		
Predicted Endpoint	Toxic	Toxic	Toxic		
Distance	0.515	0.642	0.651		
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	J Toxicol Environ Health 17:405-417; 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	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	2054891299	[*]S(=[*])(=[*))[c]1: [cH]:[cH]:[cH]:1	0.271	1 out of 1		

SCFP_6	1892882306	CI N CI N CI N CI [']NS(=0)(=0)(=)1:(cH]:(cH):(c)(-)(cH):[0.271	1 out of 1
SCFP_6	-1247518081	CHET CI N CI [*]NS(=0)(=0)(e)1:(eH [*](eH)(*)(eH)(eH) [*](eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH)(eH)(eH)(eH) [*](eH)(eH)(eH)(eH)(eH)(eH)(eH)(eH)(eH)(eH)	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	60071962	[*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	-0.422	0 out of 1
SCFP_6	-1286592310	CI NH CI HN NH CI S:0 S:0 [*]:n:[c](CI):n:[*]	-0.422	0 out of 1
SCFP_6	136686699	CI NH CI HN NH CI S:0 [*]:[c](:[*])C	-0.316	7 out of 19



C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.545
Enrichment: 1.04
Bayesian Score: -0.06
Mahalanobis Distance: 12.1

Mahalanobis Distance p-value: 2.32e-005

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

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

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

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

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

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

Name	Amsacrine	Ochratoxin a	Benomyl
Structure	N. THOMAS NO.	OH MANH OH MANH HOW	NH H
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.515	0.641	0.644
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	J Toxicol Environ Health 17:405-417; 1986

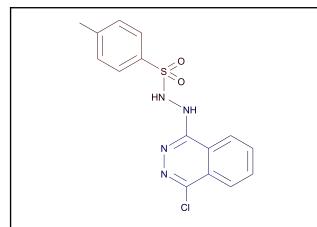
Model Applicability

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.

	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	-783770208	F[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.322	4 out of 5		

SCFP_6	-1247518081	CI N CI N CI (*)NS(=0)(=0)(=)!.;eH (*)[eH];(*)[eH];eH]	0.271	1 out of 1
SCFP_6	-1380395165	[*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*]) (=[*])[*]	0.255	3 out of 4
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	60071962	[*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	-0.422	0 out of 1
SCFP_6	-1286592310	CI NH CI HN NH CI F::0 S::0 [*]:n:[c](CI):n:[*]	-0.422	0 out of 1
SCFP_6	-1379591900	[*]:[cH]:[cH]:1	-0.282	33 out of 84



 $C_{15}H_{13}CIN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.456

Enrichment: 0.867 Bayesian Score: -2.59

Mahalanobis Distance: 15.9

Mahalanobis Distance p-value: 2.43e-013

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 C	Compounds
----------------------	-----------

Name	Amsacrine	Sulfonylurea Gliclazide	Benomyl
Structure	N N N N N N N N N N N N N N N N N N N	NH ON NH	NH H
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.547	0.565	0.594
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	Yakuri to Chiryo 9:3551- 3571; 1981	J Toxicol Environ Health 17:405-417; 1986

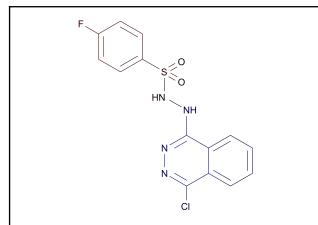
Model Applicability

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.

	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	1892882306	NN N N N N N N N N	0.271	1 out of 1		

SCFP_6	2054891299	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](C):[cH]:[cH]:1	0.271	1 out of 1
SCFP_6	795925860	[*][c]1:[cH]:[cH]:[c] (C):[cH]:[cH]:1	0.271	1 out of 1
		tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-601362767	[*]:n:[c](CI):[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	[*][c](:[*]):n:n:[*]	-0.448	5 out of 16
SCFP_6	2097816882	[*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[cH]:[c H]:[cH]:[c]:1:2	-0.422	0 out of 1



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.462
Enrichment: 0.879
Bayesian Score: -2.39
Mahalanobis Distance: 16

Mahalanobis Distance p-value: 1.37e-013

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 C	Compounds
----------------------	-----------

Name	Amsacrine	Benomyl	Sulfonylurea Gliclazide
Structure	N. THE NAME OF THE PARTY OF THE	NH H	NH ON NH ON NH
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.568	0.598	0.602
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	J Toxicol Environ Health 17:405-417; 1986	Yakuri to Chiryo 9:3551- 3571; 1981

Model Applicability

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

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

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	-783770208	F[c]1:[cH]:[cH]:1	0.322	4 out of 5	

SCFP_6	-1247518081	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
SCFP_6	-1380395165	[*][cH]:[*];[cH]:[cH] [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*]) (=[*])[*]	0.255	3 out of 4
	Top Feat	tures for negative	contribution)
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-601362767	[*]:n:[c](Cl):[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	[*][c](:[*]):n:n:[*]	-0.448	5 out of 16
SCFP_6	-705181820	[*]N[c]1:n:n:[c]([*]) :[*]:[c]:1:[*]	-0.422	0 out of 1

C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.426
Enrichment: 0.81
Bayesian Score: -3.52
Mahalanobis Distance: 15.7

Mahalanobis Distance p-value: 8.89e-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	Sulfonylurea Gliclazide	Benomyl	Amsacrine		
Structure	Atta. NH ONH ONH ONH	NH H	N. The state of th		
Actual Endpoint	Toxic	Toxic	Toxic		
Predicted Endpoint	Toxic	Toxic	Toxic		
Distance	0.584	0.586	0.592		

Model Applicability

Reference

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

J Toxicol Environ Health

17:405-417; 1986

Fundam Appl Toxicol

7(2):214-20; 1986

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

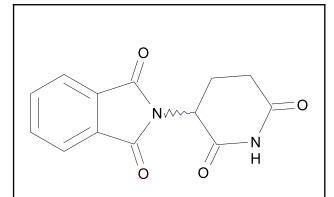
Yakuri to Chiryo 9:3551-

3571; 1981

Feature Co	Feature Contribution					
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set		
SCFP_6	1655803608	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]: [cH]:	0.271	1 out of 1		

SCFP_6	-1247518081	[']NS(=O)(=O)[c]1:[cH	0.271	1 out of 1
SCFP_6	-1380395165	[*][cH]:[:icH]:[cH] [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*]) (=[*][*]	0.255	3 out of 4
	Top Feat	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-601362767	[*]:n:[c](Cl):[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	[*][c](:[*]):n:n:[*]	-0.448	5 out of 16
SCFP_6	-705181820	[*]N[c]1:n:n:[c]([*]) :[*]:[c]:1:[*]	-0.422	0 out of 1

TOPKAT_Developmental_Toxicity_Potential



 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.536

Enrichment: 1.02

Bayesian Score: -0.317 Mahalanobis Distance: 7.73

Mahalanobis Distance p-value: 0.731

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

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

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

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

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

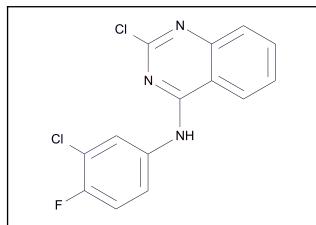
Structural Similar Compounds					
Name	11-oxo-11H-Pyrido(2;1- b)quinazoline-2- carboxylic Acid	Theobromine	Bropirimine		
Structure	N OH	HN N	Br NH NH 2		
Actual Endpoint	Toxic	Toxic	Non-Toxic		
Predicted Endpoint	Toxic	Toxic	Non-Toxic		
Distance	0.539	0.577	0.598		
Reference	Teratology 38(4):351-67; 1988	Food Chem Toxicol 24(5):375-82; 1986	Teratology 38(1):7-14; 1988		

Model Applicability

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

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set	
SCFP_6	1257084377	[*]N1[*][*]:[c](:[*]) C1=O	0.362	14 out of 18	

SCFP_6	1725056539	[*]N1C(=[*])[c]2:[cH] :[cH]:[cH]:[c]: 2C1=0	0.271	1 out of 1
SCFP_6	2049973914	[*]N1C(=[*])[c]2:[cH] :[*]:[cH]:[cH]:[c]:2 C1=0	0.271	1 out of 1
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	399659969	[*]C([*])N1C(=[*])[*] :[*]C1=[*]	-0.526	3 out of 11
SCFP_6	-1946889102	[,]CC(N([,])[,])C(=[,,])[,]	-0.512	4 out of 14
SCFP_6	-1379591900	[*]:[c]1:[*]:[cH]:[cH]:1	-0.282	33 out of 84



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227
Enrichment: 0.707
Bayesian Score: -2.35
Mahalanobis Distance: 9.9

Mahalanobis Distance p-value: 0.491

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Mestranol	Levonorgestrel
Structure		HO HO	OH NO OH

Carcinogen

Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.603

Carcinogen

Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.606

Model Applicability

Actual Endpoint

Distance

Reference

Predicted Endpoint

Structural Similar Compounds

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

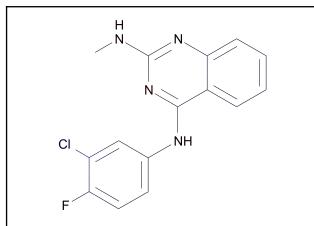
Res.) Sept. 1997

Carcinogen

Carcinogen

0.602

ECFP_6	-177077903	CI NH CI NH [*]N[c](:[cH]:[*]):[c H]:[*]	0.279	4 out of 9
ECFP_6	888054369	CI NH CI NH [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.234	3 out of 7
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1335691903	CI NH NH [*][c](:[*]):[c](CI): [cH]:[*]	-0.669	3 out of 22
ECFP_6	-1242906247	[*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2
ECFP_6	99947387	CI NH CI NH [*]:[c](:[*])CI	-0.461	9 out of 48



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.228
Enrichment: 0.712
Bayesian Score: -2.2
Mahalanobis Distance: 10.3

Mahalanobis Distance p-value: 0.302

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

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

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

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

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

Structural Similar Compounds

Name	Diclofenac	Phenolphthalein	Etodolac	
Structure	OH OH NNH	НООН	HO O H	
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen	
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen	
Distance	0.553	0.588	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.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1240852525: [*]:[c](:[*])NC

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_6	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.424	1 out of 1		

ECFP_6	-177077903	[*]N[c](:[cH]:[*]):[c H]:[*]	0.279	4 out of 9
ECFP_6	888054369	[*]N[c]1:[cH]:[c] ([*]):[cH]:[cH]:1	0.234	3 out of 7
	Top Fea	tures for negative of	ontribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-0.669	3 out of 22
ECFP_6	-1242906247	[*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2
ECFP_6	99947387	CI NH [*]:[c](:[*])CI	-0.461	9 out of 48

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

H N N	
CINH	
F III	

C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.221 Enrichment: 0.691 Bayesian Score: -2.84 Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.12

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

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

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

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

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

Structural Similar Compounds						
Name	ne Diclofenac Mef		Etodolac			
Structure	OH OH NNH	F F N OH	HO O H			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen			
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen			
Distance	0.552	0.611	0.624			
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997			

Model Applicability

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

1. OPS PC25 out of range. Value: -3.1285. Training min, max, SD, explained variance: -2.879, 2.6681, 1.105, 0.0124.

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_6	491100606	CI NH NH	0.424	1 out of 1		

ECFP_6	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.424	1 out of 1
ECFP_6	-177077903	[*]N[c](:[cH]:[*]):[c H]:[*]	0.279	4 out of 9
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-0.669	3 out of 22
ECFP_6	-950223878	CI NH NH [*]NCC	-0.482	0 out of 2
ECFP_6	-1242906247	[*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

0.643

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

CI N
CI NH CI

C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.216
Enrichment: 0.673
Bayesian Score: -3.49
Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.245

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

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

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

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

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

Structural Similar Compounds				
Name	Mestranol	Quazepam	Loratidine	
Structure	HO WO	F S N F T T T T T T T T T T T T T T T T T T	Cow N	
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-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.

0.638

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

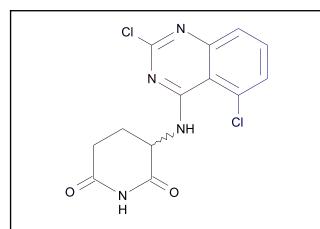
Eval.& Res./Off. Testing &

2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

Res.) Sept. 1997

0.631

ECFP_6	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.424	1 out of 1
ECFP_6	-177077903	[*]N[c](:[cH]:[*]):[c H]:[*]	0.279	4 out of 9
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1641317964	CI NH CI NH CI [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	-0.789	1 out of 11
ECFP_6	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-0.669	3 out of 22
ECFP_6	2007300961	CI NH CI [*][c]1:[*]:[c](:[*]) :[cH]:[cH]:[cH]:1	-0.652	5 out of 34



 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227 Enrichment: 0.709 Bayesian Score: -2.28 Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.0156

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Metolazone	Indapamide	Naltrexone
Structure	H ₂ N S the state of the state	HN H ₂ N O	HO the last of the
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.559	0.578	0.579
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_6	51876938	CI NH CI NH CI (*)CCC(=[*])[*]	0.473	16 out of 31		

ECFP_6	-1661653144	C NH 0 [*][c](:[*]):[c](:[c] ([*]):[*]):[c](:[*]) :[*]	0.442	2 out of 3
ECFP_6	1049768340	CI NH CI CI NH CI	0.424	1 out of 1
		ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1641317964	CI NH	-0.789	1 out of 11
ECFP_6	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-0.669	3 out of 22
ECFP_6	2007300961	CI NH CI O H [*][c]1:[*]:[c](:[*]) :[cH]:[cH]:[cH]:1	-0.652	5 out of 34

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

CI N CI NH NH HN NH 2

C₉H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.238
Enrichment: 0.744
Bayesian Score: -1.43
Mahalanobis Distance: 10.1

Mahalanobis Distance p-value: 0.392

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Pyrimethamine Guanfacine		
Structure	NH 2 N N	CI NH H ₂ N NH	CI 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	0.617	0.620	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: 1126732433: [*]:n:[c](Cl):n:[*]

Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set ECFP_6 432952415 CI NH ON N

ECFP_6	1626972527	CI NH CI NH NH CI NH2	0.424	1 out of 1
ECFP_6	-469184004	CI NH CI NH CI NH 2 [*]NNC(=O)N	0.424	1 out of 1
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	577592657	CI NH	-0.586	3 out of 20
ECFP_6	-176494269	CI NH CI NH CI NH2 [*]:[cH]:[c](CI):[cH] :[*]	-0.476	5 out of 28
ECFP_6	99947387	CI NH CI HN NH O NH 2 [*]:[c](:[*])CI	-0.461	9 out of 48

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

CI N CI
CINH

C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.213
Enrichment: 0.664
Bayesian Score: -3.86
Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.036

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Quazepam	Mestranol	Loratidine
Structure	F S N F 2 CI	HO THO	Z ₁ , N
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.627	0.631	0.634
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

ECFP_6	-177077903	CI NH CI [*]N[c](:[cH]:[*]):[c H]:[*]	0.279	4 out of 9
ECFP_6	888054369	CI NH CI [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.234	3 out of 7
	Top Fea	tures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-0.669	3 out of 22
ECFP_6	577592657	[*]:[c]1:[*]:[cH]:[cH]:1	-0.586	3 out of 20
ECFP_6	-1242906247	CI NH CI NH [*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.579

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.583

CI N CI NH O H

C₁₃H₁₀Cl₂N₄O₂

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.229
Enrichment: 0.714
Bayesian Score: -2.15
Mahalanobis Distance: 12.8

Mahalanobis Distance p-value: 0.00107

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Metolazone	Naltrexone	Indapamide	
Structure	H ₂ N S 1 NH	HO IN NOH	HN LZ SO	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	

Model Applicability

Predicted Endpoint

Distance

Reference

Structural Similar Compounds

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

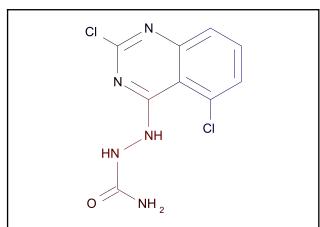
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

Non-Carcinogen

Res.) Sept. 1997

0.564

ECFP_6	1049768340	CI NH O [*]N[c](:n:[*]):[c](: [*]):[*]	0.424	1 out of 1
ECFP_6	-1699286547	[*]C(=[*])NC(=[*])[*]	0.297	12 out of 28
	Top Fea	tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	577592657	CI NH CI NH (CI NH CI NH	-0.586	3 out of 20
ECFP_6	-176494269	CI NH CI NH (*):[cH]:[c](CI):[cH]:[*]	-0.476	5 out of 28
ECFP_6	99947387	CI NH CI NH (CI	-0.461	9 out of 48



C₉H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.236
Enrichment: 0.738
Bayesian Score: -1.57
Mahalanobis Distance: 8.74

Mahalanobis Distance p-value: 0.918

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Guanfacine	Lamotrigine	Guanabenz
Structure	NH NH	H ₂ N ² N NH ₂	CI NH H ₂ N NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.586	0.597	0.608
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

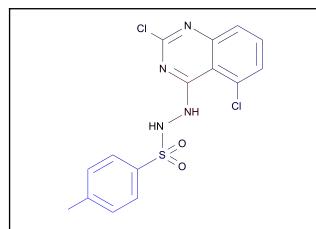
Model Applicability

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: 1126732433: [*]:n:[c](Cl):n:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1661653144	[*][c](:[*]):[c](:[*]) :[*]	0.442	2 out of 3

ECFP_6	432952415	CI N NH CI NH CI NH2 [*]NC(=O)N	0.442	2 out of 3
ECFP_6	-1238415266	CI NH CI HN NH CI NH ₂ [*]NN[c](:[*]):[*]	0.424	1 out of 1
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1641317964	CI NH2 CI NH2 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	-0.789	1 out of 11
ECFP_6	1335691903	CI NH CI NH½ [*][c](:[*]):[c](CI): [cH]:[*]	-0.669	3 out of 22
ECFP_6	2007300961	CI NH ₂ [*][c]1:[*]:[c](:[*]) :[cH]:[cH]:[cH]:1	-0.652	5 out of 34



 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.213
Enrichment: 0.664
Bayesian Score: -8.77
Mahalanobis Distance: 9.67

Mahalanobis Distance p-value: 0.599

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Torsemide
Structure	CIANOH NOOH	H ₂ N ₀ O	HN NH NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.634	0.646	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

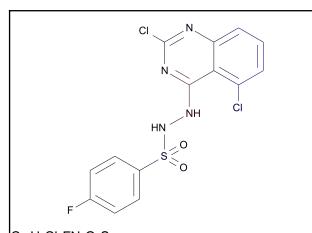
Model Applicability

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: 1126732433: [*]:n:[c](Cl):n:[*]
- 3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

eature Contribution				
Top features for positive contribution				
Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
-1661653144	[*][c](:[*]):[c](:[c] ([*]):[*]):[c](:[*]) :[*]	0.442	2 out of 3	
	Top fea Bit/Smiles	Top features for positive of Bit/Smiles Feature Structure -1661653144 [*][c](:[*]):[c](:[c] ([*]):[*]):[c](:[*])	Top features for positive contribution Bit/Smiles Feature Structure Score -1661653144 0.442	

ECFP_6	-1238415266	CI NH CI S; 0 S;	0.424	1 out of 1
ECFP_6	1049768340	[*]N[c](:n:[*]):[c](: [**]):[*]	0.424	1 out of 1
		tures for negative of	ontributio	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-179515162	[*]:[cH]:[c](C):[cH]: [*]	-1.41	0 out of 10
ECFP_6	-210573707	[*][c]1:[*]:[cH]:[c](C):[cH]:[cH]:1	-1.25	0 out of 8
ECFP_6	-1926229349	C N C N C N C N C (C):[cH]:[cH]:[c] (C):[cH]:[cH]:1	-1.05	0 out of 6



C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.22 Enrichment: 0.686 Bayesian Score: -2.98 Mahalanobis Distance: 11.5

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.

	Niclosamide	Indapamide
F O S HO ATT	CI _A	HN 72 SO
	nu o	O D O D O D O D O D O D O D O D O D O D

		0 - N 0 -	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.609	0.624	0.629
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

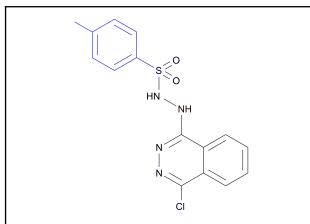
Structural Similar Compounds

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: 1126732433: [*]:n:[c](Cl):n:[*]
- 3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	-1661653144	[*][c](:[*]):[c](:[c] ([*]):[*]):[c](:[*]) :[*]	0.442	2 out of 3	

ECFP_6	-1238415266	CI NH CI S: 0 S: 0 S: 0 [*]NN[c](:[*]):[*]	0.424	1 out of 1
ECFP_6	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.424	1 out of 1
		tures for negative of	contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1641317964	[*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	-0.789	1 out of 11
ECFP_6	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-0.669	3 out of 22
ECFP_6	2007300961	[*][c]1:[*]:[c](:[*]) :[cH]:[cH]:[cH]:1	-0.652	5 out of 34



C₁₅H₁₃CIN₄O₂S

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.21 Enrichment: 0.654 Bayesian Score: -8.24 Mahalanobis Distance: 9.28

Mahalanobis Distance p-value: 0.766

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

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

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

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

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

Name	Tolazamide	Indapamide	Niclosamide
Structure	H P P P P P P P P P P P P P P P P P P P	HN H ₂ N S	Clary OH

Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.562	0.578	0.580
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

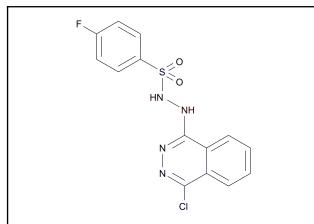
Structural Similar Compounds

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: 944235823: [*]NNS(=[*])(=[*])[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	-1238415266	[*]NN[c](:[*]):[*]	0.424	1 out of 1	

ECFP_6	1049768340		0.424	1 out of 1
		%. ZI		
		[*]N[c](:n:[*]):[c](: [*]):[*]		
	Top Featur	es for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-179515162	Q _e	-1.41	0 out of 10
		7.2 T.7.1 5.2 T.7.1		
		ći [*]:[cH]:[c](C):[cH]: [*]		
ECFP_6	-210573707	O.	-1.25	0 out of 8
		Z I Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z		
		ci [*][c]1:[*]:[cH]:[c](C):[cH]:[cH]:1		
ECFP_6	-1926229349	So z z z z z z z z z z z z z z z z z z z	-1.05	0 out of 6
		[*][c]1:[cH]:[cH]:[c] (C):[cH]:[cH]:1		



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.233
Enrichment: 0.725
Bayesian Score: -1.85
Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.0895

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Acetohexamide	Niclosamide
Structure	HN H ₂ N	O HN O	CIANOH

Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.567	0.583	0.583
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

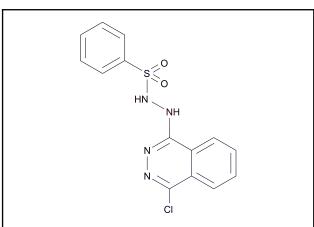
Structural Similar Compounds

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: 944235823: [*]NNS(=[*])(=[*])[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_6	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.424	1 out of 1	

ECFP_6	-1238415266	[*]NN[c](:[*]):[*]	0.424	1 out of 1
ECFP_6	-296909061	F[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.141	3 out of 8
		tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	99947387	[*]:[c](:[*])CI	-0.461	9 out of 48
ECFP_6	1050351974	[*]:n:[c](Cl):[c](:[*]):[*]	-0.27	0 out of 1
ECFP_6	74189795	[*]S(=[*])(=[*])(c]1: [cH]:[cH]:[c](F):[cH]:[cH]:1	-0.27	0 out of 1



C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227 Enrichment: 0.709 Bayesian Score: -2.28 Mahalanobis Distance: 9.73

Mahalanobis Distance p-value: 0.569

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Mebendazole	Acetohexamide	Indapamide
Structure	N H NH	HN ON THE STATE OF	HN H ₂ N O

Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.562	0.566	0.569
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: 944235823: [*]NNS(=[*])(=[*])[*]

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_6	-1238415266	[*]NN[c](:[*]):[*]	0.424	1 out of 1		

ECFP_6	1049768340	C LY Z	0.424	1 out of 1
		[*]N[c](:n:[*]):[c](: [*]):[*]		
	Top Fea	atures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	99947387	(*):[c](:[*])CI	-0.461	9 out of 48
ECFP_6	1050351974	[*]:n:[c](Cl):[c](:[*]):[*]	-0.27	0 out of 1
ECFP_6	2102150379	[*]S(=[*])(=O)[*]	-0.263	5 out of 22

Thalidomide

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

0 0

 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.244
Enrichment: 0.761
Bayesian Score: -1.06
Mahalanobis Distance: 11

Mahalanobis Distance p-value: 0.107

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

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

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

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

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

Structural Similar Compounds					
Name	Milrinone	Nalidixic acid	Theophylline		
Structure	N NH	HOO	N N H O		
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Distance	0.597	0.627	0.630		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997		

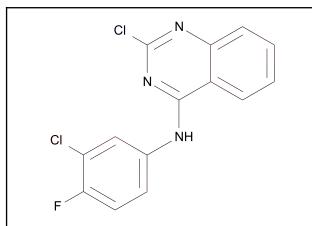
Model Applicability

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

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

Top featu Bit/Smiles	res for positive co	ontribution	
Rit/Smiles			
Divolliles	Feature Structure	Score	Carcinogen in training set
51876938	[*]CCC(=[*])[*]	0.473	16 out of 31
	1876938		

ECFP_6	-1699286547		0.297	12 out of 28
		N N N N N N N N N N N N N N N N N N N		
		[*]C(=[*])NC(=[*])[*]		
ECFP_6	2106656448		0.254	31 out of 77
		[*]C(=O)[*]		
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-484970154		-0.805	0 out of 4
		[*]CC(N([*))[*])C(=[*])[*]		
ECFP_6	-427397688		-0.476	5 out of 28
		[*]=C1[*][*][c](:[*]) :[c]1:[cH]:[*]		
ECFP_6	1984293270	N N N N N N N N N N N N N N N N N N N	-0.311	2 out of 10
		[*]=C1[*][*][c]2:[cH] :[cH]:[cH]:[cH]:[c]1 :2		



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.353
Enrichment: 1.2
Bayesian Score: 1.62
Mahalanobis Distance: 13

Mahalanobis Distance p-value: 0.000298

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

Probability: The esimated probability that the 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	Mestranol	Nafenopin	Levonorgestrel
Structure	HO MO	O O O O O O O O O O O O O O O O O O O	OH NAME OF THE PROPERTY OF THE
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.605	0.609	0.609
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
FCFP_6	-387072142	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[c] :1:[*]	0.477	4 out of 8		

FCFP_6	-2132756387	CI NH NH [*][c]1:[cH]:[cH]:[c] (F):[c](CI):[cH]:1	0.46	1 out of 1
FCFP_6	-773983804	CI NH CI NH [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.409	10 out of 24
		es for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1293778554	CI NH CI NH	-0.719	0 out of 4
FCFP_6	551850122	CI NH CI NH [*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	CI NH H P P P P P P P P P	-0.423	0 out of 2

C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.346
Enrichment: 1.17
Bayesian Score: 1.37
Mahalanobis Distance: 12.4

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	Diclofenac	Phenolphthalein	Etodolac
Structure	OH OH NH	НО	HO O H
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.530	0.590	0.614
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

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

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
FCFP_6	960546407	CI NH (*)[c]1:[*][c](:[*])	0.517	2 out of 3		
		:n:[c](NC):n:1				

FCFP_6	-387072142	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[c] :1:[*]	0.477	4 out of 8
FCFP_6	-2132756387	[*][c]1:[cH]:[c] (F):[c](CI):[cH]:1	0.46	1 out of 1
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	[*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	[*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]: [c]:1:2	-0.423	0 out of 2

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen

C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025
Rotatable Bonds: 4
Acceptors: 4

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.358
Enrichment: 1.22
Bayesian Score: 1.77
Mahalanobis Distance: 13.3

Mahalanobis Distance p-value: 0.000109

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Diclofenac Mefloquine In		Indomethacin	
Structure	OH O	F F N OH	HO	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Distance	0.536	0.592	0.621	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	

Model Applicability

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

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

Top fea	atures for positive c	ontribution	
t/Smiles			
y Onnie 3	Feature Structure	Score	Carcinogen in training set
0546407	[*][c]1:[*]:[c](:[*])	0.517	2 out of 3
	0546407	CI	[*][c]1:[*])

FCFP_6	-387072142	C NH N NH N NH	0.477	4 out of 8
FCFP_6	-2132756387	[*][c]1:[cH]:[cH]:[c] (F):[c](CI):[cH]:1	0.46	1 out of 1
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	[*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	[*]N[o]1:n:[*]:n:[c]2 ::[cH]:[*]:[cH]: [o]:1:2	-0.423	0 out of 2

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen

CI NH CI

C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.286
Enrichment: 0.972
Bayesian Score: -0.75
Mahalanobis Distance: 13

Mahalanobis Distance p-value: 0.0003

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

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

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

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

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

Structural Simila	ar Compounds		
Name	Mestranol	Loratidine	Nafenopin
Structure	HO WO	CI N N N O	OH O
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.633	0.638	0.646
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: -1374842118: [*]:n:[c](CI):n:[*]

FCFP_6	-773983804	[*]N[c]1:[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 of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	CI NH CI NH CI [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	CI NH CI NH CI [*]:[c](:[*])CI	-0.406	10 out of 59

CI N N CI O N H

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.275
Enrichment: 0.935
Bayesian Score: -1.19
Mahalanobis Distance: 12.8

Mahalanobis Distance p-value: 0.000579

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	Metolazone	Tolazamide	Indapamide		
Structure	H ₂ N S THE STATE OF THE STATE	T Z T Z T Z T Z T Z T Z T Z T Z T Z T Z	HN H ₂ N O		
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		
Distance	0.550	0.566	0.568		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &		

Model Applicability

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

Res.) Sept. 1997

Res.) Sept. 1997

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](CI):n:[*]

Res.) Sept. 1997

Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set FCFP_6 566058135 CI NH CI N

FCFP_6	1294255210	CI NH	0.441	12 out of 28
FCFP_6	-1151884458	[*]N[c](:n:[*]):[c](: [*]):[*]	0.348	6 out of 15
	Top Feat	tures for negative of	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	[*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	CI N N CI N N CI N N CI N N N CI N N N N	-0.406	10 out of 59
FCFP_6	367998008	CI NH CI NH CI NH CI CI CH]:[*][c](:[*]):[c](F):[c]	-0.374	10 out of 57

CI N CI NH NH HN O NH 2

C₀H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.34
Enrichment: 1.15
Bayesian Score: 1.17
Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.15

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Lamotrigine	Guanfacine	
Structure	NH 2 N N N N N N N N N N N N N N N N N N N	H ₂ N ^M NH ₂	NH H ₂ N NH	
Actual Endpoint	tual Endpoint Non-Carcinogen		Non-Carcinogen	
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Distance	0.595	0.607	0.611	
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	

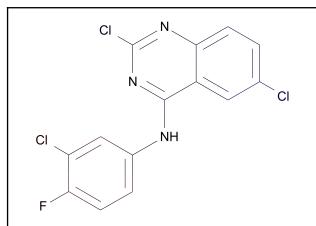
Model Applicability

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: -1374842118: [*]:n:[c](Cl):n:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885461129	CI NH	0.547	3 out of 5

FCFP_6	1294344583	CI NH	0.517	2 out of 3
FCFP_6	-1995759737	CI NH	0.46	1 out of 1
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	CI NH2 HNNH ONH2 [*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	CI NH CI CI CI NH CI	-0.423	0 out of 2
FCFP_6	71476542	CI NH CI NH ONH 2 [*]:[c](:[*])CI	-0.406	10 out of 59



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3 Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.273 Enrichment: 0.927 Bayesian Score: -1.29 Mahalanobis Distance: 12

Mahalanobis Distance p-value: 0.00584

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	Loratidine	Mestranol	Levonorgestrel	
Structure	CI N O	HO THO	OH NAME OH	
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen	
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen	
Distance	0.633	0.633	0.649	
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.
- Unknown FCFP 2 feature: -1374842118: [*]:n:[c](CI):n:[*]

Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set FCFP 6 -2132756387 0.46 out of 1 [*][c]1:[cH]:[cH]:[c] (F):[c](CI):[cH]:1

FCFP_6	-773983804	CI NH CI NH CI NH (I*]:[c] ([*]):[cH]:[cH]:1	0.409	10 out of 24
FCFP_6	590925877	CI NH [*]N[c](:[cH]:[*]):[c H]:[*]	0.369	13 out of 33
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1293778554	CI NH CI NH [*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	CI NH CI NH [*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	[*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]: [c]:1:2	-0.423	0 out of 2

CI N CI NH O H

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.262 Enrichment: 0.892 Bayesian Score: -1.73 Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.00356

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

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

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

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

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

Structural Sillinal Compounds						
Name	Metolazone	Tolazamide	Indapamide			
Structure	H ₂ N S 1 NH	N H Z H	No.			

Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.556	0.566	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

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: -1374842118: [*]:n:[c](Cl):n:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	566058135	CI NH CI NH (CI N) NH (CI	0.447	17 out of 40	

FCFP_6	1294255210	CI NH CI NH (CI NH C) NH (CI NH	0.441	12 out of 28
FCFP_6	-1151884458	[*]N[c](:n:[*]):[c](: [*]):[*]	0.348	6 out of 15
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	CI NH	-0.433	8 out of 49
FCFP_6	1122741451	CI NH	-0.423	0 out of 2
FCFP_6	71476542	CI NH CI NH (CI NH C) (CI NH (CI NH C) (CI NH	-0.406	10 out of 59

CI N CI NH CI NH 2

C₀H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.356
Enrichment: 1.21
Bayesian Score: 1.71
Mahalanobis Distance: 12.3

Mahalanobis Distance p-value: 0.00285

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Guanfacine	Pyrimethamine
Structure	CI	CI NH	H ₂ N ^M

Actual Endpoint Non-Carcinogen Non-Carcinogen Non-Carcinogen Predicted Endpoint Non-Carcinogen Non-Carcinogen Non-Carcinogen 0.599 0.602 0.604 Distance Reference US FDA (Centre for Drug US FDA (Centre for Drug US FDA (Centre for Drug Eval.& Res./Off. Testing & Eval.& Res./Off. Testing & Eval.& Res./Off. Testing & Res.) Sept. 1997 Res.) Sept. 1997 Res.) Sept. 1997

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: -1374842118: [*]:n:[c](CI):n:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	-885461129	CI N NH CI N	0.547	3 out of 5	

FCFP_6	1294344583	CI NH	0.517	2 out of 3
FCFP_6	1572236312	CI N NH CI HN NH CI (1):[e]:[:]:NNC(=O)N	0.46	1 out of 1
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	CI NH CI NHY CI CI CI ([*][c]1:[*]:[c]([*]):[c]([*]):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	CI NH CI NH CI NH CI (I CI (I))))))))))	-0.406	10 out of 59
FCFP_6	367998008	CI NHO CI	-0.374	10 out of 57

CI N NH CI HN O S O

 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227
Enrichment: 0.773
Bayesian Score: -3.4
Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 0.000761

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:

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Indomethacin	Indapamide
Structure		HO	HN 2N O
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.604	0.611	0.624
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: -1374842118: [*]:n:[c](Cl):n:[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	-885461129	CI NH CI NH CI S :: 0 S :: 0 [*]NNC(=[*])[*]	0.547	3 out of 5	

FCFP_6	1294344583	CI NH CI NH CI S;0	0.517	2 out of 3
FCFP_6	-1151884458	[*]NN[c](:[*]):[*] c N N N C (*]):[c](: [*]):[f]	0.348	6 out of 15
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 10
FCFP_6	632767364	CI N N CI N N CI S(=[*])(=[*])[c]1: [cH]:[c](C):[cH]:[cH]:1	-1.04	0 out of 7
FCFP_6	2109043264	C N C C C C C C C C C C C C C	-0.947	0 out of 6

CI N NH CI HN O S O

C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.307 Enrichment: 1.04

Bayesian Score: 0.0577 Mahalanobis Distance: 9.9

Mahalanobis Distance p-value: 0.421

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Bicalutamide	Indapamide	
Structure	CI AND HIN OH	HN 4h	HN 2 SO	
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen	

Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.601

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.601

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.

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](CI):n:[*]

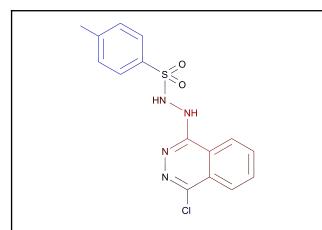
Non-Carcinogen

Res.) Sept. 1997

0.586

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set FCFP_6 -885461129 CI AND CI

FCFP_6	1294344583	CI NH CI HN NH CI S:0 S:0 S:0 [*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	-1151884458	[*]N[c](:n:[*]):[c](: [*]):[*]	0.348	6 out of 15
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	[*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	CI NH CI NH CI S::0 [*]:[c](:[*])CI	-0.406	10 out of 59
FCFP_6	367998008	[*][c](:[*]):[c](F):[cH]:[*]	-0.374	10 out of 57



 $C_{15}H_{13}CIN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.289
Enrichment: 0.983
Bayesian Score: -0.629
Mahalanobis Distance: 18.8

Mahalanobis Distance p-value: 4.69e-016

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Mebendazole	
Structure	CIANOH HN OH	HN H ₂ N O	N H O N H	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.567

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.575

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.

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

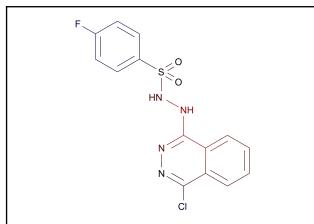
Non-Carcinogen

Res.) Sept. 1997

0.556

Feature Contribution					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
FCFP_6	-885461129	[*]NNC(=[*])[*]	0.547	3 out of 5	

FCFP_6	1294344583		0.517	2 out of 3
FCFP_6	-387072142	[*]NN[c](:[*]):[*] [*][c](:[*]):[c]f ² :[cH]:[cH]:[cH]:[c] :1:[*]	0.477	4 out of 8
	Top Fear	tures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 10
FCFP_6	632767364	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](C):[cH]:[cH]:1	-1.04	0 out of 7
FCFP_6	2109043264	[*][c]1:[cH]:[cH]:1	-0.947	0 out of 6



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.344
Enrichment: 1.17
Bayesian Score: 1.32
Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.189

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Indapamide	Mebendazole
Structure	CIANTO H	HN H ₂ N O	Z T T T T T T T T T T T T T T T T T T T
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.539	0.539	0.576
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

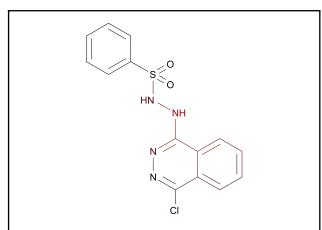
Res.) Sept. 1997

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	-885461129	[*]NNC(=[*])[*]	0.547	3 out of 5		

FCFP_6	1294344583		0.517	2 out of 3
FCFP_6	-387072142	[*]NN[c](:[*]):[*] [*][c](:[*]):[c]f:[cH]:[cH]:[cH]:[c] :1:[*]	0.477	4 out of 8
	Top Fea	tures for negative	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	[*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	-1374812327	[*]:n:[c](Cl):[c](:[*]):[*]	-0.423	0 out of 2
FCFP_6	71476542	[*]:[c](:[*])CI	-0.406	10 out of 59



 $C_{14}H_{11}CIN_4O_2S$

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.322 Enrichment: 1.09 Bayesian Score: 0.569 Mahalanobis Distance: 19.8

Mahalanobis Distance p-value: 2.3e-018

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Mebendazole	Mebendazole Indapamide	
Structure	Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z	HN H ₂ N O	CI MOH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.542	0.547	0.554
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	-885461129	(*]NNC(=[*])[*]	0.547	3 out of 5		

FCFP_6	1294344583	[*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	-387072142	[*][c](:[*]):[cH]:[cH]:[c] :1:[*]	0.477	4 out of 8
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1374812327	[*]:n:[c](CI):[c](:[*]):[*]	-0.423	0 out of 2
FCFP_6	71476542	[*]:[c](:[*])CI	-0.406	10 out of 59
FCFP_6	-1698724694	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]: [cH]:1	-0.22	15 out of 72

Thalidomide

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen

0 $N \wedge N \rightarrow 0$ $N \rightarrow 0$ $N \rightarrow 0$
0 0

 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.317
Enrichment: 1.08
Bayesian Score: 0.389
Mahalanobis Distance: 8.54

Mahalanobis Distance p-value: 0.928

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

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

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

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

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

Structural Similar Compounds					
Name	Milrinone	Phenobarbital	Dapsone		
Structure	NH NH	HN O	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	Carcinogen	Non-Carcinogen		
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen		
Distance	0.575	0.587	0.618		
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		

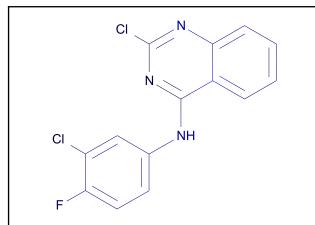
Model Applicability

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	2139882011	[*]N1C(=[*])[c]2:[cH] :[*]:[cH]:[cH]:[c]:2 C1=0	0.46	1 out of 1	

FCFP_6	1393189956	[*]N1C(=[*])[c]2:[cH] :[cH]:[cH]:[c]2:[cH]	0.46	1 out of 1
FCFP_6	566058135	[*]CC(=O)N[*]	0.447	17 out of 40
	Top Feat	tures for negative of	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1553874037	[*]C([*])N1C(=[*])[*] :[*]C1=[*]	-0.45	5 out of 32
FCFP_6	-1043310069	[,]CC(N([,])[,])C(=[,	-0.406	10 out of 59
FCFP_6	-228300541	[*]C1[*]NC(=0)CC1	-0.233	0 out of 1



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.147 Enrichment: 0.49 Bayesian Score: -10.8 Mahalanobis Distance: 13

Mahalanobis Distance p-value: 9.61e-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	Nafenopin	Mestranol	Levonorgestrel		
Structure	OH	HO WANTED TO THE REAL PROPERTY OF THE PROPERTY	OH NAME OF STREET		
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen		
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen		
Distance	0.622	0.623	0.625		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997		

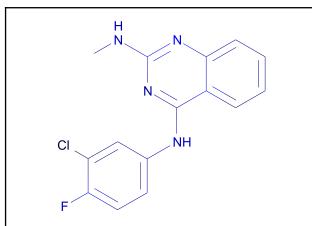
Model Applicability

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

- 1. OPS PC7 out of range. Value: -4.0266. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
- 2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](CI):n:[*]
- 3. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution					
Top Features for negative contribution Fingerprint Bit/Smiles Feature Structure Score Multiple-Carcinogen in					
FCFP_12	-1151884458	CI NH NH [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	training set 0 out of 6	

FCFP_12	590925877	CI NH NH [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	178336375	CI NH NH [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	-0.994	0 out of 5



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4 Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.143 Enrichment: 0.474 Bayesian Score: -14.3 Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.000707

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	Phenolphthalein	Oxazepam	Nafenopin	
Structure	НО	CI NOH	OH OH	
Actual Endpoint	Actual Endpoint Multiple-Carcinogen		Single-Carcinogen	
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen	
Distance	0.593	0.684	0.710	
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.
- Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
FCFP_12	1294255210	CI NH [*]:[c](:[*])NC	-1.63	0 out of 12	

FCFP_12	-1151884458	[*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6
FCFP_12	590925877	[*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

H N N	
CINH	

C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.138
Enrichment: 0.46
Bayesian Score: -14.8
Mahalanobis Distance: 15

Mahalanobis Distance p-value: 2.71e-006

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

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

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

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

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

Structural Similar Compounds					
Name	Phenolphthalein	Nafenopin	Diethylstilbesterol		
Structure	НО	OH OH	но		
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen		
Predicted Endpoint	Predicted Endpoint Multiple-Carcinogen		Single-Carcinogen		
Distance	0.657	0.695	0.702		
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](:[*]):[*]

Fingerprint	Top fea	atures for positive of Feature Structure	contribution Score	L
Fingerprint	Bit/Smiles	Feature Structure	Caara	L
			Score	Multiple- Carcinogen in training set
FCFP_12	136597326	H N N NH [*]CC	0.0722	18 out of 49

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	1294255210	I NH	-1.63	0 out of 12
		[*]:[c](:[*])NC		
FCFP_12	-1151884458	CI NH	-1.11	0 out of 6
		[*]N[c](:n:[*]):[c](: [*]):[*]		
FCFP_12	590925877	[*]N[c](:[cH]:[*]):[c	-0.998	1 out of 13
		[]N[c](.[cH].[]).[c H]:[*]		

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

N	CI N
CI NH CI	NH CI

C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.147
Enrichment: 0.488
Bayesian Score: -10.6
Mahalanobis Distance: 13.2

Mahalanobis Distance p-value: 6.75e-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	Mestranol	Nafenopin	Levonorgestrel
Structure	HO WAS TO THE STATE OF THE STAT	OH OH	O CH
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.647	0.654	0.658
Reference	erence 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 PC7 out of range. Value: -4.1247. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
- 2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]
- 3. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Top Features for negative contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1151884458	CI NH CI [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6

FCFP_12	590925877	CI NH CI [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	178336375	CI NH CI [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	-0.994	0 out of 5

CI N N CI N N H

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.149 Enrichment: 0.495 Bayesian Score: -13.6 Mahalanobis Distance: 11.2

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	Phenolphthalein	Sulfamethazine	Oxazepam
Structure	но	HN PP NH 2	CI THE

Single-Carcinogen

Single-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.657

Single-Carcinogen

Single-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.678

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.
- 2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

Multiple-Carcinogen

Multiple-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.652

Top Features for negative contribution

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

CI N CI
NH
HN
O NH ₂

C₉H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.151
Enrichment: 0.5
Bayesian Score: -7.23
Mahalanobis Distance: 11.7

Mahalanobis Distance p-value: 0.00105

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

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

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

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

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

Name	Phenobarbital	Danthron	Sulfamethazine	
Structure	HN O	ОМОН	HN AT N	
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen	
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen	
Distance	0.706	0.708	0.711	
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: -1374842118: [*]:n:[c](Cl):n:[*]

Feature Contribution Top features for positive contribution					
FCFP_12	-885461129	CI NH CI NH CI NH 2 [*]NNC(=[*])[*]	0.683	3 out of 3	

CI NH CI

C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.147
Enrichment: 0.488
Bayesian Score: -10.5
Mahalanobis Distance: 12

Mahalanobis Distance p-value: 0.000594

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Mestranol	Levonorgestrel	Nafenopin
Structure	HO HO		
	, no see	**************************************) OH

Single-Carcinogen

Res.) Sept. 1997

0.658

Multiple-Carcinogen

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Single-Carcinogen

Single-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.659

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 PC7 out of range. Value: -4.4725. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
- 2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]
- 3. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Multiple-Carcinogen

Multiple-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.647

Feature Contribution Top Features for negative contribution					
FCFP_12	-1151884458	CI NH CI [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6	

CI N CI NH O H

 $C_{13}H_{10}CI_{2}N_{4}O_{2}$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.149
Enrichment: 0.496
Bayesian Score: -13.6
Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.00447

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Sulfamethazine	Phenolphthalein	Oxazepam
Structure	HN MH 2	но	CI THE WOH
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.657	0.657	0.679
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

Feature Co	ntribution				
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
FCFP_12	-1043250487	[,]CC(N[,])C(=[,])[,] H O H CI NH CI	0.0691	7 out of 19	
	Top Features for negative contribution				

US FDA (Centre for Drug

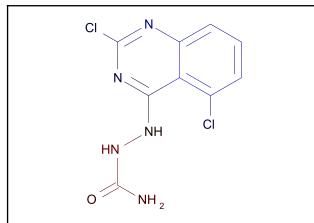
Res.) Sept. 1997

Eval.& Res./Off. Testing &

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



C₉H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.15 Enrichment: 0.5 Bayesian Score: -7.27 Mahalanobis Distance: 12.5

Mahalanobis Distance p-value: 0.000231

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

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

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

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

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

Name	Phenobarbital	Danthron	Sulfamethazine
Structure	H N O	O MOH	HN HN N
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.698	0.700	0.711
·	· · ·	 	Ť

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

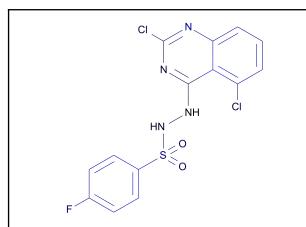
Eval.& Res./Off. Testing &

2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

Res.) Sept. 1997

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
FCFP_12	-885461129	CI N NH CI NH CI NH ₂ [*]NNC(=[*])[*]	0.683	3 out of 3	

FCFP_12	-1995759737	CI NH	0.4	1 out of 1
FCFP_12	1572236312	CI NH CI NH CI NH ₂ C'1:[c](:[*])NNC(=0)N	0.4	1 out of 1
		ures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1151884458	CI NH CI NH ₂ [*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6
FCFP_12	178336375	CI NHE ONHE [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	-0.994	0 out of 5
FCFP_12	367998008	CI NH CI NH _{NH} CI NH ₂ [*][c](:[*]):[c](F):[CH]:[*]	-0.789	1 out of 10



C₁₄H₀Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.15 Enrichment: 0.499 Bayesian Score: -13.5 Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 0.000165

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Bicalutamide	Phenolphthalein	Lansoprazole
Structure	HO HO AND	HO OH	SO NAME OF THE SECONDARY OF THE SECONDAR

Actual Endpoint Single-Carcinogen Multiple-Carcinogen Single-Carcinogen Predicted Endpoint Single-Carcinogen Multiple-Carcinogen Single-Carcinogen 0.629 0.680 0.721 Distance US FDA (Centre for Drug US FDA (Centre for Drug US FDA (Centre for Drug Reference Eval.& Res./Off. Testing & Eval.& Res./Off. Testing & Eval.& Res./Off. Testing & Res.) Sept. 1997 Res.) Sept. 1997 Res.) Sept. 1997

Model Applicability

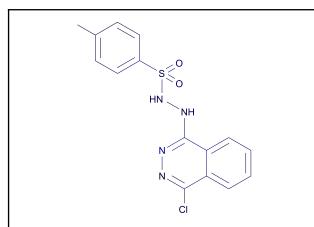
Structural Similar Compounds

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

Fingerprint Bit/Smiles Feature Structure Score Multiple-Carcinogen in training set FCFP_12 -885461129 CI AND CI A

FCFP_12	1294344583	CI NH CI	0.174	1 out of 2
		\$\frac{1}{5}\frac{1}{5		
		[*]NN[c](:[*]):[*]		
		es for negative c		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-1151884458	[*]N[c](:n:[*]):[c](: [*]):[*]	-1.11	0 out of 6
FCFP_12	178336375	[*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	-0.994	0 out of 5
FCFP_12	-1096219292	C FN C N C N H C ']NS(=0)(=0)(c](:[*]):[*]	-0.859	0 out of 4



C₁₅H₁₃CIN₄O₂S

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.149
Enrichment: 0.496
Bayesian Score: -11.3
Mahalanobis Distance: 17

Mahalanobis Distance p-value: 6.46e-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 substit.

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Phenolphthalein	Bicalutamide	Sulfamethazine
Structure	но	O DI S HO HN AM	HN THE N

Predicted Endpoint Multiple-Carcinogen Single-Carcinogen Single-Carcinogen 0.657 0.667 0.723 Distance Reference US FDA (Centre for Drug US FDA (Centre for Drug US FDA (Centre for Drug Eval.& Res./Off. Testing & Eval.& Res./Off. Testing & Eval.& Res./Off. Testing & Res.) Sept. 1997 Res.) Sept. 1997 Res.) Sept. 1997

Single-Carcinogen

Single-Carcinogen

Model Applicability

Actual Endpoint

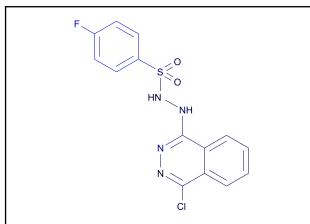
Structural Similar Compounds

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1374812327: [*]:n:[c](Cl):[c](:[*]):[*]

Multiple-Carcinogen

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
FCFP_12	-885461129	[*]NNC(=[*])[*]	0.683	3 out of 3		



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248
Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.145 Enrichment: 0.483 Bayesian Score: -14

Mahalanobis Distance: 14.9

Mahalanobis Distance p-value: 2.99e-006

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Bicalutamide	Phenolphthalein	Sulfamethazine		
Structure	Ho the state of th	но	O HN M		

	%	ОН	NH ₂
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.638	0.671	0.710
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

Structural Similar Compounds

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1374812327: [*]:n:[c](Cl):[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set	
FCFP_12	-885461129	[*]NNC(=[*])[*]	0.683	3 out of 3	

 $C_{14}H_{11}CIN_4O_2S$

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.148
Enrichment: 0.493
Bayesian Score: -11.1
Mahalanobis Distance: 17.6

Mahalanobis Distance p-value: 2.27e-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	Phenolphthalein	Sulfamethazine	Bicalutamide
Structure	НО	HN MM 2	HO JANA HO JANA F
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.667	0.673	0.681
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: -1374812327: [*]:n:[c](Cl):[c](:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
FCFP_12	-885461129	[*]NNC(=[*])[*]	0.683	3 out of 3

Thalidomide

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

O O

 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.219
Enrichment: 0.727
Bayesian Score: -1.62
Mahalanobis Distance: 8.2

Mahalanobis Distance p-value: 0.215

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

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

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

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

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

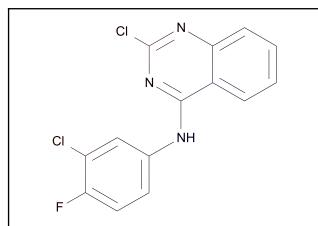
Structural Similar Compounds				
Name	Phenobarbital	Nitroacetophenetide	Metronidazole	
Structure	HN O	O No.	O N N N N N N N N N N N N N N N N N N N	
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen	
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen	
Distance	0.648	0.673	0.695	
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
FCFP_12	-1549163031	[*]N1[*][*]:[c](:[*]) C1=O	0.683	3 out of 3

			ı	l i
FCFP_12	-989213044	1/20 1/20	0.679	5 out of 6
FCFP_12	-1043310069	[,]cc(v([,])[,])c(=[,,])[,]	0.597	7 out of 10
	Top Foatur	os for nogativo c	ontribution	
Fin mannint		es for negative c		Multiple
		Feature Structure		Multiple- Carcinogen in training set
FCFP_12	991735244	[*]:[c]1:[*]:[cH]:[cH	-0.562	5 out of 28
]:[cH]:[cH]:1		
FCFP_12	566058135	[*]CC(=O)N[*]	-0.528	3 out of 17
FCFP_12	-1698724694		-0.423	3 out of 15
		[*]S(=[*])(=[*])[o]1: [cH]:[cH]:[cH]: [cH]:1		



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mild Probability: 0.813 Enrichment: 1.18

Bayesian Score: -0.466 Mahalanobis Distance: 6.91

Mahalanobis Distance p-value: 0.999

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

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

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

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

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

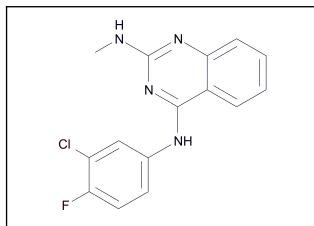
Structural Similar Compounds			
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-
Structure	The state of the s	CION	O N N N N O O
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.552	0.574	0.608
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72

Model Applicability

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

Feature Contribution				
Top fea	atures for positive o	ontribution		
Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
2047994594	CI N H E N H [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3	
	Top fea	Top features for positive of Bit/Smiles Feature Structure 2047994594 CI C	Top features for positive contribution Bit/Smiles Feature Structure Score 2047994594 CI	

FCFP_10	-1716224640	CI NH E*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	-2132756387	CI NH NH NH [*][c]1:[cH]:[cH]:[c] (F):[c](CI):[cH]:1	0.294	3 out of 3
	Top Feat	ures for negative of	ontribution	n _
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1320007763	CI NH H [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	-0.316	19 out of 40
FCFP_10	713358128	CI NH NH [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Mild
Probability: 0.768
Enrichment: 1.11
Bayesian Score: -2.04
Mahalanobis Distance: 6.81

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

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

Enrichment: An 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	PHENOL;2;2'- METHYLENEBIS(4- CHLORO-	P-PHENYLENEDIAMINE; N;N'-DIPHENYL-	P-PHENYLENEDIAMINE; N-PHENYL-N'- CYCLOHEXYL-	
Structure	Cland	NH NH	HN H	
Actual Endpoint	Moderate_Severe	Mild	Mild	
Predicted Endpoint	Moderate_Severe	Mild	Mild	
Distance	0.654	0.679	0.681	

Model Applicability

Reference

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

28ZPAK-;73;72

28ZPAK-;73;72

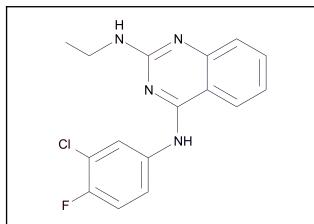
1. All properties and OPS components are within expected ranges.

28ZPAK-;82;72

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

Feature Contribution				
Top fea	ntures for positive o	ontribution		
Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
-2132756387	[*][c]1:[cH]:[cI]	0.294	3 out of 3	
	Top fea Bit/Smiles	Bit/Smiles Feature Structure -2132756387	Top features for positive contribution Bit/Smiles Feature Structure Score -2132756387	

FCFP_10	2047994594	CI	0.294	3 out of 3
FCFP_10	-1716224640	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1320007763	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	-0.316	19 out of 40
FCFP_10	713358128	[*]:n:[e]1:[cH]:[cH]: [cH]:[*]:[e]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	[*]:[c](:[*])N[c](:[*]	-0.304	9 out of 19



C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Mild Probability: 0.801 Enrichment: 1.16

Bayesian Score: -0.984 Mahalanobis Distance: 7.95

Mahalanobis Distance p-value: 0.934

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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;2;2'- METHYLENEBIS(4- CHLORO-	P-PHENYLENEDIAMINE; N-PHENYL-N'- CYCLOHEXYL-	P-PHENYLENEDIAMINE; N;N'-DIPHENYL-		
Structure	Cland	HN N	NH N		
Actual Endpoint	Moderate_Severe	Mild	Mild		
Predicted Endpoint	Moderate_Severe	Mild	Mild		
Distance	0.674	0.681	0.682		
Reference	28ZPAK-;82;72	28ZPAK-;73;72	28ZPAK-;73;72		

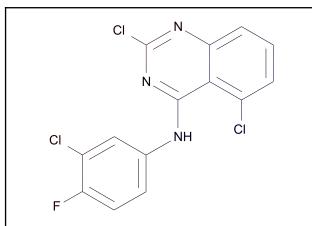
Model Applicability

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

Feature Co	Feature Contribution				
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	2047994594	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3	

FCFP_10	-2132756387	[*][c]1:[cH]:[c] (F):[c](CI):[cH]:1	0.294	3 out of 3
FCFP_10	-1716224640	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1320007763	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	-0.316	19 out of 40
FCFP_10	713358128	[*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mild Probability: 0.801 Enrichment: 1.16

Bayesian Score: -0.954 Mahalanobis Distance: 7.04

Mahalanobis Distance p-value: 0.998

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

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

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

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

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

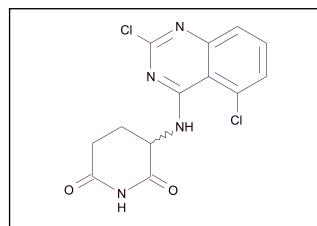
Structural Similar Compounds				
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	
Structure	The state of the s	CI	O NAME O O O O O O O O O O O O O O O O O O O	
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild	
Predicted Endpoint	Mild	Moderate_Severe	Mild	
Distance	0.475	0.610	0.645	
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72	

Model Applicability

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

Feature Co	Feature Contribution				
	Top fea	tures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-1716224640	CI N	0.294	3 out of 3	

FCFP_10	2047994594	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	-2132756387	CI NH CI NH CI (*][c]1:[cH]:[cH]:[c] (F):[c](CI):[cH]:1	0.294	3 out of 3
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	[*][c](:[*]):[c]1:[c] (CI):[cH]:[cH]:[*]:[c]:1:[*]	-0.842	0 out of 2
FCFP_10	713358128	CI NH CI [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19



 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Moderate_Severe

Probability: 0.825 Enrichment: 1.2

Bayesian Score: 0.195 Mahalanobis Distance: 8.96

Mahalanobis Distance p-value: 0.531

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 part of the sample.

Enrichment: An 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.

Caracterial Campounds						
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	1-AMINO-4-HYDROXY-5- CHLORANTHRAQUINONE			
Structure	HO the property of the second	HNm	HO MNH ₂			

Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.562	0.570	0.597
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;83;72

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: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-1716224640	CI NHO [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3	

FCFP_10	2047994594	CI N CI H	0.294	3 out of 3
FCFP_10	-922480536	CI N N CI N N CI N CI N CI N CI N CI N	0.256	2 out of 2
	Top Fea	tures for negative of	ontributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	CI N CI H	-0.842	0 out of 2
FCFP_10	713358128	CI NH CI NH CI NH CI (*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	307419094	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	-0.29	21 out of 43

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe

CI NH NH NH NH 2

C₉H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Mild Probability: 0.806 Enrichment: 1.17

Bayesian Score: -0.782 Mahalanobis Distance: 6.62

Mahalanobis Distance p-value: 1

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

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

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

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

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

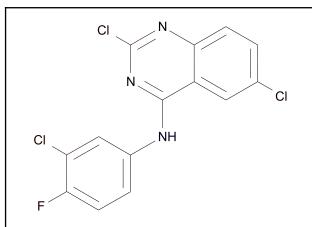
Structural Similar Compounds				
Name	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE	Benzimidazole; 6-amino-2- (3'-aminophenyl)-; dihydrochloride	
Structure	OH OH OH	O OH WOH	NH ₂	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance	0.583	0.594	0.613	
Reference	28ZPAK 190;72	28ZPAK-;103;7	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;831;86	

Model Applicability

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

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-745491832	CI NH NH ONH ₂ [*]:[c]1:[*]:[cH]:[cH]:[c](CI):[cH]:1	0.304	29 out of 32

FCFP_10	-1716224640	CI N CI N H O NH ₂ [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	1070061035	CI NH CI NH CI NH 2 [*]C(=[*])N	0.239	284 out of 338
		tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1294344583	CI NH CI CI NH 2 [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	1499521844	CI NH	-0.4	1 out of 3
FCFP_10	713358128	CI NH CI NH CI NH. (CI NH.) NH. (CI NH.) NH. (CI NH.) (*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Moderate_Severe

Probability: 0.823 Enrichment: 1.19

Bayesian Score: 0.0833 Mahalanobis Distance: 7.04

Mahalanobis Distance p-value: 0.998

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-		
Structure	The Hard And And And And And And And And And An	CI	O 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	Mild		
Predicted Endpoint	Mild	Moderate_Severe	Mild		
Distance	0.481	0.604	0.650		
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72		

Model Applicability

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

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-745491832	CI NH EF: (C) NH [*]:[c]1:[*]:[cH]:[cH]:[c](CI):[cH]:1	0.304	29 out of 32

FCFP_10	-2132756387	CI NH CI NH [*][c]1:[cH]:[c] (F):[c](CI):[cH]:1	0.294	3 out of 3
FCFP_10	-1716224640	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	713358128	CI NH	-0.307	8 out of 17
FCFP_10	1293778554	[*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19
FCFP_10	-773983804	CI NH CI NH [*]N[c]1:[cH]:[c] ([*]):[cH]:[cH]:1	-0.294	50 out of 102

CI N CI NH O H

C₁₃H₁₀Cl₂N₄O₂

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Moderate_Severe

Probability: 0.838 Enrichment: 1.22 Bayesian Score: 1.23

Mahalanobis Distance: 8.96

Mahalanobis Distance p-value: 0.531

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	1-AMINO-4-HYDROXY-5- CHLORANTHRAQUINONE
Structure	HO the MO	HN _M	HO mNH ₂

Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.567	0.576	0.602
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;83;72

Model Applicability

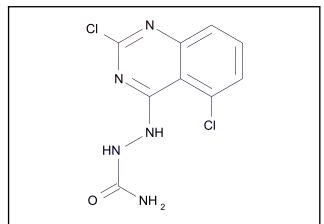
Structural Similar Compounds

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: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-745491832	CI NH	0.304	29 out of 32	

FCFP_10	-1716224640	[*][c]1:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	-922480536	CI NH CI NH CI (*)=C1[*]CCC(=O)N1	0.256	2 out of 2
		ntures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	713358128	CI NH OHHOO [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	307419094	CI NH NH (*)[[c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	-0.29	21 out of 43
FCFP_10	1294255210	CI NH CI NH (CI NH)	-0.218	20 out of 38



C₉H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Mild Probability: 0.776

Enrichment: 1.13
Bayesian Score: -1.82

Mahalanobis Distance: 6.62

Mahalanobis Distance p-value: 1

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

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

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

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

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

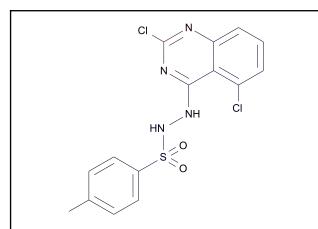
Structural Sim	ilar Compounds 2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7-	1;2;4-TRIHYDROXY ANTHRAQUINONE	Benzimidazole; 6-amino-2- (3'-aminophenyl)-; dihydrochloride
Structure	(METHÝLAMINO)-		NH ,
	OH OH	O OH OH	H ₂ N
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.583	0.588	0.607
Reference	28ZPAK 190;72	28ZPAK-;103;7	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;831;86

Model Applicability

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

Feature Contribution				
	Top fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1716224640	CI N CI	0.294	3 out of 3

FCFP_10	2047994594	CI NH ₂ [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	1070061035	CI N N CI N NH CI NH 2 [*]C(=[*])N	0.239	284 out of 338
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	CI N N CI N N N N N N N N N N N N N N N	-0.842	0 out of 2
FCFP_10	1294344583	CI NH CI HN NH CI NH ₂ [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	1499521844	CI NH CI NH CI NH 2 [*]NC(=O)N	-0.4	1 out of 3



 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild Probability: 0.772 Enrichment: 1.12 Bayesian Score: -1.92

Mahalanobis Distance: 5.56 Mahalanobis Distance p-value: 1

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

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

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

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

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

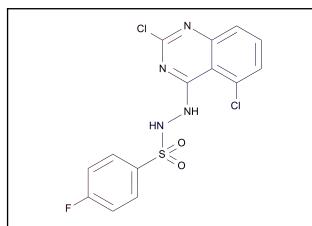
Structural Similar Compounds				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-	
Structure	HO stands of the NH 2	HN MAN THE NH 2	OHCI CI CI OH	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance	0.617	0.669	0.705	
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72	

Model Applicability

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

	Feature Contribution Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set		
FCFP_10	2047994594	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3		

FCFP_10	-1716224640	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	1789442672	CI N CI N CI N CI (7)NS(=0)(=0)[c]1:[cH]:[cH]:[cH]:1	0.256	2 out of 2
	Top Fea	tures for negative of	ontributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	[*][c](:[*]):[c]1:[c] (CI):[cH]:[cH]:[*]:[c]:1:[*]	-0.842	0 out of 2
FCFP_10	632767364	[*]S(=[*])(=[*])[c]1: [cH]:[c](C):[cH]:[cH]:1	-0.6	1 out of 4
FCFP_10	1294344583	CI NH CI NH CI S; 0 S; 0 [*]NN[c](:[*]):[*]	-0.507	0 out of 1



C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild Probability: 0.802 Enrichment: 1.16 Bayesian Score: -0.921

Mahalanobis Distance: 5.48

Mahalanobis Distance p-value: 1

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

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

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

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

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

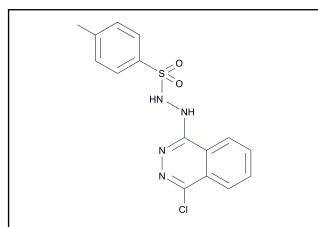
Structural Similar Compounds				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-	
Structure	HO MANU NA	HN MAN NH 2	OHCI CI CI OH	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance	0.601	0.650	0.682	
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72	

Model Applicability

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

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-149636017	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]:1	0.352	7 out of 7

FCFP_10	-1508180856	[*][cH]:[cH]:[c] (F):[cH]:1	0.329	16 out of 17
FCFP_10	-745491832	[*]:[c](CI):[cH]:1	0.304	29 out of 32
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	[*][c](:[*]):[c]1:[c] (CI):[cH]:[cH]:[*]:[c]:1:[*]	-0.842	0 out of 2
FCFP_10	1294344583	CI NH CI HN NH CI S:0 S:0 [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	713358128	[*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17



C₁₅H₁₃CIN₄O₂S

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild Probability: 0.73 Enrichment: 1.06 Bayesian Score: -2.95

Mahalanobis Distance: 5.53

Mahalanobis Distance p-value: 1

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

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

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

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

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

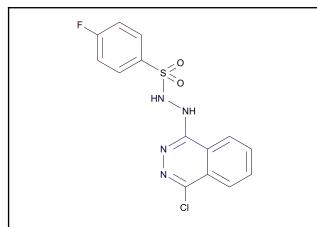
Structural Similar Compounds					
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-		
Structure	HO summer to the NH 2	HN Physical	OHCI CI CI OH		
Actual Endpoint	Mild	Mild	Moderate_Severe		
Predicted Endpoint	Mild	Mild	Moderate_Severe		
Distance	0.528	0.591	0.651		
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72		

Model Applicability

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

Feature Co	ntribution			
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1789442672	[*]NS(=0)(=0)[c]1:[cH]:[cH]:c[cH]:[cH]:1	0.256	2 out of 2

FCFP_10	136120670		0.206	53 out of 65
FCFP_10	-1374812327	[*]:[c](:[*])C	0.186	1 out of 1
	Top Feat	ures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	[*][c](:[*]):n:n:[*]	-1.29	0 out of 4
FCFP_10	632767364	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](C):[cH]:[cH]:1	-0.6	1 out of 4
FCFP_10	1294344583	[*]NN[c](:[*]):[*]	-0.507	0 out of 1



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248
Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.771 Enrichment: 1.12 Bayesian Score: -1.97

Mahalanobis Distance: 5.46

Mahalanobis Distance p-value: 1

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

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

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

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

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

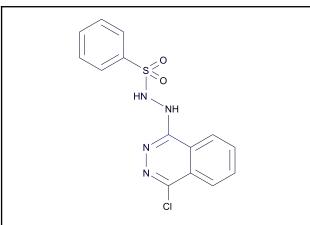
Structural Similar Compounds				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-	
Structure	HO MANUAL	HN MAN TANH 2	OHCI CI CI OH	
Actual Endpoint	Mild	Mild	Moderate_Severe	
Predicted Endpoint	Mild	Mild	Moderate_Severe	
Distance	0.534	0.583	0.638	
Reference	287PAK 239:72	287PAK-:124:72	287PAK-:92:72	

Model Applicability

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

Feature Contribution				
	Top fea	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-149636017	[*]S(=[*])(=[*])(c] [cH]:[cH]:[cH]:1	0.352	7 out of 7

FCFP_10	-1508180856	[*][c]1:[cH]:[cH]:[c] (F):[cH]:[cH]:1	0.329	16 out of 17
FCFP_10	-745491832	[*]:[c](CI):[cH]:1	0.304	29 out of 32
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	[*][c](:[*]):n:n:[*]	-1.29	0 out of 4
FCFP_10	1294344583	[*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	-1320007763	[*][c](:[*]):[c] ^f :[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	-0.316	19 out of 40



C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild Probability: 0.704 Enrichment: 1.02

Bayesian Score: -3.46
Mahalanobis Distance: 5.5

Mahalanobis Distance p-value: 1

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

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

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

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

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

Structural Similar Compounds

·				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	1-AMINO-4-HYDROXY-5- CHLORANTHRAQUINONE	
Structure	HO to NH 2	HN MH 2	HO the NH 2	
Actual Endpoint	Mild	Mild	Mild	
Predicted Endpoint	Mild	Mild	Mild	
Distance	0.525	0.570	0.635	
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: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

Top features for positive contribution					
-1374812327	[*]:n:[c](Cl):[c](:[*]):[*]	0.186	1 out of 1		
	Top fea Bit/Smiles	Top features for positive of Bit/Smiles Feature Structure -1374812327	Top features for positive contribution Bit/Smiles Feature Structure Score -1374812327 0.186		

FCFP_10	3		0.165	383 out of 491
FCFP_10	32	[*]N[*] oooluber	0.0821	145 out of 202
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	[*][c](:[*]):n:n:[*]	-1.29	0 out of 4
FCFP_10	1294344583	[*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	-1320007763	[*][c](:[*]):[cfl:[cH]:[cH]:[cH]:[*]	-0.316	19 out of 40

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Moderate
Probability: 0.581

Enrichment: 0.938
Bayesian Score: -2.63

Mahalanobis Distance: 10.3

Mahalanobis Distance p-value: 0.0189

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

negative rate

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

Enrichment: An 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.

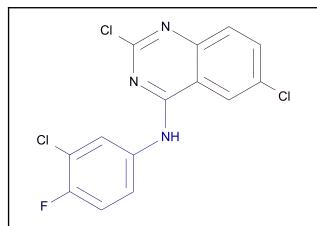
Structural Similar Compounds				
Name	5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-	s-TRIAZINE; 2-CHLORO- 4-ETHYLAMINO-6- ISOPROPYLAMINO-	2- NAPHTHALENESULFONI C ACID;5-AMINO-6- ETHOXY-	
Structure	OHCI CI CI OH	CI NI NH	HO S NH ₂	
Actual Endpoint	Severe	Severe	Moderate	
Predicted Endpoint	Severe	Moderate	Moderate	
Distance	0.624	0.644	0.660	
Reference	28ZPAK-;92;72	CIGET* -;-;77	28ZPAK-;191;72	

Model Applicability

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

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set	
SCFP_12	1631785938	[']C(=['])NC(=['])[']	0.218	1 out of 1	

SCFP_12	622342378	CI NHO [*][c]1:[*]:[c]([*]): [c]2:[c]([*]):[*]:[c H]:[cH]:[c]:2:n:1	0.213	4 out of 5
SCFP_12	403834996	CI NH CI NH O [*][c]1:[*]:[c]([*]): [c]2:[c]([*]):[cH]:[cH]:[cH]:[c]:2:n:1	0.213	4 out of 5
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1072897324	CI NH	-0.8	3 out of 13
SCFP_12	1256995004	CI NH CI NH CI	-0.483	12 out of 33
SCFP_12	18117904	[*]C([*])N[c](:[*]):[*]	-0.481	3 out of 9



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Prediction: Moderate
Probability: 0.479
Enrichment: 0.772
Bayesian Score: -4.46

Mahalanobis Distance: 5.84 Mahalanobis Distance p-value: 1

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

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

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

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

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

Structural Similar Compounds				
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	2';6'-Acetoxylidide; 2-(p-chlorobenzyl(2-(pyrrolidinyl)ethyl)amino)-;	
Structure	Man H	CI		
Actual Endpoint	Moderate	Severe	Severe	
Predicted Endpoint	Moderate	Severe	Severe	
Distance	0.484	0.615	0.704	
Reference	28ZPAK-;242;72	CIGET* -;-;77	Arzneimittel-Forschung 9;167;59	

Model Applicability

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

Feature Contribution					
	Top fe	atures for positive of	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set	
SCFP_12	622342378	CI NH	0.213	4 out of 5	

SCFP_12	-1378360678	CI NH CI NH CI CI CI CI CI CI CI C	0.211	22 out of 29
SCFP_12	-116109291	CI NH CI	0.174	19 out of 26
		tures for negative o		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	951581613	CI NH CI NH [*]:[c](:[*])N[c](:[*]):[*]	-0.769	2 out of 9
SCFP_12	1334669481	CI NH CI (:[cH]:[*]):[c H]:[*]	-0.685	28 out of 93
SCFP_12	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	-0.561	2 out of 7

CI NH O NH

C₁₃H₁₀Cl₂N₄O₂

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Moderate
Probability: 0.579
Enrichment: 0.934
Bayesian Score: -2.67

Mahalanobis Distance: 10.3

Mahalanobis Distance p-value: 0.0189

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

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

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

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

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

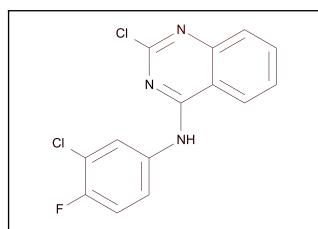
Structural Similar Compounds				
Name	5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-	s-TRIAZINE; 2-CHLORO- 4-ETHYLAMINO-6- ISOPROPYLAMINO-	2- NAPHTHALENESULFONI C ACID;5-AMINO-6- ETHOXY-	
Structure	OHCI CI CI OH	HN NH	HO S NH ₂	
Actual Endpoint	Severe	Severe	Moderate	
Predicted Endpoint	Severe	Moderate	Moderate	
Distance	0.624	0.645	0.660	
Reference	28ZPAK-;92;72	CIGET* -;-;77	28ZPAK-;191;72	

Model Applicability

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

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set	
SCFP_12	1631785938	CI NH CI	0.218	1 out of 1	

SCFP_12	622342378	CI NH NH CI [*][c]1:[*]:[c]([*]): [c]2:[c]([*]):[*]:[c H]:[cH]:[c]:2:n:1	0.213	4 out of 5
SCFP_12	-1378360678	CI NH CI NH (CI NH CI NH	0.211	22 out of 29
	Top Featu	res for negative o	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1072897324	CI NH CI	-0.8	3 out of 13
SCFP_12	1256995004	CI NH CI NH (*) CI (*)	-0.483	12 out of 33
SCFP_12	18117904	CI NH CI	-0.481	3 out of 9



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1 Enrichment: 1.18 Bayesian Score: 3.13

Mahalanobis Distance: 6.09

Mahalanobis Distance p-value: 1

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

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

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

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

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

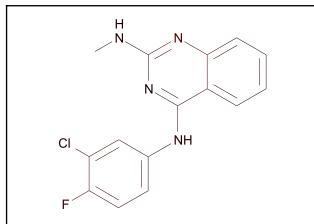
Structural Similar Compounds				
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	
Structure	TANK H	CI	O N N N O O O O O O O O O O O O O O O O	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.545	0.563	0.599	
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72	

Model Applicability

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

Feature Co	Feature Contribution				
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	1747237384	CI NH [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44	

FCFP_12	178336375	[*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	CI NH NH Sin:[c]1:[cH]:[cH]: [cH]:[c]:1:[t]	0.2	17 out of 17
	Top Fea	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	307419094	C NH *][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52
FCFP_12	-773983804	CI NH CI NH [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121
FCFP_12	991735244	CI NH NH Sign (C) 1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1 Enrichment: 1.18 Bayesian Score: 3.49

Mahalanobis Distance: 4.96

Mahalanobis Distance p-value: 1

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

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

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

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

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

Name	PHENOL;2;2'- METHYLENEBIS(4- CHLORO-	P-PHENYLENEDIAMINE; N;N'-DIPHENYL-	s-TRIAZINE; 2-CHLORO- 4;6- BIS(ISOPROPYLAMINO)-
Structure	Clanton	NH NH	CI N N N N N N N N N N N N N N N N N N N

	ОН		
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.647	0.674	0.675
Reference	28ZPAK-;82;72	28ZPAK-;73;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.

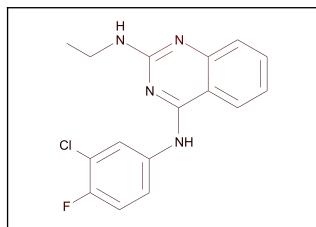
- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

FCFP_12	178336375	CI NH	0.202	19 out of 19
FCFP_12	713358128	[*]:[cH]:[c](:n:[*]):	0.2	17 out of 17
	Top Featu	res for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	307419094	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52
FCFP_12	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	-773983804	[*]N[c]1:[cH]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121

Irritant

0.677

28ZPAK-;73;72



C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1

Enrichment: 1.18

Bayesian Score: 3.53 Mahalanobis Distance: 5.94

Mahalanobis Distance p-value: 1

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

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

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

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

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

Structural Similar Compounds				
Name	PHENOL;2;2'- METHYLENEBIS(4- CHLORO-	P-PHENYLENEDIAMINE; N-PHENYL-N'- CYCLOHEXYL-	P-PHENYLENEDIAMINE; N;N'-DIPHENYL-	
Structure	CluntoH	H N N	NH N	
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

0.677

28ZPAK-;73;72

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

Irritant

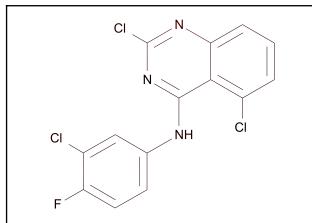
0.665

28ZPAK-;82;72

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

Feature Contribution Top features for positive contribution					
1747237384	[*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44		
	Top fea	Bit/Smiles Feature Structure 1747237384 Top features for positive of the p	Top features for positive contribution Bit/Smiles Feature Structure Score 1747237384		

FCFP_12	178336375	[*]:[cH]:[c](:n:[*]):	0.202	19 out of 19
FCFP_12	713358128	[c](:[*]):[*] H N CI NH [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17
	Top Fea	tures for negative	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	991735244	[*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	-773983804	[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121
FCFP_12	0	H N N N N N N N N N N N N N N N N N N N	0	1184 out of 1397



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Probability: 1

Model Prediction

Prediction: Irritant

Enrichment: 1.18

Bayesian Score: 2.64 Mahalanobis Distance: 6.25

Mahalanobis Distance p-value: 1

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

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

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

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

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

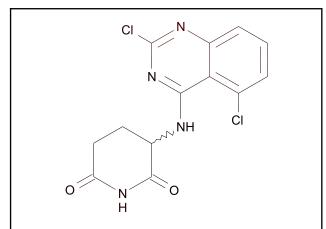
Structural Similar Compounds				
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-	
Structure	The Hard of the Ha	CI	O Name o	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.476	0.593	0.628	
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72	

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. 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	[*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44		

FCFP_12	178336375	[*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	CI NH CI [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	[*][c](:[*]):[c]1:[c] (CI):[cH]:[cH]:[*]:[c]:1:[*]	-0.132	2 out of 3
FCFP_12	991735244	CI NH CI NH CI (1):[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	-773983804	[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121



 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1

Enrichment: 1.18
Bayesian Score: 2.85

Mahalanobis Distance: 6.88

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

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

Enrichment: An 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	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	Anthraquinone; 1-amino- 2-bromo-4-hydroxy-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE
Structure	HO MANH 2	HO my Br NH 2	HN MN NH 2
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.552	0.562	0.567
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: -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	[*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44		

FCFP_12	178336375	[*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	[*]:n:[c]1:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	CI N CI H CI	-0.132	2 out of 3
FCFP_12	1	CI N N CI N N CI N N N CI N N N N CI N N N N	0	872 out of 1051
FCFP_12	991735244	CI NH CI NH CI (CI) (CI	0	237 out of 291

CI NH NH NH NH 2

C₀H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Irritant

Probability: 1 Enrichment: 1.18

Bayesian Score: 2.46

Mahalanobis Distance: 4.47

Mahalanobis Distance p-value: 1

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

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

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

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

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

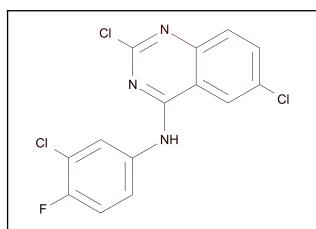
Structural Similar Compounds				
Name	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE	Benzimidazole; 6-amino-2- (3'-aminophenyl)-; dihydrochloride	
Structure	OH NH	O OH OH	NH ₂	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.575	0.588	0.605	
Reference	28ZPAK 190;72	28ZPAK-;103;7	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;831;86	

Model Applicability

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

Feature Contribution Top features for positive contribution					
FCFP_12	1747237384	CI NH CI NH CI NH2 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44	

FCFP_12	178336375	CI HN NH O NH ₂ [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	CI HN NH O NH; [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17
	Top Fea	atures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	0	CI NH	0	1184 out of 1397
FCFP_12	1	CI NH CI NH O NH 2 [*]=0	0	872 out of 1051
FCFP_12	307419094	CI NHH ONH2 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1

Enrichment: 1.18
Bayesian Score: 2.8

Mahalanobis Distance: 6.25 Mahalanobis Distance p-value: 1

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

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

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

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

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

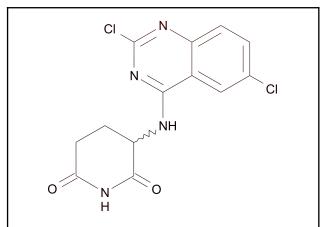
Structural Similar Compounds					
Name	ANTHRAQUINONE;1- (2;4;6- TRIMETHYLPHENYLAMIN O)-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5- CHLORO-1- ANTHRAQUINONYL)-		
Structure	The state of the s	CI	O NAME O O O O O O O O O O O O O O O O O O O		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Irritant	Irritant	Irritant		
Distance	0.481	0.587	0.633		
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72		

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. 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	CI NH CI NH [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44		

FCFP_12	178336375	CI NH CI [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	CI NH CI NH (CI NH CI NH	0.2	17 out of 17
	Top Fea	atures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	307419094	CI NH CI NH ([*][c](:[*]):[c](:[cH	0	43 out of 52
]:[*]):[c](:[*]):[*]		



 $C_{13}H_{10}CI_{2}N_{4}O_{2}$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1 Enrichment: 1.18 Bayesian Score: 3

Mahalanobis Distance: 6.88

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

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

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

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

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

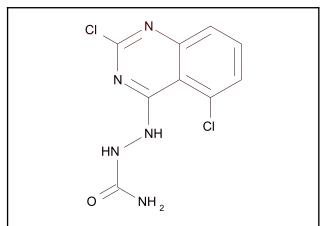
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	Anthraquinone; 1-amino- 2-bromo-4-hydroxy-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE
Structure	HO stands of the NH 2	HO 11 NH 2	HN MH 2
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.556	0.568	0.572
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: -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	CI NH CI NH (CI NH) (C	0.208	44 out of 44		

FCFP_12	178336375	CI NH ON H [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	CI NH CI NH CI NH (CI NH CI NH	0.2	17 out of 17
	Top Fea	tures for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1	CI NH	O	872 out of 1051
FCFP_12	307419094	CI NH OI NH (I'):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52
FCFP_12	-1272798659	CI NH CI NH (*)[*][*]	0	517 out of 643



C₉H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Irritant
Probability: 1

Enrichment: 1.18
Bayesian Score: 2.3

Mahalanobis Distance: 4.47 Mahalanobis Distance p-value: 1

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

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

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

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

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

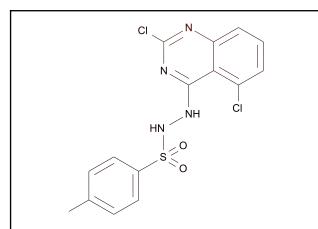
Structural Similar Compounds			
Name	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE	Benzimidazole; 6-amino-2- (3'-aminophenyl)-; dihydrochloride
Structure	OH OH OH	O OH OH OH OH	NH ₂
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.574	0.582	0.599
Reference	28ZPAK 190;72	28ZPAK-;103;7	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;831;86

Model Applicability

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

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	1747237384	[*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44		

FCFP_12	178336375	CI NH2 CI NH2 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	Ci NH ₂ [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17
	Top Featu	res for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	CI N CI CI N CI	-0.132	2 out of 3
FCFP_12	1	CI NH CI NH CI NH 2 [*]=O	0	872 out of 1051
FCFP_12	307419094	CI NH2 CI NH2 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52



 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1
Enrichment: 1.18
Payesian Secret 2

Bayesian Score: 2.72

Mahalanobis Distance: 3.95 Mahalanobis Distance p-value: 1

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

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

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

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

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

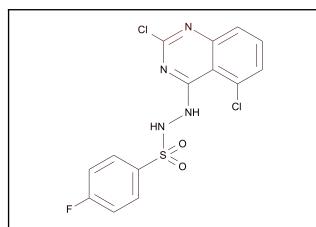
Structural Similar Compounds				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-	
Structure	HO MANUAL	HN str. NH 2	OHCI CI CI OH	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.605	0.651	0.681	
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72	

Model Applicability

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

ntribution			
Top fea	atures for positive o	ontribution	
Bit/Smiles	Feature Structure	Score	Irritant in training set
1747237384	[*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44
	Top fea Bit/Smiles	Top features for positive of Bit/Smiles Feature Structure 1747237384 CI C	Top features for positive contribution Bit/Smiles Feature Structure Score 1747237384 0.208

FCFP_12	178336375	[*]:[cH]:[c](:n:[*]):	0.202	19 out of 19
FCFP_12	713358128	[c](:[*]):[*] c	0.2	17 out of 17
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	[*][c](:[*]):[c]1:[c] (CI):[cH]:[cH]:[*]:[c]:1:[*]	-0.132	2 out of 3
FCFP_12	1	CI NH CI NH CI S::0	0	872 out of 1051
FCFP_12	307419094	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52



C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1

Enrichment: 1.18

Bayesian Score: 2.81

Mahalanobis Distance: 3.89 Mahalanobis Distance p-value: 1

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

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

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

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

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

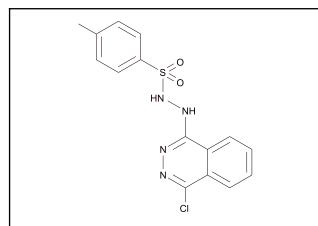
Structural Similar Compounds				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	5-NORBORNENE-2;3- DICARBOXYLIC ACID; 1;4;5;6;7;7- HEXACHLORO-	
Structure	HO sample of the NH 2	O HN MNH 2	OHCI CI CI OH	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.592	0.635	0.662	
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72	

Model Applicability

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

Feature Contribution					
	Top fea	atures for positive o	ontribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	1747237384	[*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44	

FCFP_12	178336375	[*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	[*]:n:[c]1:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	[*][c](:[*]):[c]1:[c] (CI):[cH]:[cH]:[*]:[c]:1:[*]	-0.132	2 out of 3
FCFP_12	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	1872154524	CI NH CI NH CI	0	563 out of 690



C₁₅H₁₃CIN₄O₂S

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1
Enrichment: 1.18
Payosian Socre: 1

Bayesian Score: 1.69
Mahalanobis Distance: 3.74

Mahalanobis Distance p-value: 1

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

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

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

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

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

Structural Similar Compounds

	<u>•</u>		
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	2;2';-Dihydroxy-4;4'- dimethoxybenzophenone
Structure	HO show the NH 2	HN MNH 2	OH OH HOW THO
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.524	0.584	0.609
Reference	28ZPAK 239;72	28ZPAK-;124;72	J. Am. Coll. Toxicol. 2(5):35;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: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

Feature Contribution

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	17	[*]:n:[*]	0.189	48 out of 49	

FCFP_12	71476542		0.175	81 out of 84
FCFP_12	4427049	[*]:[c](:[*])Cl	0.167	4 out of 4
	Top Fea	tures for negative of	contribution)
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	632767364	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](C):[cH]:[cH]:1	0	4 out of 5
FCFP_12	-453677277	[*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*]) (=[*])[*]	0	264 out of 323
FCFP_12	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	0	237 out of 291

C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248
Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1

Enrichment: 1.18
Bayesian Score: 2.11

Mahalanobis Distance: 3.69

Mahalanobis Distance p-value: 1

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

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

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

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

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

Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	Anthraquinone; 1-amino- 2-bromo-4-hydroxy-
Structure	HO the NH 2	O HN MNH 2	HO MANH 2
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.530	0.579	0.602
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: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

Feature Co	ntribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	-1508180856	[*][c]1:[cH]:[cH]:[c] (F):[cH]:[cH]:1	0.2	17 out of 17		

FCFP_12	17	F Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z	0.189	48 out of 49
FCFP_12	-149636017	[*]:n:[*]	0.184	7 out of 7
	Top Fea	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-453677277	[*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*]) (=[*])[*]	0	264 out of 323
FCFP_12	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	203677720	[*]S(=[*])(=[*])[c](: [cH]:[*]):[cH]:[*]	O	319 out of 382

C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1

Enrichment: 1.18

Bayesian Score: 1.17

Mahalanobis Distance: 3.64 Mahalanobis Distance p-value: 1

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

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

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

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

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

Structural Similar Compounds				
Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	1-AMINO-4- BENZOYLAMINO- ANTHRAQUINONE	Anthraquinone; 1-amino- 2-bromo-4-hydroxy-	
Structure	HO the hand of the hold of the	O HN MH 2	HO MNH 2	
Actual Endpoint	Irritant	Irritant	Non-Irritant	
Predicted Endpoint	Irritant	Irritant	Non-Irritant	
Distance	0.521	0.567	0.582	
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: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

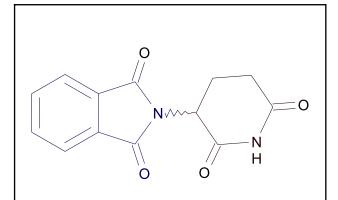
Feature Co	ntribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	17	[*]:n:[*]	0.189	48 out of 49		

FCFP_12	71476542	S O O N N N N N N N N N N N N N N N N N	0.175	81 out of 84
FCFP_12	4427049	[*]:[c](:[*])CI	0.167	4 out of 4
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]: [cH]:1	-0.0964	107 out of 146
FCFP_12	-453677277	[*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*]) (=[*])[*]	0	264 out of 323
FCFP_12	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	0	237 out of 291

Prehled Prumyslove Toxikologie; Organicke

Latky; Marhold; J. pp

868.86



 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Structural Similar Compounds				
ETHANOL;2- (PHENYLSULFONYL)-	P-ACETOPHENETIDIDE;3-NITRO-	3(2H)-Pyridazinone; 5- amino-4-chloro-2-phenyl-		
OH OO	O No -	N N N N N N N N N N N N N N N N N N N		
Irritant	Irritant	Irritant		
Irritant	Irritant	Irritant		
0.570	0.582	0.611		
	ETHANOL;2- (PHENYLSULFONYL)-	ETHANOL;2-(PHENYLSULFONYL)- P-ACETOPHENETIDIDE;3- NITRO- Irritant Irritant Irritant Irritant Irritant		

Model Prediction

Prediction: Non-Irritant

Probability: 0.971 Enrichment: 1.14 Bayesian Score: -1.01 Mahalanobis Distance: 6.67

Mahalanobis Distance p-value: 1

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

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

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

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

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

Model Applicability

Reference

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

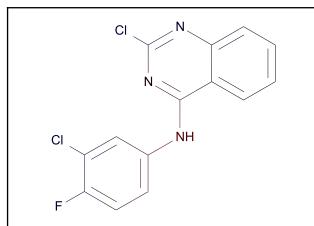
28ZPAK-;115;72

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

28ZPAK-;200;72

Feature Contribution				
Top fea	ntures for positive o	ontribution		
Bit/Smiles	Feature Structure	Score	Irritant in training set	
-885550502	[*]C(=[*])NC(=[*])[*]	0.18	64 out of 66	
	Bit/Smiles	Bit/Smiles Feature Structure -885550502	-885550502 0.18	

FCFP_12	566058135	[*]CC(=O)N[*]	0.163	23 out of 24
FCFP_12	-1553874037	[*]C([*])N1C(=[*])[*] :[*]C1=[*]	0.163	42 out of 44
		res for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1393189956	[*]N1C(=[*])[c]2:[cH] :[cH]:[cH]:[cH]:[c]: 2C1=O	-0.782	5 out of 15
FCFP_12	2139882011	[*]N1C(=[*])[c]2:[cH] :[*]:[cH]:[cH]:[c]:2 C1=O	-0.782	5 out of 15
FCFP_12	-1549163031	[*]N1[*][*]:[c](:[*]) C1=O	-0.623	16 out of 38



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.251 Enrichment: 0.78 Bayesian Score: -2.7

Mahalanobis Distance: 9.27

Mahalanobis Distance p-value: 0.714

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Mestranol	Levonorgestrel
Structure	OH	HO MAN TO THE OWN THE	OH NAME OH
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.607	0.610	0.611
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

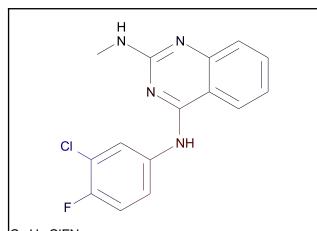
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

Feature Contribution

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	-177077903	CI NH CI NH [*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10	

ECFP_12	888054369	CI NH [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	1049768340	CI NH CI NH [*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	CI NH NH [*][c](:[*]):[c](CI): [cH]:[*]	-1.11	2 out of 26
ECFP_12	99947387	CI NH NH [*]:[c](:[*])CI	-0.817	8 out of 62
ECFP_12	-817402818	CI NH NH [*]CI	-0.368	17 out of 80



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.258
Enrichment: 0.802
Bayesian Score: -2.25
Mahalanobis Distance: 9.5

Mahalanobis Distance p-value: 0.607

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

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

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

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

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

Structural Similar Compounds

<u> </u>				
Name	Meclofenamate	Meclofenamate Diclofenac		
Structure	OH HN CCI	OH OH NNH	НО	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen	
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen	
Distance	0.540	0.566	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.
- 2. Unknown ECFP_2 feature: -1240852525: [*]:[c](:[*])NC

Feature Contribution

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-177077903	[*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10

ECFP_12	888054369	H NH NH	0.454	5 out of 9
		[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1		
ECFP_12	1049768340	H NH	0.421	1 out of 1
		[*]N[c](:n:[*]):[c](: [*]):[*]		
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	CI NH	-1.11	2 out of 26
		[*][c](:[*]):[c](CI): [cH]:[*]		
ECFP_12	99947387	H	-0.817	8 out of 62
		CI NH [*]:[c](:[*])CI		
ECFP_12	-817402818	H N N N N N N N N N N N N N N N N N N N	-0.368	17 out of 80
		[*]CI		

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.251 Enrichment: 0.779 Bayesian Score: -2.72 Mahalanobis Distance: 10.1

Mahalanobis Distance p-value: 0.341

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Meclofenamate	Mefloquine	
Structure	OH OH	OH HN CI	F OH	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Distance	0.568	0.569	0.625	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	

Model Applicability

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

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

Feature Contribution Top features for positive contribution				
ECFP_12	-177077903	[*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10

ECFP_12	888054369	[*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	1434334340	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.421	1 out of 1
	Top Feat	ures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-1.11	2 out of 26
ECFP_12	99947387	CI NH NH [*]:[c](:[*])CI	-0.817	8 out of 62
ECFP_12	491100606	H N N N N N N N N N	-0.485	0 out of 2

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

CI N
CI NH CI

C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.228
Enrichment: 0.708
Bayesian Score: -4.41
Mahalanobis Distance: 8.07

Mahalanobis Distance p-value: 0.984

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

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

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

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

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

Structural Simil	ar Compounds		
Name	Mestranol	Nafenopin	Levonorgestrel
Structure	HO NO	OH OH	O THE STATE OF THE
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.639	0.649	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: 1126732433: [*]:n:[c](Cl):n:[*]

ECFP_12	888054369	CI NH CI [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	-1661653144	[*][c](:[*]):[c](:[c] ([*]):[*]):[c](:[*]) :[*]	0.437	2 out of 3
	Top Featur	es for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	CI NH CI NH CI [*][c](:[*]):[c](CI): [cH]:[*]	-1.11	2 out of 26
ECFP_12	1641317964	CI NH CI NH CI [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	-0.929	1 out of 13
ECFP_12	99947387	CI NH CI	-0.817	8 out of 62
		[*]:[c](:[*])Cl		

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Eval.& Res./Off. Testing &

Res.) Sept. 1997

CI N N CI O N H

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227
Enrichment: 0.705
Bayesian Score: -4.5
Mahalanobis Distance: 10.8

Mahalanobis Distance p-value: 0.0952

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

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

Enrichment: An 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	Metolazone	Naltrexone	Indapamide
Structure	H ₂ N State of the NH	HO MANOH	HN H ₂ N O
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.574	0.589	0.590
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. OPS PC30 out of range. Value: 3.7537. Training min, max, SD, explained variance: -2.9582, 2.682, 0.9684, 0.0098.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](CI):n:[*]

Res.) Sept. 1997

Eval.& Res./Off. Testing &

Feature Contribution					
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	-742538367	CI N N CI N N N N CI N N N N N N N N N N	0.445	3 out of 5	

ECFP_12	-1661653144	[*][c](:[*]):[c](:[*]) :[*]	0.437	2 out of 3
ECFP_12	1049768340	CI NH CI NH CI [*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1
	Top Feat	ures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-1.11	2 out of 26
ECFP_12	1641317964	CI NH	-0.929	1 out of 13
ECFP_12	99947387	CI NH CI NH CI NH CI (**):[c](:[*])CI	-0.817	8 out of 62

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

CI N CI
 NH
HN
O NH 2

C₀H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227
Enrichment: 0.704
Bayesian Score: -4.51
Mahalanobis Distance: 9.86

Mahalanobis Distance p-value: 0.434

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	Pyrimethamine	Lamotrigine	Guanfacine
Structure	NH 2 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	O NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.627	0.635	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

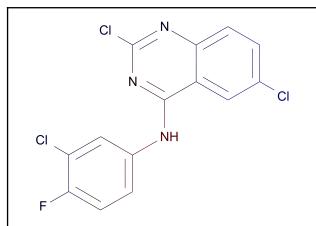
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
- 3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]

Feature Contribution					
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	1049768340	CI NH CI NH CI NH2 [*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1	

ECFP_12	432952415	CI NH CI NH CI NH 2 [*]NC(=O)N	0.208	1 out of 2
ECFP_12	-1074141656	CI NH CI NH CI NH 2 [*]=0	0.103	86 out of 248
		atures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	99947387	CI NH	-0.817	8 out of 62
ECFP_12	1854732111	CI NH NH NH ₂ [*]:[c]1:[*]:[cH]:[c] (CI):[cH]:[cH]:1	-0.816	4 out of 33
ECFP_12	-176494269	CI NH NH NH ₂ [*]:[cH]:[c](CI):[cH]	-0.714	5 out of 36

:[*]



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.217 Enrichment: 0.673 Bayesian Score: -5.4

Mahalanobis Distance: 9.72

Mahalanobis Distance p-value: 0.503

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Mestranol	Loratidine	Levonorgestrel
Structure	HO THE STATE OF TH	T _t	O OH
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.639	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

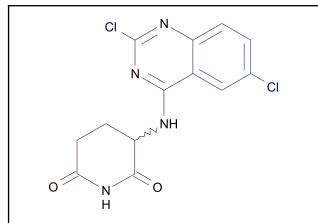
Model Applicability

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: 1126732433: [*]:n:[c](Cl):n:[*]

Feature Contribution Top features for positive contribution					
-177077903	[*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10		
	Top fea	Top features for positive of Bit/Smiles Feature Structure -177077903	Top features for positive contribution Bit/Smiles Feature Structure Score -177077903 0.529		

ECFP_12	888054369	CI NH CI N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	1434334340	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.421	1 out of 1
	Top Fea	ntures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-1.11	2 out of 26
ECFP_12	99947387	CI NH CI NH [*]:[c](:[*])CI	-0.817	8 out of 62
ECFP_12	1854732111	CI NH CI NH (CI):[c] 1:[c]1:[c]1:[cH]:[cH]:1	-0.816	4 out of 33



 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227
Enrichment: 0.707
Bayesian Score: -4.46
Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.0199

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

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

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

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

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

otructural offinial compounds					
Name	Metolazone	Naltrexone	Tolazamide		
Structure	H ₂ N NI	HO stands of the			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.589

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.594

Model Applicability

Predicted Endpoint

Distance

Reference

Structural Similar Compounds

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

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

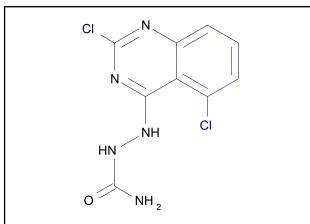
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

Non-Carcinogen

Res.) Sept. 1997

0.578

ECFP_12	1049768340	CI NH CI NH CI NH (I N)	0.421	1 out of 1
ECFP_12	51876938	CI NH CI NH (*)CCC(=[*])[*]	0.232	18 out of 45
		tures for negative o	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	99947387	CI NH CI NH (CI	-0.817	8 out of 62
ECFP_12	1854732111	CI NH OI NH	-0.816	4 out of 33
ECFP_12	-176494269	CI NH	-0.714	5 out of 36



C₉H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.226 Enrichment: 0.703 Bayesian Score: -4.56 Mahalanobis Distance: 8.39

Mahalanobis Distance p-value: 0.957

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Guanfacine	Guanabenz
Structure	H ₂ N ^M NH ₂	CI NH H ₂ N NH	CI NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.608	0.609	0.628
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

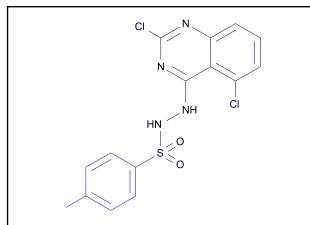
Model Applicability

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: 1126732433: [*]:n:[c](CI):n:[*]
- 3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	-1661653144	[*][c](:[*]):[c](:[*]) :[*]	0.437	2 out of 3	

ECFP_12	1049768340	CI NH2 CI ONH2 [*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1
ECFP_12	432952415	CI NH CI NH CI NH 2 [*]NC(=O)N	0.208	1 out of 2
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	CI NH CI NH CI NH _E [*][c](:[*]):[c](CI): [cH]:[*]	-1.11	2 out of 26
ECFP_12	1641317964	CI NH2 CI NH2 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	-0.929	1 out of 13
ECFP_12	99947387	CI NH CI HN NH CI NH ₂ [*]:[c](:[*])CI	-0.817	8 out of 62



C₁₅H₁₂Cl₂N₄O₂S

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.189
Enrichment: 0.588
Bayesian Score: -8.62
Mahalanobis Distance: 8.88

Mahalanobis Distance p-value: 0.858

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Niclosamide	Indan

Name	Niclosamide	Indapamide	Torsemide
Structure	CIANOH NO CIANON O	HN THE NOTICE OF THE NAME OF T	HN NH NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.647	0.663	0.670
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: 1126732433: [*]:n:[c](CI):n:[*]
- 3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
- 4. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

Top features for positive contribution					
-1661653144	[*][c](:[*]):[c](:[c] ([*]):[*]):[c](:[*]) :[*]	0.437	2 out of 3		
	Top fea	Top features for positive of Bit/Smiles Feature Structure -1661653144 -1661653144 -1661653144	Top features for positive contribution Bit/Smiles Feature Structure Score -1661653144		

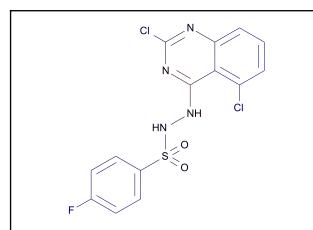
ECFP_12	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1
ECFP_12	-1074141656	CI NH CI HN OH CI S:0 [*]=0	0.103	86 out of 248
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-1.11	2 out of 26
ECFP_12	-533780882	C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.06	0 out of 6
ECFP_12	-1926229349	[*][c]1:[cH]:[c] (C):[cH]:[cH]:1	-1.06	0 out of 6

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Eval.& Res./Off. Testing &

Res.) Sept. 1997



C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.202
Enrichment: 0.627
Bayesian Score: -6.95
Mahalanobis Distance: 9.01

Mahalanobis Distance p-value: 0.816

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

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

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

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

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

Structural Similar Compounds				
Name	Bicalutamide	Niclosamide	Indapamide	
Structure	HO HO ATANA TANANA TANA	CIAN OH	HN H ₂ N O	
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen	
Distance	0.632	0.637	0.645	
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 &

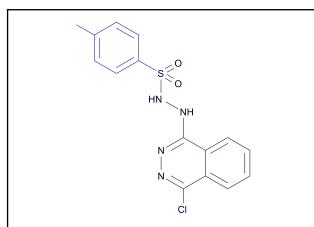
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](CI):n:[*]

Res.) Sept. 1997

- 3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
- 4. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	-1661653144	[*][c](:[*]):[c](:[c] ([*]):[*]):[c](:[*]) :[*]	0.437	2 out of 3	

ECFP_12	74189795	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:[cH]:	0.421	1 out of 1
ECFP_12	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1
	Top Fea	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	[*][c](:[*]):[c](CI): [cH]:[*]	-1.11	2 out of 26
ECFP_12	1641317964	[*]:[c]1:[*]:[cH]:[cH	-0.929	1 out of 13
ECFP_12	99947387	CI NH CI HN NH CI S:00 S:00 [*]:[c](:[*])CI	-0.817	8 out of 62



C₁₅H₁₃CIN₄O₂S

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.202 Enrichment: 0.628 Bayesian Score: -6.93 Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.3

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 C	Compounds
----------------------	-----------

	I	I	I
Name	Niclosamide	Indapamide	Tolazamide
Structure	CIANDO CI	HN H ₂ N O	T T N O T N
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.592	0.594	0.605
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

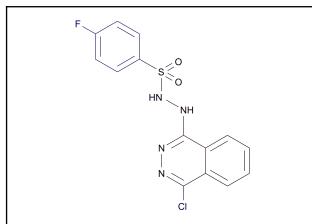
Model Applicability

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.
- Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
- Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

i catalo oo					
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
ECFP_12	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1	

ECFP_12	1306977740	[*][c](:[*]):[c]f ¹ [cH]:[cH]:[cH]:[c] :1:[*]	0.271	4 out of 9
ECFP_12	1639858918	[*][c](:[*]):[c] ^{fl} [cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.119	4 out of 11
	Top Feat	tures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1926229349	[*][c]1:[cH]:[cH]:[c] (C):[cH]:[cH]:1	-1.06	0 out of 6
ECFP_12	-533780882	C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.06	0 out of 6
ECFP_12	99947387	[*]:[c](:[*])CI	-0.817	8 out of 62



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248
Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.224
Enrichment: 0.695
Bayesian Score: -4.77
Mahalanobis Distance: 9.96

Mahalanobis Distance p-value: 0.385

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Acetohexamide
Structure	H ₂ N _C O	CI AND CI	HN HN O
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.583	0.595	0.598
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
- 3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

i oatai o oo	Catal C Contribution					
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1		

ECFP_12	74189795	[*]S(=[*])(=[*])(^C]1: [cH]:[cH]:[c](F):[cH]:[cH]:1	0.421	1 out of 1
ECFP_12	1306977740	[*][c](:[*]):[c] ^f .[cH]:[cH]:[cH]:[c] :1:[*]	0.271	4 out of 9
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	99947387	[*]:[c](:[*])CI	-0.817	8 out of 62
ECFP_12	-296909061	F[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.56	1 out of 8
ECFP_12	-1659169698	[*][c]1:[cH]:[cH]:1	-0.56	1 out of 8

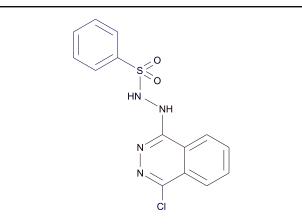
Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.584



C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.223
Enrichment: 0.693
Bayesian Score: -4.83
Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.137

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

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

Enrichment: An 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	Acetohexamide	Mebendazole	Indapamide
Structure	HN	Z T Z T Z T Z T Z T Z T Z T Z T Z T Z T	HN 2N O
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen

Model Applicability

Predicted Endpoint

Distance

Reference

Structural Similar Compounds

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

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.581

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

2. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]

Non-Carcinogen

Res.) Sept. 1997

0.579

3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

Feature Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	1049768340	[*]N[c](:n:[*]):[c](: [*]):[*]	0.421	1 out of 1		

ECFP_12	1306977740	[*][c](:[*]):[cf]:[cH]:[c]:1:[*]	0.271	4 out of 9
ECFP_12	1639858918	[*][c](:[*]):[cf1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.119	4 out of 11
	Top Feat	tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	99947387	(*]:[c](:[*])CI	-0.817	8 out of 62
ECFP_12	-281505363	[*][c]1:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	1571214559	[*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64

Thalidomide

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

O //
$N \sim N$
O O H

C₁₃H₁₀N₂O₄

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.283
Enrichment: 0.879
Bayesian Score: -0.795
Mahalanobis Distance: 8.89

Mahalanobis Distance p-value: 0.853

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

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

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

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

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

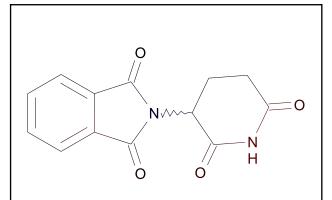
Structural Similar Compounds					
Name	Pemoline	Nalidixic acid	Milrinone		
Structure	NH ₂	HOOO	N NH NH		
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen		
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen		
Distance	0.629	0.638	0.647		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & 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 Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
ECFP_12	-742538367	[*]CC(=O)N[*]	0.445	3 out of 5		

ECFP_12	51876938	[*]CCC(=[*])[*]	0.232	18 out of 45
ECFP_12	-2105515128	[*]=C1[*]CCC(=O)N1	0.208	1 out of 2
		res for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-867777309	[*]NC(=O)C([*])[*]	-0.661	0 out of 3
ECFP_12	1997021792	[*]:[cH]:[cH]:[*	-0.296	36 out of 156
ECFP_12	-957084426	[*]C([*])N1C(=[*])[*] :[*]C1=[*]	-0.272	0 out of 1



 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.524
Enrichment: 1.4
Bayesian Score: 0.842

Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 0.000809

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

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

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

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

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

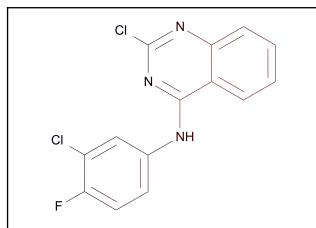
Name	Aminoglutethimide	Phenobarbital	Nalidixic acid
Structure	NH ₂	HN O	HOO
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.579	0.599	0.599
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Contribution						
	Top fe	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
SCFP_4	387787917	[*]=C1[*][c]2:[*]: [cH]:[cH]:[c]1:	0.449	6 out of 11		

SCFP_4	1257084377	[*]N1[*][*]:[c](:[*]) C1=O	0.44	3 out of 5
SCFP_4	345243876	[*]=C1[*]CCC(=O)N1	0.419	1 out of 1
		tures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_4	-1946889102	[,]cc(N([,])[,])c(=[,,])C(=[,,]) H	-0.816	0 out of 4
SCFP_4	399659969	[*]C([*])N1C(=[*])[*] :[*]C1=[*]	-0.666	0 out of 3
SCFP_4	-1272798659	[*]CCC([*])[*]	-0.421	10 out of 50



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.454
Enrichment: 1.36
Bayesian Score: 3.08
Mahalanobis Distance: 13.9

Mahalanobis Distance p-value: 5.62e-005

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

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

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

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

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

Structural Similar Compounds					
Name	Meclofenamate	Nafenopin	Mestranol		
Structure	OH HN CI	OH	HO THO		

Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.605	0.609	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

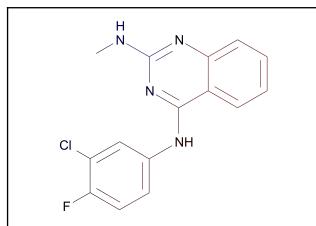
Model Applicability

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

1. OPS PC15 out of range. Value: 5.9999. Training min, max, SD, explained variance: -4.0246, 5.5336, 1.388, 0.0180.

Fingerprint Bit/Smiles Feature Structure Score Carcinogen in training set SCFP_6 1651620003 CILIT OF THE STRUCTURE SCORE CARCINOGEN IN TRAINING SET TO UT OF THE STRUCTURE SCORE CARCINOGEN IN TRAINING SET TO UT OF THE STRUCTURE SCORE CARCINOGEN IN TRAINING SET TO UT OF THE STRUCTURE SCORE CARCINOGEN IN TRAINING SET TO UT OF THE STRUCTURE SCORE SC

SCFP_6	-1379673609	CI NH E**[[c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.526	11 out of 19
SCFP_6	1655199790	CI NH CI NH (*]:n:[c]1:[cH]:[cH]: [cH]:[cH]:[c]:1:[*]	0.52	5 out of 8
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	CI NH CI NH [*]:[c](:[*])CI	-0.315	14 out of 61
SCFP_6	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.3	14 out of 60
SCFP_6	403834996	CI NH H H H H H H H H H H H H H H H H H H	-0.264	1 out of 5



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4 Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.417 Enrichment: 1.25 Bayesian Score: 1.98 Mahalanobis Distance: 13.9

Mahalanobis Distance p-value: 5.82e-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

	-	Table 1	T
Name	Meclofenamate	Diclofenac	Phenolphthalein
Structure	OH HN CCI	OH OH NNH	НО
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.525	0.544	0.592
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

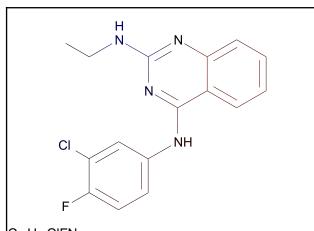
Model Applicability

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

1. OPS PC15 out of range. Value: 5.5915. Training min, max, SD, explained variance: -4.0246, 5.5336, 1.388, 0.0180.

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	1651620003	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[c] :1:[*]	0.643	7 out of 10	

SCFP_6	-1379673609	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.526	11 out of 19
SCFP_6	1655199790	[*]:n:[c]1:[cH]:[cH]: [cH]:[cH]:[c]:1:[*]	0.52	5 out of 8
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	18117904	CI NH NH [*]:[c](:[*])NC	-0.578	1 out of 8
SCFP_6	-1545804258	[*]N[c]1:n:[*]:[c](:[*]):[c](N[*]):n:1	-0.578	1 out of 8
SCFP_6	649648475	H N N N N N N N N N N N N N N N N N N N	-0.48	2 out of 12



C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.399
Enrichment: 1.19
Bayesian Score: 1.42
Mahalanobis Distance: 16

Mahalanobis Distance p-value: 4.53e-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.

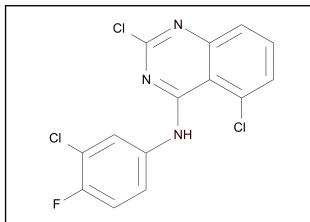
Name	ilar Compounds Diclofenac	•		
Structure	OH OH	OH HN N	F F F NH	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	
Distance	0.550	0.554	0.610	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	

Model Applicability

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

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	1651620003	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[c] :1:[*]	0.643	7 out of 10		

SCFP_6	-1379673609	[*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.526	11 out of 19
SCFP_6	1655199790	[*]:n:[c]1:[cH]:[cH]: [cH]:[cH]:[c]1:[*]	0.52	5 out of 8
		ures for negative of	-	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1545804258	[*]N[c]1:n:[*]:[c](:[*]):[c](N[*]):n:1	-0.578	1 out of 8
SCFP_6	18117904	CI NH NH [*]:[c](:[*])NC	-0.578	1 out of 8
SCFP_6	649648475	H N N N N N N N N N N N N N N N N N N N	-0.48	2 out of 12



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.364
Enrichment: 1.09
Bayesian Score: 0.273
Mahalanobis Distance: 13.5

Mahalanobis Distance p-value: 0.000258

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

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

Enrichment: An 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	Mestranol	Loratidine	Meclofenamate
Structure	HO THE STATE OF TH	Z _t	OH HN
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.642	0.642	
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	g US FDA (Centre for Drug & US FDA (Centre fo Eval.& Res./Off. Testing & Res.) Sept. 1997 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	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.415	1 out of 1		

SCFP_6	1334878018	CI NH CI [*]N[c](:n:[*]):[c](: [*]):[*]	0.333	8 out of 17
SCFP_6	-1380909229	CI NH CI [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.287	17 out of 39
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	CI NH CI NH CI [*]:[c](:[*])CI	-0.315	14 out of 61
SCFP_6	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	CI NH CI NH CI (1):[c]1:[c]1:[c]1:[c]1:[c]1:[c]1:[c]1:[c]1	-0.289	12 out of 51

Res.) Sept. 1997

CI N N CI O N H

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.366 Enrichment: 1.1

Bayesian Score: 0.331 Mahalanobis Distance: 14.1

Mahalanobis Distance p-value: 3.26e-005

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

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

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

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

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

Name	Metolazone	Indapamide	Tolazamide
Structure	H ₂ N S THE STATE OF THE STATE	HN 12N O	H 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	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.548	0.566	0.573
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &	US FDA (Centre for Drug Eval.& Res./Off. Testing &

Model Applicability

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

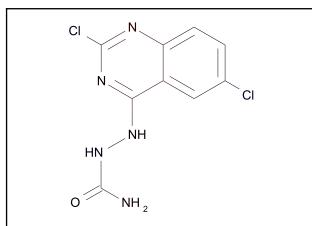
Res.) Sept. 1997

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

Res.) Sept. 1997

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	-1533214013	CI N N CI N N CI N N CI N CI N CI N CI	0.415	1 out of 1		

SCFP_6	395945879	CI N NH CI NH CI CI NH CI CI NH CI N	0.415	1 out of 1
SCFP_6	345243876	CI N N CI N N CI N CI N CI N CI N CI N	0.415	1 out of 1
	Top Fea	atures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	18117904	CI NH CI NH CI (**).	-0.578	1 out of 8
SCFP_6	-52074512	CI NH CI NH CI H CI (**)CI (**)CI	-0.315	14 out of 61
SCFP_6	-601571304	CI NH CI NH CI NH CI (CI): [cH]:[*]	-0.3	14 out of 60



C₀H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.348 Enrichment: 1.04 Bayesian Score: -0.291

Mahalanobis Distance: 14.3

Mahalanobis Distance p-value: 1.37e-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 C	compounds
----------------------	-----------

Name	Pyrimethamine	Lamotrigine	Guanfacine
Structure	H ₂ N ^M	H ₂ N ² N ² NH ₂	NH H ₂ N NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.603	0.608	0.622
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

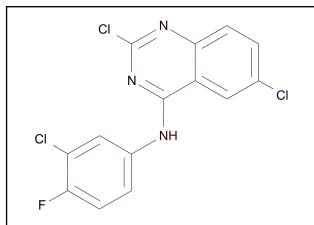
Model Applicability

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

Feature	Contr	ibution
----------------	-------	---------

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	-1357949052	CI NH	0.494	8 out of 14	

SCFP_6	1334878018	CI NH2 CI NH2 (in:[*]):[c](: [*]):[*]	0.333	8 out of 17
SCFP_6	112346096	CI NH	0.276	13 out of 30
		ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	CI NH CI NH CI NH 2 [*]:[c](:[*])CI	-0.315	14 out of 61
SCFP_6	-601571304	CI NH CI NH CI NH½ [*][c](:[*]):[c](CI): [cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	CI NH NH NH ₂ [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	-0.289	12 out of 51



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.365 Enrichment: 1.09 Bayesian Score: 0.284 Mahalanobis Distance: 13.6

Mahalanobis Distance p-value: 0.000236

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Loratidine	Mestranol	Meclofenamate
Structure	CI V N O	HO TO THE TOTAL TOTAL TO THE TOTAL TOTAL TO THE TOTAL TOTAL TO THE TOTAL	OH HN CI
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.637	0.642	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 8

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.

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	-300914917	[*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	0.415	1 out of 1		

SCFP_6	1334878018	CI NH CI [*]N[c](:n:[*]):[c](: [*]):[*]	0.333	8 out of 17
SCFP_6	-1380909229	CI NH CI NH CI NH (I*)N[c]1:[cH]:[c] ([*]):[cH]:[cH]:1	0.287	17 out of 39
	Top Fea	tures for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	CI NH CI [*]:[c](:[*])CI	-0.315	14 out of 61
SCFP_6	-601571304	CI NH CI NH [*][c](:[*]):[c](CI): [cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	CI NH CI NH (*):[c]1:[*]:[cH]:[cH]:[cH]:[c]:1CI	-0.289	12 out of 51

CI N CI NH O H

 $C_{13}H_{10}CI_{2}N_{4}O_{2}$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.366 Enrichment: 1.1

Bayesian Score: 0.343 Mahalanobis Distance: 14.4

Mahalanobis Distance p-value: 9.01e-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	Metolazone	Indapamide	Tolazamide
Structure	H ₂ N S 1 NH	HN 12 N O	H H N 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	Non-Carcinogen	Non-Carcinogen
Distance	0.554	0.572	0.573
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 8

Model Applicability

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

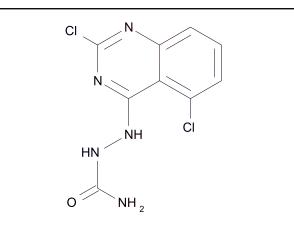
Res.) Sept. 1997

Res.) Sept. 1997

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

Feature Contribution						
	Top fe	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	345243876	CI NH	0.415	1 out of 1		

SCFP_6	395945879	CI NH CI NH ("]C1["]CC(=0)NC1=0	0.415	1 out of 1
SCFP_6	-1533214013	CI NH CI NH CI STATE CI STATE CI CI CI CI CI CI CI C	0.415	1 out of 1
		tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	18117904	CI NH CI NH (F)	-0.578	1 out of 8
SCFP_6	-52074512	CI NH CI NH (CI	-0.315	14 out of 61
SCFP_6	-601571304	CI NH CI NH (CI NH) (C	-0.3	14 out of 60



C₀H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.348 Enrichment: 1.04

Bayesian Score: -0.303 Mahalanobis Distance: 13.8

Mahalanobis Distance p-value: 9.34e-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	Lamotrigine	Pyrimethamine	Guanfacine	
Structure	CI	NH 2 N N	CI	

	H ₂ NNNNNH ₂	CI	H ₂ N NH
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.600	0.611	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 Co	Feature Contribution				
Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set	
SCFP_6	-1357949052	CI N N CI N N CI N N CI N N N CI N N N N	0.494	8 out of 14	

SCFP_6	1334878018	CI NH2 CI NH2 [*]N[c](:n:[*]):[c](: [*]):[*]	0.333	8 out of 17
SCFP_6	112346096	CI NH2 CI NH2 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.276	13 out of 30
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	CI N N CI N N N CI N N N CI N N N N N N	-0.315	14 out of 61
SCFP_6	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	CI NH2 CI NH2 CI NH2 (CI NH2 CI NH2 C	-0.289	12 out of 51

Carcinogen

Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug Eval.& Res./Off. Testing &

0.630

 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.284 Enrichment: 0.849 Bayesian Score: -2.82 Mahalanobis Distance: 14.5

Mahalanobis Distance p-value: 5.33e-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	Indapamide	Indomethacin	Torsemide
Structure	HN 22 S	O HO HO	HN NH NH

Non-Carcinogen

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.628

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

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	1334878018	[*]N[c](:n:[*]):[c](: [*]):[*]	0.333	8 out of 17		

SCFP_6	136686699	CI N NH CI NH CI S: 0 S: 0 CI S: (a) [*]:[c](:[*])C	0.287	17 out of 39
SCFP_6	112346096	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.276	13 out of 30
	Top Feat	tures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1892882306	CI N CI N CI	-0.957	0 out of 5
SCFP_6	2054891299	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](C):[cH]:[cH]:1	-0.496	0 out of 2
SCFP_6	795925860	C K C K N C K H [*][c]1:[cH]:[cH]:[c] (C):[cH]:[cH]:1	-0.38	1 out of 6

C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.347 Enrichment: 1.04

Bayesian Score: -0.338 Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 7.06e-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 cutoff

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Bicalutamide	Indapamide	Niclosamide	
Structure	HO HO HN MAN THE F	HN H ₂ N O	CI _M	
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen	
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen	
Distance	0.612	0.613	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	

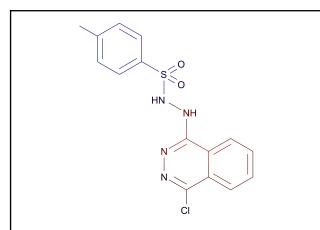
Model Applicability

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

1. OPS PC1 out of range. Value: -5.807. Training min, max, SD, explained variance: -5.694, 8.348, 2.949, 0.0811.

Feature Co	Feature Contribution					
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	2144224020	[*]S(=[*])(=[*])(c]1: [cH]:[cH]:[cH]:1	0.415	1 out of 1		

SCFP_6	1334878018	[*]N[c](:n:[*]):[c](: [*]):[*]	0.333	8 out of 17
SCFP_6	112346096	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.276	13 out of 30
	Top Feat	ures for negative of		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	CI NH CI NH CI S::0 [*]:[c](:[*])CI	-0.315	14 out of 61
SCFP_6	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	[*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.289	12 out of 51



 $C_{15}H_{13}CIN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.349 Enrichment: 1.04

Bayesian Score: -0.258 Mahalanobis Distance: 25.5

Mahalanobis Distance p-value: 8.64e-034

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Tolazamide	Niclosamide	
Structure	HN 22 S O	N H N N N N N N N N N N N N N N N N N N	CI AND TOH	
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen	

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.569

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.581

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.

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

US FDA (Centre for Drug

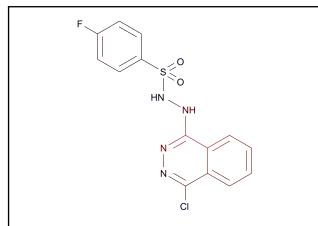
Eval.& Res./Off. Testing &

Non-Carcinogen

Res.) Sept. 1997

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	1651620003	[*][c](:[*]):[c]f ¹ .[cH]:[cH]:[cH]:[c] :1:[*]	0.643	7 out of 10		

SCFP_6	149212520	SOO H H H	0.543	9 out of 15
SCFP_6	-1379673609	[*][c](:[*]):n:n:[*] [*][c](:[*]):[c]f ¹ [cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.526	11 out of 19
	Top Feat	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1892882306	[*]NS(=O)(=O)(c)1:[cH]:[cH]:[c)(C);[cH]: ch]:1	-0.957	0 out of 5
SCFP_6	2054891299	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](C):[cH]:[cH]:1	-0.496	0 out of 2
SCFP_6	795925860	[*][c]1:[cH]:[cH]:[c] (C):[cH]:[cH]:1	-0.38	1 out of 6



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.437
Enrichment: 1.31
Bayesian Score: 2.56
Mahalanobis Distance: 25.8

Mahalanobis Distance p-value: 1.04e-034

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	Acetohexamide		
Structure	HN H ₂ N S	O CIANON ON O	HN O		
Actual Endpoint	Non-Carcinogen	Non-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.

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.585

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.587

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

US FDA (Centre for Drug

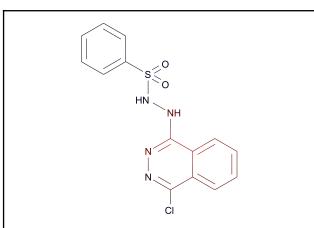
Eval.& Res./Off. Testing &

Non-Carcinogen

Res.) Sept. 1997

Feature Co	Feature Contribution					
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	1651620003	[*][c](:[*]):[c] f ² [cH]:[cH]:[cH]:[c] :1:[*]	0.643	7 out of 10		

SCFP_6	149212520	[*][c](:[*]):n:n:[*]	0.543	9 out of 15
SCFP_6	-1379673609	[*][c](:[*]):[c]f ² [cH]:[cH]:[cH]:[*]	0.526	11 out of 19
	Top Fea	tures for negative of	contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	[*]:[c](:[*])CI	-0.315	14 out of 61
SCFP_6	21	[*]S(=[*])(=[*])[*]	-0.283	7 out of 30
SCFP_6	1311429347	[*]S(=[*])(=O)[*]	-0.283	7 out of 30



C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.386 Enrichment: 1.16 Bayesian Score: 1

Mahalanobis Distance: 25.7

Mahalanobis Distance p-value: 1.97e-034

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Acetohexamide	Mebendazole		
Structure	HN SOO	HN	T N H N N H		
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen		

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.563

Non-Carcinogen

Res.) Sept. 1997

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

0.576

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.

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

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Non-Carcinogen

Res.) Sept. 1997

Feature Co	Feature Contribution					
	Top fea	tures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	1651620003	[*][c](:[*]):[c]f1:[cH]:[cH]:[cH]:[c] :1:[*]	0.643	7 out of 10		

SCFP_6	149212520	[*][c](:[*]):n:n:[*]	0.543	9 out of 15
SCFP_6	-1379673609	[*][c](:[*]):[c]1:[c]: 1:[*]	0.526	11 out of 19
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1653911926	[*][c]1:[cH]:[cH]:[cH]:1	-0.504	12 out of 64
SCFP_6	-52074512	[*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	1311429347	[*]S(=[*])(=O)[*]	-0.283	7 out of 30

Thalidomide

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen

0 $N \sim N$ $N \sim N$ $N \sim N$
O O

 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.332 Enrichment: 0.995 Bayesian Score: -0.857 Mahalanobis Distance: 13

Mahalanobis Distance p-value: 0.00164

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

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

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

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

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

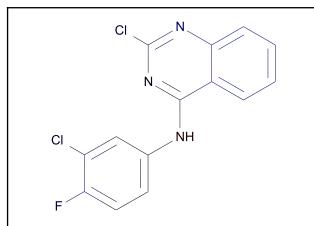
Structural Similar Compounds					
Name	Milrinone	Aminoglutethimide	Theophylline		
Structure	N NH	NH ₂	N N N N N N N N N N N N N N N N N N N		
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen		
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen		
Distance	0.582	0.597	0.614		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997		

Model Applicability

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

Feature Contribution						
	Top fea	tures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set		
SCFP_6	-1533214013	[*]C1CCC(=O)NC1=O	0.415	1 out of 1		

SCFP_6	345243876		0.415	1 out of 1
		[*]=C1[*]CCC(=O)N1		
SCFP_6	395945879		0.415	1 out of 1
		[*]C1[*]CC(=O)NC1=O		
		es for negative c		
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1946889102	N N N N N N N N N N N N N N N N N N N	-0.885	1 out of 12
		[*]CC(N((*))(*))C(=[*])(*]		
SCFP_6	399659969	N N N N N N N N N N N N N N N N N N N	-0.578	1 out of 8
		[*]C([*])N1C(=[*])[*] :[*]C1=[*]		
SCFP_6	1257084377		-0.436	4 out of 21
		[*]N1[*][*]:[c](:[*]) C1=O		



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.436 Enrichment: 1.05 Bayesian Score: -5.26 Mahalanobis Distance: 17.2

Mahalanobis Distance p-value: 9.6e-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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds					
Name	Mestranol	Nafenopin	Norethindrone		
Structure	HO TO THE TOTAL PART OF THE TO	OH OH	O H		
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen		
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen		
Distance	0.620	0.621	0.686		
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.

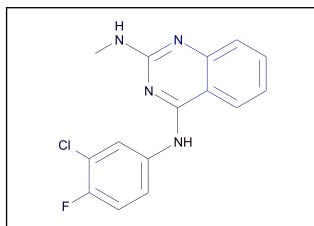
Res.) Sept. 1997

Res.) Sept. 1997

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

Feature Co	ntribution			
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	951581613	CI NH *]:[c](:[*])N[c](:[*)):[*]	0.383	1 out of 1

SCFP_8	-300914917	CI N H H H H H H H H H H H H H H H H H H	0.383	1 out of 1
SCFP_8	10	CI NH NH [*]N[*]	0.226	18 out of 39
		es for negative c		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-601571304	CI NH NH [*][c](:[*]):[c](CI): [cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	CI NH CI NH [*]:[c](:[*])CI	-0.707	2 out of 14
SCFP_8	-1381862798		-0.572	1 out of 7



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.412
Enrichment: 0.995
Bayesian Score: -5.85
Mahalanobis Distance: 18.4

Mahalanobis Distance p-value: 8.86e-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.

Structural Similar Compounds				
Name Phenolphthalein Doxefazepam				
Structure	^			

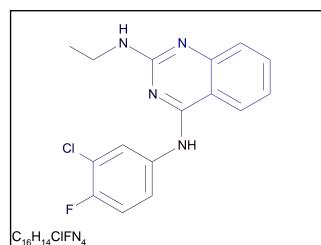
Name	Phenolphthalein	Doxefazepam	Diethylstilbesterol
Structure	НО	OH OOH	НО
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.599	0.679	0.699
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

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

Feature Contribution				
	Top fe	atures for positive of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	951581613	[*]:[c](:[*])N[c](:[*]):[*]	0.383	1 out of 1

SCFP_8	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.383	1 out of 1
SCFP_8	10	H N N N N N N N N N	0.226	18 out of 39
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	CI NH [*]:[c](:[*])CI	-0.707	2 out of 14
SCFP_8	-1381862798	[*]:n:[c]1:[cH]:[cH]: [cH]:[*]:(c]:1:[*]	-0.572	1 out of 7



Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.391 Enrichment: 0.945 Bayesian Score: -6.32 Mahalanobis Distance: 19.2

Mahalanobis Distance p-value: 1.65e-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	Phenolphthalein	Nafenopin	Diethylstilbesterol		
Structure	но	O OH	но		
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen		
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen		
Distance	0.665	0.686	0.686		
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997		

Model Applicability

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

Feature Co	ntribution			
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	951581613	[*]:[c](:[*])N[c](:[*]):[*]	0.383	1 out of 1

SCFP_8	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.383	1 out of 1
SCFP_8	10	H N N N N N N N N N	0.226	18 out of 39
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-52074512	CI NH [*]:[c](:[*])CI	-0.707	2 out of 14
SCFP_8	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.707	2 out of 14
SCFP_8	-1381862798	[*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.572	1 out of 7

CI NH CI

C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.417 Enrichment: 1.01 Bayesian Score: -5.74 Mahalanobis Distance: 18.5

Mahalanobis Distance p-value: 7.37e-009

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	Mestranol	Nafenopin	Sertraline
Structure	HO THE STATE OF TH	OH	CI MIN H
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.653	0.662	0.706
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	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.383	1 out of 1

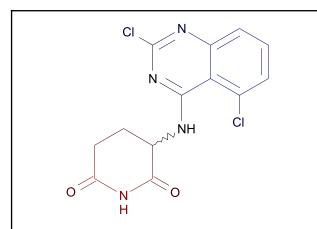
SCFP_8	951581613	[*]:[c](:[*])N[c](:[*]):[*]	0.383	1 out of 1
SCFP_8	10	CI NH CI [*]N[*]	0.226	18 out of 39
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	CI NH CI NH CI [*]:[c](:[*])CI	-0.707	2 out of 14
SCFP_8	-1378360678	CI NH CI NH CI CI NH CI CI NH CI CI NH CI CI NH CI N	-0.58	2 out of 12

0.666

US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997



 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.535 Enrichment: 1.29 Bayesian Score: -2

Mahalanobis Distance: 17.4

Mahalanobis Distance p-value: 6.3e-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	Doxefazepam	oxefazepam Phenolphthalein	
Structure	OH OH	НО	CI THE WOH
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-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.

0.663

US FDA (Centre for Drug

Res.) Sept. 1997

Eval.& Res./Off. Testing &

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

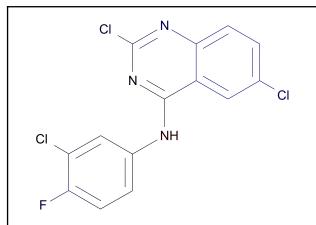
US FDA (Centre for Drug

Eval.& Res./Off. Testing &

Res.) Sept. 1997

Feature Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
SCFP_8	1631785938	(*)C(=[*))NC(=[*))[*]	0.574	4 out of 5		

SCFP_8	345243876	CI N N CI N N CI N CI N CI N CI N CI N	0.383	1 out of 1
SCFP_8	-1072897324	CI NH	0.383	1 out of 1
	Top Feat	tures for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	CI NH CI NH CI (**):[c](:[*])CI	-0.707	2 out of 14
SCFP_8	-1378360678	CI NH	-0.58	2 out of 12



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.413
Enrichment: 0.999
Bayesian Score: -5.82
Mahalanobis Distance: 18.5

Mahalanobis Distance p-value: 7.12e-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.

Name	Mestranol	Nafenopin	Chlorpromazine	
Structure	HO THOUSE THE PARTY OF THE PART	OH	No we consider the constant of	
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen	
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen	
Distance	0.653	0.667	0.707	
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	Multiple- Carcinogen in training set			
SCFP_8	-300914917	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.383	1 out of 1			

SCFP_8	951581613	CI N	0.383	1 out of 1
		CI		
		CINH		
		[*]:[c](:[*])N[c](:[*		
CCED 0	10]):[*]	0.000	10 out of 20
SCFP_8	10	CI	0.226	18 out of 39
		CINH		
		[*]N[*]		
		es for negative c		
•		Feature Structure		Multiple- Carcinogen in training set
SCFP_8	-601571304	CI N	-0.707	2 out of 14
		CI. NH		
		[*][c](:[*]):[c](CI): [cH]:[*]		
SCFP_8	-52074512	CI	-0.707	2 out of 14
		CI		
		[*]:[c](:[*])Cl		
SCFP_8	-1378360678	CI NO CI	-0.58	2 out of 12
		CINH		
		[*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl		

CI N CI NH O H

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.533 Enrichment: 1.29 Bayesian Score: -2.08 Mahalanobis Distance: 17

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 negative rate.

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Doxefazepam	Oxazepam	Phenolphthalein
Structure	OH OH	CI N OH N OH	НО
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.597	0.666	0.668
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 8

Model Applicability

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

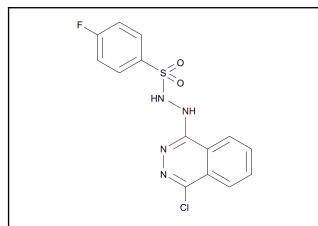
Res.) Sept. 1997

Res.) Sept. 1997

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

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
SCFP_8	1631785938	CI NH OH	0.574	4 out of 5		

SCFP_8	345243876	CI NH CI	0.383	1 out of 1
		[*]=C1[*]CCC(=O)N1		
SCFP_8	-1072897324	CI NH	0.383	1 out of 1
	Top Fea	tures for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-52074512	CI NH	-0.707	2 out of 14
SCFP_8	-601571304	[*][c](:[*]):[c](CI): [cH]:[*]	-0.707	2 out of 14
SCFP_8	-1378360678	CI NH	-0.58	2 out of 12



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248
Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.522
Enrichment: 1.26
Bayesian Score: -2.59
Mahalanobis Distance: 14.7

Mahalanobis Distance p-value: 1.95e-005

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

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

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

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

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

Name	Doxefazepam	Bicalutamide	Torsemide
Structure	OH OH	HO HO HN MAN	HN ONH NH
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.615	0.635	0.669
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

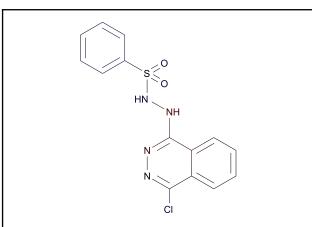
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

i cature co						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
SCFP_8	2144224020	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](F):[cH	0.383	1 out of 1		

SCFP_8	-705181820	[*]N[c]1:n:n:[c]([*]) :[*]:[c]:1:[*]	0.383	1 out of 1
SCFP_8	-1038366601	[*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[c]:1:2	0.383	1 out of 1
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	-1247518081	[]NS(=0)(=0)(c)1:(cH 1:(cH):(1:(cH);(cH)	-0.737	0 out of 3
SCFP_8	-52074512	[*]:[c](:[*])CI	-0.707	2 out of 14
SCFP_8	-730654023	[*][c](:[*]):[c](F):[cH]:[*]	-0.463	1 out of 6



C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.522
Enrichment: 1.26
Bayesian Score: -2.6
Mahalanobis Distance: 15.3

Mahalanobis Distance p-value: 5.78e-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 C	Compounds
----------------------	-----------

	-	_	_
Name	Doxefazepam	Torsemide	Omeprazole
Structure	OH OH	HN O NH NH NH NH	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	redicted Endpoint Single-Carcinogen		Single-Carcinogen
Distance	0.636	0.666	0.688
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

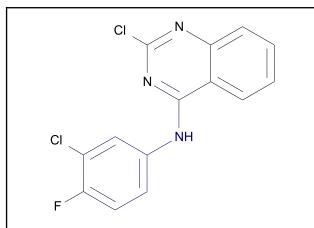
Model Applicability

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

Feature	Contribution
----------------	--------------

i cature co	iiiiibatioii					
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set		
SCFP_8	2097816882	[*]N[c]1:n:[*] [*] [c]([*]):[c]2:[cH]:[cH]:[c H]:[cH]:[c]:1:2	0.383	1 out of 1		

SCFP_8	-705181820	[*]N[c]1:n:n:[c]([*]) :[*]:[c]:1:[*]	0.383	1 out of 1
SCFP_8	-1038366601	[*]N[c]1:n:[*]*[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.383	1 out of 1
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple- Carcinogen in training set
SCFP_8	1653911926	[*][c]1:[cH]:[cH]:[cH]:1	-0.985	1 out of 12
SCFP_8	-1247518081		-0.737	0 out of 3
SCFP_8	-52074512	[*]:[c](:[*])CI	-0.707	2 out of 14



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.916
Enrichment: 0.995
Bayesian Score: -2.55
Mahalanobis Distance: 7.24

Mahalanobis Distance p-value: 0.989

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

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

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

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

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

Structural Similar Compounds					
Name	Aniline, 2,4-bis(o- methylphenoxy)-	ne, 2,4-bis(o- ylphenoxy)- Anthraquinone, 1-(2,4,6- trimethylphenylamino)-			
Structure	H ₂ N ^M	H H W N N N N N N N N N N N N N N N N N	HN N N CI		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant		
Distance	0.566	0.568	0.584		
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 Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,242,1	34ZIAG* -,235,69		

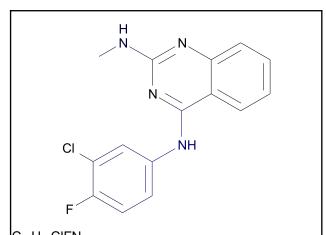
Model Applicability

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	-1185376954	CI NH CI NH F[c]1:[cH]:[cH]:[*]:[cH]:[c]:1CI	0.0756	6 out of 6
FCFP_12	-2069374281	CI NH [*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	0.0703	4 out of 4
FCFP_12	-2132756387	CI NH NH [*][c]1:[cH]:[cH]:[c] (F):[c](CI):[cH]:1	0.0703	4 out of 4

Top Features for negative contribution					
Bit/Smiles	Feature Structure	Score	Irritant in training set		
839741273	[*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[cH]:1	-0.708	4 out of 10		
-773983804	CI NH CI NH [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.444	46 out of 79		
	839741273	839741273 [*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	839741273		



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.547
Enrichment: 0.594
Bayesian Score: -4.42
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)-	Phenol, 2,2'- methylenebis(4-chloro-	Aniline, 2,4-bis(o- methylphenoxy)-		
Structure	HN N N CI	CIMOH	H ₂ N ^M		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant		
Distance	0.639	0.676	0.707		
Reference 34ZIAG* -,235,69		85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,533,1986	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	Feature Contribution							
	Top features for positive contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set				
FCFP_12	-1185376954	F[c]1:[cH]:[cH]:[*]:[cH]:[c]:1CI	0.0756	6 out of 6				

FCFP_12	-2132756387	CI NH	0.0703	4 out of 4
FCFP_12	-2050771350	[*][c]1:[cH]:[c] (F):[c](CI):[cH]:1	0.0583	2 out of 2
FGFF_12	-2030771330	H NH	0.0363	2 out or 2
		[*]N[c]1:[cH]:[cH]:[c](F):[c](CI):[cH]:1		
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	NH KN	-0.708	4 out of 10
		[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1		
FCFP_12	1294255210	H	-0.486	12 out of 22
		CI NH		
FCFP_12	136686699	[*]:[c](:[*])NC	-0.484	3 out of 6
		[*]NC		

C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4 Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.732
Enrichment: 0.794
Bayesian Score: -3.82
Mahalanobis Distance: 8.93

Mahalanobis Distance p-value: 0.456

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

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

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

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

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

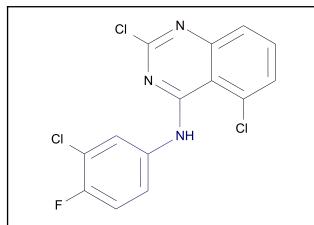
Structural Similar Compounds				
Name	s-Triazine, 2,4-dichloro-6- (o-chloroanilino)-	Sulfide, bis(4-t-butyl-m-cresyl)-	Aniline, 2,4-bis(o- methylphenoxy)-	
Structure	HN N N CI	OH OH	H ₂ N ^M	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant	
Distance	0.674	0.689	0.699	
Reference	34ZIAG* -,235,69	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	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 fea	tures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1185376954	F[c]1:[cH]:[cH]:[*]:[cH]:[c]:1CI	0.0756	6 out of 6

FCFP_12	-2132756387	[*][c]1:[cH]:[cH]:[c] (F):[c](CI):[cH]:1	0.0703	4 out of 4
FCFP_12	-1716224640	[*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.0583	2 out of 2
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	-0.708	4 out of 10
FCFP_12	1294255210	H N N N N N N N N N N N N N N N N N N N	-0.486	12 out of 22
FCFP_12	-773983804	[*]N[c]1:[cH]:[t]:[c] ([*]):[cH]:[cH]:1	-0.444	46 out of 79



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.926 Enrichment: 1.01 Bayesian Score: -2.4 Mahalanobis Distance: 7.4

Mahalanobis Distance p-value: 0.981

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	Aniline, 2,4-bis(o- methylphenoxy)-	s-Triazine, 2,4-dichloro-6- (o-chloroanilino)-
Structure	The state of the s	H ₂ N ^N r	HN N CI
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.513	0.610	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		85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986	34ZIAG* -,235,69

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 c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1185376954	CI NH CI NH CI CI CH]:[cH]:[cH]:[cH]:[c]:1CI	0.0756	6 out of 6
FCFP_12	-2069374281	CI NH CI NH CI [*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	0.0703	4 out of 4
FCFP_12	-1374842118	CI NH CI NH CI [*]:n:[c](CI):n:[*]	0.0703	4 out of 4

Top Features for negative contribution				
Bit/Smiles	Feature Structure	Score	Irritant in training set	
839741273	[*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.708	4 out of 10	
-773983804	[*]N[c]1:[cH]:[cH]:1	-0.444	46 out of 79	
	Bit/Smiles 839741273	Bit/Smiles Feature Structure 839741273 [*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	Bit/Smiles Feature Structure Score 839741273 C	

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.953 Enrichment: 1.03 Bayesian Score: -1.8

Mahalanobis Distance: 8.93

Mahalanobis Distance p-value: 0.459

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

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

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

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

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

Structural Sim	1-Amino-2-bromo-4-hydroxyanthraquinone	1-Amino-4-hydroxy-5- chloroanthraquinone	5-Norbornene-2,3- dicarboxylic acid, 1,4,5,6,7,7-hexachloro-
Structure	HO MNH 2	HO 11 NH ₂	OHC CI CI OH
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.590	0.625	0.655
Reference	28ZPAK -,83,72	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986

Model Applicability

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

	Top fe	atures for positive of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-2069374281	CI NH CI NH CI	0.0703	4 out of 4
FCFP_12	-1374842118	CI NH	0.0703	4 out of 4
FCFP_12	-1716224640	CI N	0.0583	2 out of 2

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294255210	CI N NH CI N	-0.486	12 out of 22
FCFP_12	566058135	CI N NH CI NH CI (*]CC(=O)N[*]	-0.367	13 out of 21

C₉H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4 Donors: 3

Model Prediction

Prediction: Non-Irritant

Probability: 0.958
Enrichment: 1.04
Bayesian Score: -1.63

Mahalanobis Distance: 6.44 Mahalanobis Distance p-value: 1

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

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

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

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

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

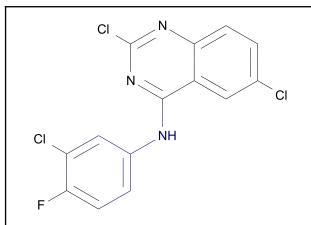
Name	8-Methylamino-4-hydroxy- 2-naphthalene sulfonic acid	Anthraquinone, 1,2,4- trihydroxy-	Phenol, 4,4'-sulfonyldi-
Structure	HO WN H	O OH WOH	HO O O O O O O O O O O O O O O O O O O
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.584	0.615	0.681
Reference	28ZPAK -,190,72	28ZPAK -,103,72	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 Co	ntribution				
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	-2069374281	CI HN NH ONH ₂ [*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	0.0703	4 out of 4	

FCFP_12	-1374842118	CI NH CI NH CI NH 2 [*]:n:[c](CI):n:[*]	0.0703	4 out of 4
FCFP_12	1499521844	CI NH	0.0658	3 out of 3
	Top Feat	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	CI NH	-0.65	0 out of 1
FCFP_12	367998008	CI NH2 CI NH2 (F):[cH]:[*]	-0.129	61 out of 76
FCFP_12	71476542	CI NH CI NH O NH 2 [*]:[c](:[*])CI	-0.12	64 out of 79



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.931 Enrichment: 1.01 Bayesian Score: -2.32 Mahalanobis Distance: 7.4

Mahalanobis Distance p-value: 0.981

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	Aniline, 2,4-bis(o- methylphenoxy)-	s-Triazine, 2,4-dichloro-6- (o-chloroanilino)-
Structure	The state of the s	H ₂ N ^N	HN N CI
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.518	0.615	0.664
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	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986	34ZIAG* -,235,69

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	-1185376954	CI NH CI	0.0756	6 out of 6
FCFP_12	-2069374281	[*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	-1374842118	CI NH CI NH [*]:n:[c](CI):n:[*]	0.0703	4 out of 4

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	[*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	-0.708	4 out of 10
FCFP_12	-773983804	CI NH	-0.444	46 out of 79

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.955
Enrichment: 1.04
Bayesian Score: -1.72

Mahalanobis Distance: 8.93

Mahalanobis Distance p-value: 0.459

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

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

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

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

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

Structural Sim	1-Amino-2-bromo-4-hydroxyanthraquinone	1-Amino-4-hydroxy-5- chloroanthraquinone	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-
Structure	HO MANH 2	HO 11 NH ₂	OH CI CI
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.596	0.631	0.655
Reference	28ZPAK -,83,72	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986

Model Applicability

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

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

		[*]:n:[c](Cl):n:[*]		
FCFP_12	-2069374281	CI NH	0.0703	4 out of 4
FCFP_12	-922480536	CI NH	0.0583	2 out of 2

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294255210	CI NH CI NH (CI NH C) (CI	-0.486	12 out of 22
FCFP_12	566058135	CI NH CI NH (CI NH CI NH	-0.367	13 out of 21

C₀H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Irritant

Probability: 0.955
Enrichment: 1.04
Bayesian Score: -1.71
Mahalanobis Distance: 6.44

Mahalanobis Distance p-value: 1

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

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

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

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

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

Name	8-Methylamino-4-hydroxy- 2-naphthalene sulfonic acid	Anthraquinone, 1,2,4- trihydroxy-	1-Amino-4-hydroxy-5- chloroanthraquinone
Structure	HO WN H	O OH OH OH	HO 11 NH ₂
Actual Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.577	0.608	0.680
Reference	28ZPAK -,190,72	28ZPAK -,103,72	28ZPAK -,83,72

Model Applicability

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

Feature Co	eature Contribution				
	Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	-1374842118	CI N NH CI HN NH CI (I) NH 2 (I) N	0.0703	4 out of 4	

FCFP_12	-2069374281	[*][c]1:[*]:[c](:[*]) :n:[c](CI):n:1	0.0703	4 out of 4
FCFP_12	1499521844	CI NH CI NH CI NH ₂ [*]NC(=O)N	0.0658	3 out of 3
		tures for negative of	-	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	CI NH CI HN NH CI HN NH CI (I 1 1 1 1 1 1 1 1 1	-0.65	0 out of 1
FCFP_12	367998008	[*][c](:[*]):[c](F):[cH]:[*]	-0.129	61 out of 76
FCFP_12	71476542	CI NH CI HN NH CI NH ₂ [*]:[c](:[*])CI	-0.12	64 out of 79

 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.927 Enrichment: 1.01 Bayesian Score: -2.38

Mahalanobis Distance: 6.03

Mahalanobis Distance p-value: 1
Prediction: Positive if the Bayesian score is above the estimated

best cutoff value from minimizing the false positive and false negative rate.

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

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

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

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

Structural Simila	ar Compounds		
Name	5-Norbornene-2,3- dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	Anthraquinone, 1,1'- iminodi-	1-Amino-2-bromo-4- hydroxyanthraquinone
Structure	OHD CI CI OH	H H H H H H H H H H H H H H H H H H H	HO MINH 2
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.722	0.734	0.755
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986	28ZPAK -,83,72

Model Applicability

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

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

FCFP_12	-2069374281	[*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	-1374842118	CI NH CI NH CI S:0 S:0 S:0 [*]:n:[c](CI):n:[*]	0.0703	4 out of 4
FCFP_12	-1716224640	[*][e]1:[*]:[c]([*]): [e]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.0583	2 out of 2
	Top Feat	ures for negative	contribution	1

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583		-0.65	0 out of 1
		[*]NN[c](:[*]):[*]		
FCFP_12	1789442672	CI N CI N CI ["]NS(=0)(=0)[c]1:[cH]:[cH]:[cH]:[cH]:1	-0.347	1 out of 2

C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.942 Enrichment: 1.02 Bayesian Score: -2.09

Mahalanobis Distance: 5.98

Mahalanobis Distance p-value: 1

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

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

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

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

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

Structural Sim	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-2-bromo-4- hydroxyanthraquinone	Anthraquinone, 1,1'- iminodi-
Structure	OHCI CI CI OH	HO 1 NH 2	o H
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.704	0.743	0.744
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986

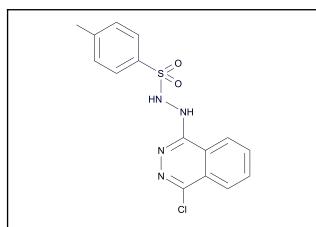
Model Applicability

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

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

FCFP_12	-1374842118	CI NH CI HN NH CI S; 0 S; 0 [*]:n:[c](CI):n:[*]	0.0703	4 out of 4
FCFP_12	-149636017	CI N N CI N N CI (H):(e[*))(c]1: [cH]:(cH):[c](F):[cH]:(cH):1	0.0703	4 out of 4
FCFP_12	-2069374281	[*][c]1:[*]:[c](:[*]) :n:[c](CI):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	1294344583	CI NH CI HN NH CI S:0 S:0 [*]NN[0](:[*]):[*]	-0.65	0 out of 1
FCFP_12	-1096219292	[']NS(=O)(=O)[c](:[']):[']	-0.229	26 out of 36



C₁₅H₁₃CIN₄O₂S

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.916 Enrichment: 0.995 Bayesian Score: -2.55 Mahalanobis Distance: 6

Mahalanobis Distance p-value: 1

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

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

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

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

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

Structural Simi	lar Compounds		
Name	1-Amino-2-bromo-4- hydroxyanthraquinone	5-Norbornene-2,3- dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-4-hydroxy-5- chloroanthraquinone
Structure	HO 1NH 2	OHCI CI CI OH	HO the NH 2
Actual Endpoint	Non-Irritant	Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.687	0.688	0.732
Reference	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72

Model Applicability

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

Feature Co			4 11 41	
	l op te	atures for positive of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	4427049		0.0734	5 out of 5
		s		
		i H H		
		N N		
		c: [*][c](:[*]):n:n:[*]		
	Ton Featur	es for negative of	ontribution	
Fingerprint		Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	Y	-0.65	0 out of 1
		s o o		
		H H		
		[*]NN[c](:[*]):[*]		
FCFP_12	1789442672	$\uparrow \uparrow \uparrow$	-0.347	1 out of 2
		N N N		
		N. T.		
		CI		
		[*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1		
FCFP_12	-1096219292	Y	-0.229	26 out of 36
		CI		
		[*]NS(=O)(=O)[c](:[*]):[*]		
		l		

C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248
Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.922 Enrichment: 1

Bayesian Score: -2.47

Mahalanobis Distance: 5.97

Mahalanobis Distance p-value: 1

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

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

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

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

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

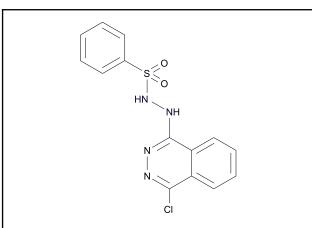
Structural Similar Compounds					
Name	1-Amino-2-bromo-4- hydroxyanthraquinone	5-Norbornene-2,3- dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-4-hydroxy-5- chloroanthraquinone		
Structure	HO MINH 2	OHCI CI CI OH	HO m		
Actual Endpoint	Non-Irritant	Irritant	Non-Irritant		
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant		
Distance	0.674	0.679	0.717		
Reference	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72		

Model Applicability

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

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

FCFP_12	4427049	[*][c](:[*]):n:n:[*]	0.0734	5 out of 5
FCFP_12	-149636017	[*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](F):[cH]:[cH]:1	0.0703	4 out of 4
		tures for negative		
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	[*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	-1096219292	[*]NS(=0)(=0)[c](:[*]	-0.229	26 out of 36
FCFP_12	367998008	[*][c](:[*]):[c](F):[cH]:[*]	-0.129	61 out of 76



C₁₄H₁₁CIN₄O₂S

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.94 Enrichment: 1.02 Bayesian Score: -2.14 Mahalanobis Distance: 6

Mahalanobis Distance p-value: 1

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

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

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

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

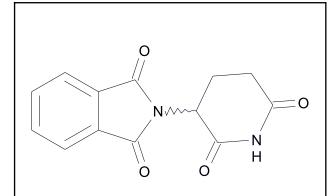
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly nonnormal X properties (e.g., fingerprints), the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

Structural Similar Compounds Name 1-Amino-2-bromo-4- 5-Norbornene-2,3- Phenol, 4,4'-sulfonyldi-					
	hydroxyanthraquinone	dicarboxylic acid, 1,4,5,6,7,7-hexachloro-			
Structure	HO MNH 2	OH CI CI OH	HO OI OI OH		
Actual Endpoint	Non-Irritant	Irritant	Irritant		
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant		
Distance	0.655	0.693	0.694		
Reference	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	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.

FCFP_12	4427049		0.0734	5 out of 5
FCFP_12	-1089199451	[*][c](:[*]):n:n:[*]	0.0658	3 out of 3
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	[*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	-1096219292	[*]NS(=0)(=0)[G](:[*]	-0.229	26 out of 36
FCFP_12	71476542	[*]:[c](:[*])CI	-0.12	64 out of 79



 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.964
Enrichment: 1.05
Bayesian Score: -1.37
Mahalanobis Distance: 11.6

Mahalanobis Distance p-value: 0.000214

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

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

Enrichment: An 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.

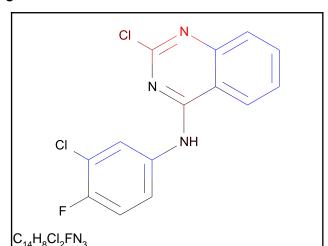
Structural Similar Compounds					
Name	1-Methylsulfonyl-2- pipecoline	1-(ethylsulfonyl)-4- pipecoline	p-Acetophenetidide, 3'- nitro-		
Structure	N S O O OH	OH ON SOO	O Number of the second of the		
Actual Endpoint	Non-Irritant	Non-Irritant	Non-Irritant		
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant		
Distance	0.591	0.592	0.627		
Reference	US ARMY	US ARMY	28ZPAK -,115,72		

Model Applicability

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

Feature Contribution						
	Top fea	atures for positive o	ontribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	2139882011	[*]N1C(=[*])[c]2:[cH] :[*]:[cH]:[cH]:[c]:2 C1=O	0.081	11 out of 11		

FCFP_12	1393189956	[*]N1C(=[*])[c]2:[cH]	0.081	11 out of 11
FCFP_12	-922480536	:[cH]:[cH]:[c]: 2C1=O	0.0583	2 out of 2
		[*]=C1[*]CCC(=O)N1		
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	566058135	N N N H	-0.367	13 out of 21
		[*]CC(=O)N[*]		
FCFP_12	392594677		-0.207	8 out of 11
		[*]C([*])N1C(=[*])[c] (:[*]):[c](:[cH]:[*])C1=O		
FCFP_12	1618154665		-0.0845	412 out of 490
		[*]:[cH]:[cH]:[eH]:[*]		



Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 10.9

Unit: mg/kg_body_weight/day Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.0088

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds					
Name	148	p-Nitrosodiphenylamine	2,3,7,8- Tetrachlorodibenzo-p- dioxin		
Structure	N NH ₂	N N N N N N N N N N N N N N N N N N N	CI		
Actual Endpoint (-log C)	3.7362	2.76567	9.31469		
Predicted Endpoint (-log C)	3.75282	3.37945	4.75869		
Distance	0.685	0.701	0.704		
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: 1126732433: [*]:n:[c](Cl):n:[*]
- 3. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

Feature Cont	Feature Contribution					
	Top features	for positive contribution	n			
Fingerprint	Bit/Smiles	Feature Structure	Score			
ECFP_6	655739385	CI NH CI NH [*]:n:[*]	0.229			
		[]-11.[]				

ECFP_6	834876373	CI NH [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	CI NH CI NH [*]CI	0.129
Fingerprint	Top Features	for negative contributio Feature Structure	n Score
ECFP_6	1996767644	[*]:[cH]:[c](:[*]):[*]	-0.251
ECFP_6	642810091	CI NH [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	CI NH F NH [*]:[cH]:[*]	-0.232

TOPKAT_Carcinogenic_Potency_TD50_Mouse

C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4 Donors: 2

Model Prediction

Prediction: 8.65

Unit: mg/kg_body_weight/day Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.00191

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	646	44			
Structure	НО	OH NN N O I N N O I N N O I N N O I N N O I N O	NH NH			
Actual Endpoint (-log C)	2.43468	0.937339	2.42163			
Predicted Endpoint (-log C)	3.66084	3.26294	2.85113			
Distance	0.646	0.696	0.698			
Reference	CPDB	CPDB	CPDB			

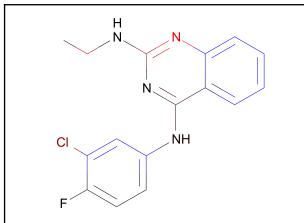
Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

Feature Contribution Top features for positive contribution						
ECFP_6	655739385	CI NH NH [*]:n:[*]	0.229			

E 0 E D . 0	bo 10=00=0		0.400
ECFP_6	834876373	T N	0.163
		CIŅH	
		[*][c](:[*]):n:[c](:[*]):[*]	
ECFP_6	-817402818	H N	0.129
		CINH	
		[*]CI	
	Top Features	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	H	-0.251
		CINNH	
		[*]:[cH]:[cH]:[c](:[*]):[*]	
ECFP_6	642810091	H	-0.247
		CINH	
		[*]:[c](:[*]):[*]	
ECFP_6	-182236392	H N	-0.232
		CINH	
		[*]:[cH]:[*]	



C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4

Acceptors: 4
Donors: 2

Model Prediction

Prediction: 5.92

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11

Mahalanobis Distance p-value: 0.00123

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Phenolphthalein	646	Chlorobenzilate
Structure	но	H OH	CI

Actual Endpoint (-log C)	2.43468	0.937339	3.53947
Predicted Endpoint (-log C)	3.66084	3.26294	3.34564
Distance	0.672	0.700	0.704
Reference	CPDB	CPDB	CPDB

Model Applicability

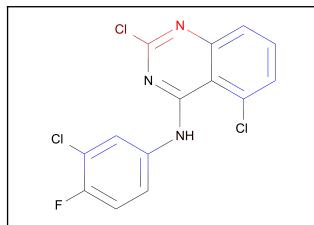
Structural Similar Compounds

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	CI NH [*]:n:[*]	0.229	

ECFP_6	1559650422	CI NH	0.203
ECFP_6	834876373	[*]C[*] CI	0.163
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*]:[cH]:[cH]:[c](:[*]):[*]	-0.251
ECFP_6	642810091	C1 NH [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	CI NH [*]:[cH]:[*]	-0.232



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 7.12

Unit: mg/kg_body_weight/day Mahalanobis Distance: 9.93

Mahalanobis Distance p-value: 0.0434

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	2,3,7,8- Tetrachlorodibenzo-p- dioxin	Dicofol	Nitrofen	
Structure	CI	HO CI CI CI CI	CI	
Actual Endpoint (-log C)	9.31469	4.05158	3.39277	
Predicted Endpoint (-log C)	4.75869	3.80707	3.41579	
Distance	0.702	0.715	0.735	
Reference	CPDB	CPDB	CPDB	

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
- 3. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

Feature Contribution				
	Top features	for positive contribution	1	
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	CI NH CI NH CI [*]:n:[*]	0.229	

ECFP_6	834876373	CI NH CI [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	CI NH CI [*]CI	0.129
Fig. 20 mg wind		for negative contributio	
Fingerprint	Bit/Smiles 1996767644	Feature Structure	Score
ECFP_6		[*]:[cH]:[cH]:[c](:[*]):[*]	-0.251
ECFP_6	642810091	CI NH CI [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	CI NH CI CI [*]:[cH]:[*]	-0.232

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49 Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 20.5

Unit: mg/kg_body_weight/day Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 5.83e-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.

Name	693	Oxazepam	Chrysazin
Structure	AND Enantiomer H N N CI	OH ON N HN	но
Actual Endpoint (-log C)	3.90356	3.90356	3.0774
Predicted Endpoint (-log C)	3.39677	3.39677	3.07832
Distance	0.610	0.610	0.615
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: 1126732433: [*]:n:[c](CI):n:[*]
- 3. Unknown ECFP_2 feature: -1237219435: [*]C([*])N[c](:[*]):[*]
- 4. Unknown ECFP_2 feature: -2097159651: [*]CC(N[*])C(=[*])[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	CI N NH CI N	0.229	

ECFP_6	834876373	[*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	CI N N CI N N N N N N N N N N N N N N N	0.129
Fingerprint	Top Features f	or negative contributio Feature Structure	on Score
ECFP_6	2106656448	CI N NH CI N	-0.275
ECFP_6	1996767644	CI NH CI NH CI	-0.251
ECFP_6	642810091	CI N N CI N N CI N N N CI N N N CI N N N N	-0.247

CI N CI NH HN NH 2

C₉H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: 10.1

Unit: mg/kg_body_weight/day Mahalanobis Distance: 12.6

Mahalanobis Distance p-value: 1.32e-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.

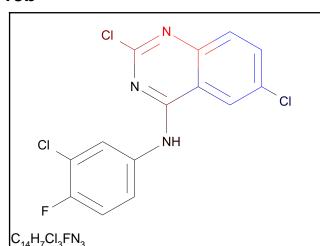
Structural Similar Compounds			
Name	3,4,4´-Triaminodiphenyl ether	2-Hydrazino-4-(p- aminophenyl) thi-azole	215
Structure	H ₂ N NH ₂	NH ₂ N NH ₂ N NH ₂ H	H ₂ N ₄ N ₁ N ₂ N ₁ N ₂ N ₃ CI
Actual Endpoint (-log C)	4.20909	4.26135	3.477
Predicted Endpoint (-log C)	3.19224	4.32504	3.97642
Distance	0.589	0.596	0.618
Reference	CPDB	CPDB	CPDB

Model Applicability

- 1. OPS PC18 out of range. Value: -3.6042. Training min, max, SD, explained variance: -3.5308, 6.5936, 1.215, 0.0184.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

Feature Contribution				
	Top features	for positive contribution	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	CI NH CI NH CI NH 2 [*]:n:[*]	0.229	

ECFP_6	1572579716	CI NH CI NH CI NH 2 [*]N	0.225
ECFP_6	834876373	CI HN NH ONH ₂ [*][c](:[*]):n:[c](:[*]):[*]	0.163
		or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	CI NH CI NH ONH [*]:[cH]:[cH]:[c](:[*]):[*]	-0.251
ECFP_6	642810091	CI NH CI NH O NH 2 [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	CI NH CI NH O NH 2 [*]:[cH]:[*]	-0.232



Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 4.57

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11.3

Mahalanobis Distance p-value: 0.00044

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

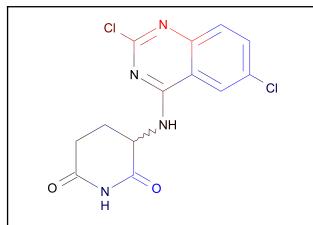
Structural Similar Compounds				
Name	Dicofol	2,3,7,8- Tetrachlorodibenzo-p- dioxin	Nitrofen	
Structure	HO CI	CI	CI	
Actual Endpoint (-log C)	4.05158	9.31469	3.39277	
Predicted Endpoint (-log C)	3.80707	4.75869	3.41579	
Distance	0.700	0.702	0.716	
Reference	CPDB	CPDB	CPDB	

Model Applicability

- 1. OPS PC16 out of range. Value: 4.2781. Training min, max, SD, explained variance: -3.1026, 4.016, 1.245, 0.0193.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
- 3. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

Feature Cont		for positive contribution	n
Fingerprint	Bit/Smiles	for positive contributio Feature Structure	Score
ECFP_6	655739385	CI NH CI (*):n:[*]	0.229

ECFP_6	834876373	CI NH CI NH [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	CI NH CI NH	0.129
Fingerprint	Top Features	for negative contribution Feature Structure	on Score
ECFP_6	1996767644	CI NH CI NH (CI NH):[c](:[*]):[*]	-0.251
ECFP_6	642810091	CI NH CI NH (*):[*]	-0.247
ECFP_6	-182236392	CI NH CI NH (*):[cH]:[*]	-0.232



 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 21.6

Unit: mg/kg_body_weight/day Mahalanobis Distance: 13.6

Mahalanobis Distance p-value: 3.06e-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 Similar	Compounds
Nama	603

Name	693	Oxazepam	Chrysazin
Structure	HO N	OH ON N N N CI	но
Actual Endpoint (-log C)	3.90356	3.90356	3.0774
Predicted Endpoint (-log C)	3.39677	3.39677	3.07832
Distance	0.592	0.592	0.628
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 PC18 out of range. Value: -3.6033. Training min, max, SD, explained variance: -3.5308, 6.5936, 1.215, 0.0184.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
- 3. Unknown ECFP_2 feature: -1237219435: [*]C([*])N[c](:[*]):[*]
- 4. Unknown ECFP_2 feature: -2097159651: [*]CC(N[*])C(=[*])[*]

Feature Contribution

ECFP_6	834876373	CI NH CI	0.163
ECFP_6	-817402818	[*][c](:[*]):n:[c](:[*]):[*]	0.129
LOIT_0	7017402010	CI NH	0.129
		o H	
		[*]Cl	
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	CI NH CI	-0.275
FOED 0	4000707044	н [*]C(=O)[*]	0.054
ECFP_6	1996767644	CI NH CI NH (CI NH):[cH]:[cH]:[cH]:[c](:[*	-0.251
ECFP_6	642810091	CI NH CI	-0.247
		н [*]:[c](:[*]):[*]	

C₉H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4
Donors: 3

Model Prediction

Prediction: 9.53

Unit: mg/kg_body_weight/day Mahalanobis Distance: 10.8

Mahalanobis Distance p-value: 0.00281

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	2-Hydrazino-4-(p- aminophenyl) thi-azole	3,4,4´-Triaminodiphenyl ether	215	
Structure	NH ₂ N-NH ₂ H	H ₂ N NH ₂	H ₂ N _{rt} N _{rt} N _r CI	
Actual Endpoint (-log C)	4.26135	4.20909	3.477	
Predicted Endpoint (-log C)	4.32504	3.19224	3.97642	
Distance	0.596	0.596	0.611	
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: 1126732433: [*]:n:[c](CI):n:[*]

Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score ECFP_6 655739385 CI NH2 (*]:n:[*]

ECFP_6	1572579716	CI NH CI NH CI NH 2 [*]N	0.225
ECFP_6	834876373	CI NH CI O NH ₂ [*][c](:[*]):n:[c](:[*]):[*]	0.163
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*]:[cH]:[cH]:[c](:[*	-0.251
ECFP_6	642810091	CI N NH CI NH CI NH 2 [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	CI N N CI N NH CI NH 2 [*]:[cH]:[*]	-0.232

 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4 Acceptors: 5

Donors: 2

Model Prediction

Prediction: 8.91

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11.3

Mahalanobis Distance p-value: 0.000538

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	646	[4-Chloro-6-(2,3-xylidino)- 2-pyri-midinylthio]acetic acid s	Phenolphthalein	
Structure	OH OH	HO O NH	НО	
Actual Endpoint (-log C)	0.937339	4.47685	2.43468	
Predicted Endpoint (-log C)	3.26294	3.8529	3.66084	
Distance	0.627	0.662	0.678	
Reference	CPDB	CPDB	CPDB	

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](CI):n:[*]
- 3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

-eature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score			
ECFP_6	655739385	CI NH CI NH CI S:00 S:00 [*]:n:[*]	0.229			

ECFP_6	834876373	[*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	CI NH CI HN CI S:0	0.129
Fig. 20 and 19 de		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*]:[cH]:[cH]:[c](:[*]):[*]	-0.251
ECFP_6	642810091	CI NH CI HN NH CI F:0 S:0 [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	CI NH CI HN CI S::0 [*]:[cH]:[*]	-0.232

C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 6.89

Unit: mg/kg_body_weight/day Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.0208

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds					
Name	646	[4-Chloro-6-(2,3-xylidino)- 2-pyri-midinylthio]acetic acid s	Phenolphthalein		
Structure	OH	HO NH NH	НО		
Actual Endpoint (-log C)	0.937339	4.47685	2.43468		
Predicted Endpoint (-log C)	3.26294	3.8529	3.66084		
Distance	0.656	0.676	0.677		
Reference	CPDB	CPDB	CPDB		

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](CI):n:[*]
- 3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

Feature Contribution						
Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score			
ECFP_6	655739385	CI NH CI NH CI S'.00 S'.00 [*]:n:[*]	0.229			

ECFP_6	834876373	[*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	CI NH CI NH CI S:0 S:0 S:0	0.129
Fig. are any sign.t		for negative contributio	
Fingerprint ECFP_6	Bit/Smiles 1996767644	Feature Structure	Score -0.251
2011_0	1336767644	[*]:[cH]:[c](:[*]	0.201
ECFP_6	642810091	CI NO NH CI NO S'O S'O S'O [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	CI NH CI HN NH CI S; O S; O [*]:[cH]:[*]	-0.232

 $C_{15}H_{13}CIN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 21.2

Unit: mg/kg_body_weight/day Mahalanobis Distance: 12.5

Mahalanobis Distance p-value: 2.15e-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	646	Phenolphthalein	C.I. pigment red 3
Structure	O - Z	НО	N N N N N N N N N N N N N N N N N N N
Actual Endpoint (-log C)	0.937339	2.43468	0.937339
Predicted Endpoint (-log C)	3.26294	3.66084	3.17837
Distance	0.618	0.659	0.660
Reference	CPDB	CPDB	CPDB

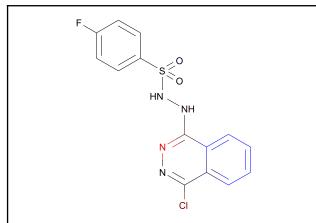
Model Applicability

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

- 1. OPS PC26 out of range. Value: -3.4447. Training min, max, SD, explained variance: -2.9667, 3.5042, 1.009, 0.0127.
- 2. Unknown ECFP_2 feature: 1050351974: [*]:n:[c](Cl):[c](:[*]):[*]
- 3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

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

-817402818		0.129
1333660716	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.0746
	Feature Structure	Score
1996767644	[*]:[cH]:[cH]:[c](:[*]):[*]	-0.251
642810091	[*]:[c](:[*]):[*]	-0.247
-182236392	[*]:[cH]:[*]	-0.232
	1333660716 Top Features 1 Bit/Smiles 1996767644 642810091	1333660716



C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248
Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 16.4

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11.3

Mahalanobis Distance p-value: 0.000448

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds						
Name	646	Phenolphthalein	C.I. pigment red 3			
Structure	OH OH	НО	O N N O			
Actual Endpoint (-log C)	0.937339	2.43468	0.937339			
Predicted Endpoint (-log C)	3.26294	3.66084	3.17837			
Distance	0.655	0.664	0.686			
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 PC26 out of range. Value: -3.2042. Training min, max, SD, explained variance: -2.9667, 3.5042, 1.009, 0.0127.
- Unknown ECFP_2 feature: 1050351974: [*]:n:[c](Cl):[c](:[*]):[*]
 Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	[*]:n:[*]	0.229	

ECFP_6	-817402818	F	0.129
ECFP_6	1333660716	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.0746
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*]:[cH]:[cH]:[c](:[*]):[*]	-0.251
ECFP_6	642810091	[*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	[*]:[cH]:[*]	-0.232

 $C_{14}H_{11}CIN_4O_2S$

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 17.2

Unit: mg/kg_body_weight/day Mahalanobis Distance: 12.3

Mahalanobis Distance p-value: 6.1e-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	646	Phenolphthalein	C.I. pigment red 3
Structure	O - Z	НО	N N N N N N N N N N N N N N N N N N N
Actual Endpoint (-log C)	0.937339	2.43468	0.937339
Predicted Endpoint (-log C)	3.26294	3.66084	3.17837
Distance	0.659	0.664	0.684
Reference	CPDB	CPDB	CPDB

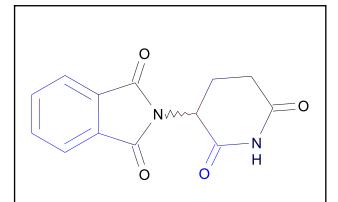
Model Applicability

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

- 1. OPS PC26 out of range. Value: -2.9692. Training min, max, SD, explained variance: -2.9667, 3.5042, 1.009, 0.0127.
- Unknown ECFP_2 feature: 1050351974: [*]:n:[c](Cl):[c](:[*]):[*]
 Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

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

ECFP_6	-817402818	[*]CI	0.129
ECFP_6	1333660716	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.0746
		or negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*]:[cH]:[cH]:[c](:[*]):[*]	-0.251
ECFP_6	642810091	[*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	[*]:[cH]:[*]	-0.232



 $C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002 Rotatable Bonds: 1 Acceptors: 4

Donors: 1

Model Prediction

Prediction: 367

Unit: mg/kg_body_weight/day Mahalanobis Distance: 9.93

Mahalanobis Distance p-value: 0.043

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

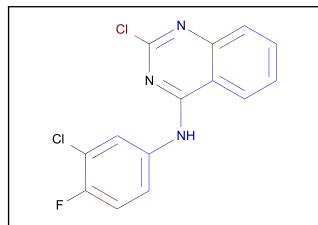
Structural Similar Compounds				
Name	581	Phenobarbital s	1-[(5- Nitrofurfuryli-dene)ami-no] hy-dantoin	
Structure	H N N Na	HN O	H N N N N N N N N N N N N N N N N N N N	
Actual Endpoint (-log C)	3.89128	4.49846	2.23074	
Predicted Endpoint (-log C)	3.31105	3.14828	3.18468	
Distance	0.590	0.638	0.660	
Reference	CPDB	CPDB	CPDB	

Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -957084426: [*]C([*])N1C(=[*])[*]:[*]C1=[*]

Top features to the street to	for positive contribution Feature Structure	Score	
	Feature Structure	Score	
S7460056			
37400030	[*]C([*])[*]	0.0596	
		N N N	N N N

ECFP_6	670515721	[*]N([*])[*]	0.00735
-		for negative contribution	
Fingerprint	Bit/Smiles		Score
ECFP_6	2106656448	[*]C(=O)[*]	-0.275
ECFP_6	1996767644	[*]:[cH]:[c](:[*]):[*]	-0.251
ECFP_6	642810091	[*]:[c](:[*]):[*]	-0.247



C₁₄H₈Cl₂FN₃

Molecular Weight: 308.13782

ALogP: 5.202 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 18.6

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11.7

Mahalanobis Distance p-value: 0.00147

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	45	357	4-(4-N-Methyl-N- nitrosamino-styryl)quinolin e	
Structure	OH N N	N N N N N N N N N N N N N N N N N N N	2-20	
Actual Endpoint (-log C)	3.92659	5.61692	5.61692	
Predicted Endpoint (-log C)	3.28325	5.49128	5.49128	
Distance	0.530	0.601	0.601	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

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

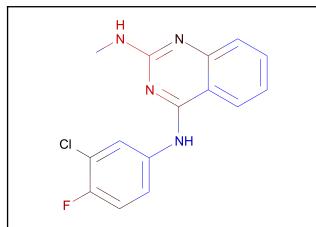
Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	32	CI NH NH [*]CI	0.154	

FCFP_6	307419094	CI NH CI NH [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121
FCFP_6		CI NH [*]N[*]	0.064
Fingerprint		egative contribution Feature Structure	Score
FCFP_6	991735244	CI NH CI NH [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	[*]:[cH]:[*]	-0.354
FCFP_6	590925877	CI NH NH [*]N[c](:[cH]:[*]):[c H]:[*]	-0.323

3.7508

0.630

CPDB



C₁₅H₁₂CIFN₄

Molecular Weight: 302.73398

ALogP: 4.677 Rotatable Bonds: 3

Acceptors: 4 Donors: 2

Model Prediction

Prediction: 12

Unit: mg/kg_body_weight/day Mahalanobis Distance: 13.9

Mahalanobis Distance p-value: 4.08e-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.

Otractarar Offinar Compounds				
Name	45	Indomethacin	Phenolphthalein	
Structure	OH N N N N N N N N N N N N N N N N N N N	OH OH	НО	
Actual Endpoint (-log C)	3.92659	5.49293	2.54766	

Model Applicability

Predicted Endpoint (-log

Distance

Reference

Structural Similar Compounds

3.28325

0.618

CPDB

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

4.9569

0.629

CPDB

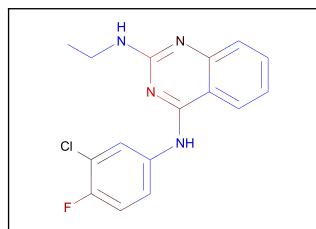
1. OPS PC18 out of range. Value: 4.9898. Training min, max, SD, explained variance: -4.1023, 4.8669, 1.305, 0.0164.

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	-1151914249	CI NH NH [*]N[c](:n:[*]):n:[*]	0.204	

FCFP_6	32 307419094	I NH NH I NH	0.154
	Top Features	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	CI NH NH [*]:[cH]:[*]	-0.354
FCFP_6	590925877	[*]N[c](:[cH]:[*]):[c H]:[*]	-0.323

0.650

CPDB



C₁₆H₁₄CIFN₄

Molecular Weight: 316.76056

ALogP: 5.025 Rotatable Bonds: 4

Acceptors: 4
Donors: 2

Model Prediction

Prediction: 35.7

Unit: mg/kg_body_weight/day Mahalanobis Distance: 14.1

Mahalanobis Distance p-value: 9.65e-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.

Name	Indomethacin	45	Phenolphthalein
Structure	OH OH	OH N N	НО
Actual Endpoint (-log C)	5.49293	3.92659	2.54766
Predicted Endpoint (-log	4.9569	3.28325	3.7508

Model Applicability

Distance

Reference

Structural Similar Compounds

0.623

CPDB

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

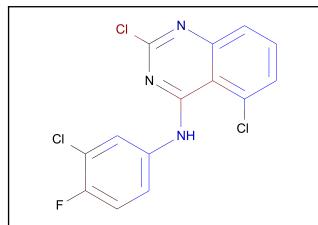
0.644

CPDB

1. OPS PC18 out of range. Value: 5.121. Training min, max, SD, explained variance: -4.1023, 4.8669, 1.305, 0.0164.

Feature Contribution Top features for positive contribution				
FCFP_6	-1151914249	H	0.204	

FCFP_6	307419094	H NH NH I	0.154
	Ton Footures f	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	
Fingerprint	Bit/Smiles	or negative contribution Feature Structure	Score
FCFP_6	-1272709286	CI NH [*]NCC	-0.526
FCFP_6	991735244	C: NH [*]:[e]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	CI NH [*]:[cH]:[*]	-0.354



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 16

Unit: mg/kg_body_weight/day Mahalanobis Distance: 14

Mahalanobis Distance p-value: 1.83e-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	45	4-(4-N-Methyl-N- nitrosamino-styryl)quinoli ne	357	
Structure	OH		N N N N N N N N N N N N N N N N N N N	
Actual Endpoint (-log C)	3.92659	5.61692	5.61692	
Predicted Endpoint (-log C)	3.28325	5.49128	5.49128	
Distance	0.596	0.651	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 Contribution				
	Top features	for positive contribution	า	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	32	CI NH CI E T CI CI CI CI CI CI CI	0.154	

FCFP_6	307419094	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121
FCFP_6	3	CI NH CI [*]N[*]	0.064
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	CI NH CI (*):[cH]:[*]	-0.354
FCFP_6	590925877	[*]N[c](:[cH]:[*]):[c H]:[*]	-0.323

F

┖

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.679

Unit: mg/kg_body_weight/day Mahalanobis Distance: 15.3

Mahalanobis Distance p-value: 4.62e-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	44	Chrysazin	
Structure	-O-N H N H N O	OH NH	но	
Actual Endpoint (-log C)	5.25509	2.85045	2.99143	
Predicted Endpoint (-log C)	3.89291	2.7768	3.29868	
Distance	0.593	0.616	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. OPS PC18 out of range. Value: 4.9912. Training min, max, SD, explained variance: -4.1023, 4.8669, 1.305, 0.0164.

Feature Contribution Top features for positive contribution				
-1043250487	[,]CC(N[,])C(=[,])[,] O H O H CI N CI	1.15		
	Top features Bit/Smiles	Bit/Smiles Feature Structure -1043250487	Top features for positive contribution Bit/Smiles Feature Structure Score -1043250487 1.15	

FCFP_6	1	CI N N CI N N CI N N N CI N N N N N N N	0.234
FCFP_6	-885550502	[']C(=['])NC(=['])[']	0.229
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	CI NH CI NH CI NH CI (CI CI C	-0.422
FCFP_6	16	CI NH CI NH CI F CI	-0.354
FCFP_6	566058135	CI NH CI NH CI (*) CC(=O)N[*]	-0.182

C₉H₇Cl₂N₅O

Molecular Weight: 272.09078

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4 Donors: 3

Model Prediction

Prediction: 11.2

Unit: mg/kg_body_weight/day Mahalanobis Distance: 13.3

Mahalanobis Distance p-value: 8.94e-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.

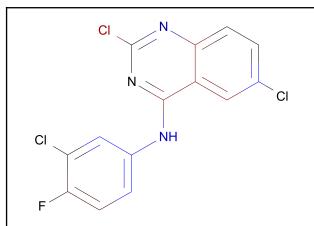
Name	215	2-Hydrazino-4-(p- aminophenyl) thi-azole	Dapsone
Structure	H ₂ N ₄ CI	NH ₂ N N-NH ₂ H	NH 2
Actual Endpoint (-log C)	2.84742	5.30159	4.04473
Predicted Endpoint (-log C)	3.32496	4.70335	4.05717
Distance	0.598	0.615	0.621
Reference	CPDB	CPDB	CPDB

Model Applicability

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

Feature Contribution					
	Top features	for positive contributio	n		
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_6	1	CI NH CI NH CI NH 2 [*]=O	0.234		

FCFP_6	307419094	NH CI NH CI NH 2 [*]CI	0.154
		CI NH2 *][c](:[*]):[c](:[cH :[*]):[c](:[*]):[*]	
Fingerprint		egative contribution Feature Structure	Score
FCFP_6	16	CI NH CI NH CI NH 2 [*]:[cH]:[*]	-0.354
FCFP_6	17	CI N CI HN NH CI (*):n:[*]	-0.149
FCFP_6	0	CI NO CI NO CI NO	-0.115



C₁₄H₇Cl₃FN₃

Molecular Weight: 342.58288

ALogP: 5.867 Rotatable Bonds: 2

Acceptors: 3
Donors: 1

Model Prediction

Prediction: 6.05

Unit: mg/kg_body_weight/day Mahalanobis Distance: 14.2

Mahalanobis Distance p-value: 4.74e-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 Similar Compounds				
Name	45	357	4-(4-N-Methyl-N- nitrosamino-styryl)quinolin e	
Structure	OH N N	N N N N N N N N N N N N N N N N N N N	2-20	
Actual Endpoint (-log C)	3.92659	5.61692	5.61692	
Predicted Endpoint (-log C)	3.28325	5.49128	5.49128	
Distance	0.603	0.664	0.664	
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	32	CI NH CI NH [*]CI	0.154		

FCFP_6	307419094	[*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121
FCFP_6	3	CI NH CI NH [*]N[*]	0.064
		for negative contribution	
Fingerprint FCFP_6	Bit/Smiles 16	Feature Structure	Score -0.354
		CI NH CI NH [*]:[cH]:[*]	
FCFP_6	590925877	CI NH CI ([cH]:[*]):[c H]:[*]	-0.323
FCFP_6	17	CI NH CI	-0.149

F

┖

 $C_{13}H_{10}CI_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.257

Unit: mg/kg_body_weight/day Mahalanobis Distance: 15.1

Mahalanobis Distance p-value: 2.07e-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	1,2-Dihydro-2-(5-nitro-2- thi-enyl) quinazolin-4(3H)- one	44	4,4´-Sulfonylbisacetanilide	
Structure	-O-N HN HN O	OH NH	HN NH	
Actual Endpoint (-log C)	5.25509	2.85045	3.77655	
Predicted Endpoint (-log C)	3.89291	2.7768	3.55337	
Distance	0.599	0.616	0.628	
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 PC18 out of range. Value: 4.9468. Training min, max, SD, explained variance: -4.1023, 4.8669, 1.305, 0.0164.

Feature Contribution						
	Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_6	-1043250487	[,]cc(N[,])c(=[,])[,] OH OH NH CI	1.15			

FCFP_6	-885550502	CI NH	0.234
		[*]C(=[*])NC(=[*])[*]	
		egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	CI NH CI NH (*):[cH]:[*]	-0.354
FCFP_6	566058135	CI NH CI NH (CI NH CI NH	-0.182
FCFP_6	17	CI NH CI NH (*):n:[*]	-0.149

C₉H₇Cl₂N₅O

Molecular Weight: 272.09077

ALogP: 2.379 Rotatable Bonds: 2

Acceptors: 4 Donors: 3

Model Prediction

Prediction: 29.6

Unit: mg/kg_body_weight/day Mahalanobis Distance: 13.6

Mahalanobis Distance p-value: 2.08e-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	215	Purpurin	2-Hydrazino-4-(p- aminophenyl) thi-azole		
Structure	H ₂ N ₄ , N ₁ CI	HO OH O OH O	NH ₂ N NH ₂ H		
Actual Endpoint (-log C)	2.84742	2.57737	5.30159		
Predicted Endpoint (-log C)	3.32496	3.49183	4.70335		
Distance	0.590	0.613	0.615		
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	1	CI NH CI NH CI NH 2 [*]=O	0.234		

FCFP_6	32	CI N NH CI NH 2 [*]CI	0.154
FCFP_6	307419094	CI NH CI NH CI NH ₂ [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.422
FCFP_6	16	CI NH CI NH CI NH 2 [*]:[cH]:[*]	-0.354
FCFP_6	17	CI NH CI NH CI NH CI NH 2 [*]:n:[*]	-0.149

3.4628

4.7324

0.609

CPDB

 $C_{15}H_{12}CI_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 10.2

Unit: mg/kg_body_weight/day Mahalanobis Distance: 14

Mahalanobis Distance p-value: 1.55e-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.

Name	Indomethacin	646	Omeprazole
Structure	OH	OH NN OI.	

2.41938

3.77987

0.603

CPDB

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.

5.49293

4.9569

0.568

CPDB

Feature Contribution				
Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	CI NH CI NH CI S: 0	0.234	

FCFP_6	32	CI N N N N N N N N N N N N N N N N N N N	0.154
FCFP_6	203677720	[*]S(=[*])(=[*])[c](: [cH]:[*]):[cH]:[*]	0.137
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.422
FCFP_6	16	CI NH CI NH CI S :: 0 S :: 0 [*]:[cH]:[*]	-0.354
FCFP_6	17	CI NH CI NH CI S::0 [*]:n:[*]	-0.149

C₁₄H₉Cl₂FN₄O₂S

Molecular Weight: 387.21626

ALogP: 4.162 Rotatable Bonds: 4

Acceptors: 5 Donors: 2

Model Prediction

Prediction: 12.2

Unit: mg/kg_body_weight/day Mahalanobis Distance: 13.5

Mahalanobis Distance p-value: 2.98e-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	Indomethacin	Omeprazole	646	
Structure	OH OH		H N N N N N N N N N N N N N N N N N N N	
Actual Endpoint (-log C)	5.49293	3.4628	2.41938	
Predicted Endpoint (-log C)	4.9569	4.7324	3.77987	
Distance	0.554	0.605	0.623	
Reference	CPDB	CPDB	CPDB	

Model Applicability

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

Feature Contribution				
	Top features for positive contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	CI NH CI NH CI S:0	0.234	

FCFP_6	32	CI NH	0.154
FCFP_6	203677720	[*]S(=[*])(=[*])(c](: [cH]:[*]):[cH]:[*]	0.137
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	CI NH NH NH NH NH NH NH NH NH NH NH NH NH	-0.422
FCFP_6	16	CI NH CI NH CI S: 0 S: 0 [*]:[cH]:[*]	-0.354
FCFP_6	17	CI NH CI HN :00 S:00 [*]:n:[*]	-0.149

 $C_{15}H_{13}CIN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 12.7

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.000885

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

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

Name	Omeprazole	Indomethacin	796
Structure	O N N N N N N N N N N N N N N N N N N N	OH OH	H ₂ N N N N N N N N N N N N N N N N N N N
Actual Endpoint (-log C)	3.4628	5.49293	2.71505
Predicted Endpoint (-log C)	4.7324	4.9569	4.45918
Distance	0.592	0.595	0.603
Reference	CPDB	CPDB	CPDB

Model Applicability

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

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

Feature Contribution

Top features for positive contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	(*) = O	0.234

FCFP_6	32	© Z T Z Z C [*]CI	0.154
FCFP_6	203677720	[*]S(=[*])(=[*])[c](: [cH]:[*]):[cH]:[*]	0.137
		for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.422
FCFP_6	16	[*]:[cH]:[*]	-0.354
FCFP_6	17	[*]:n:[*]	-0.149

C₁₄H₁₀CIFN₄O₂S

Molecular Weight: 352.7712

ALogP: 3.248 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 15.2

Unit: mg/kg_body_weight/day Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.0125

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the 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	Indomethacin	Omeprazole	796
Structure	OH OH	H N N	H ₂ N N N N N N N N N N N N N N N N N N N
Actual Endpoint (-log C)	5.49293	3.4628	2.71505
Predicted Endpoint (-log C)	4.9569	4.7324	4.45918
Distance	0.576	0.593	0.593
Reference	CPDB	CPDB	CPDB

Model Applicability

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

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

Feature Contribution

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	[*]=O	0.234	

FCFP_6	32	F*1CI	0.154
FCFP_6	203677720	[*]CI [*]S(=[*])(=[*])[c](: [cH]:[*]):[cH]:[*]	0.137
		for negative contributio	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.422
FCFP_6	16	[*]:[cH]:[*]	-0.354
FCFP_6	17	[*]:n:[*]	-0.149

 $C_{14}H_{11}CIN_4O_2S$

Molecular Weight: 334.78073

ALogP: 3.043 Rotatable Bonds: 4

Acceptors: 5
Donors: 2

Model Prediction

Prediction: 38.1

Unit: mg/kg_body_weight/day Mahalanobis Distance: 14

Mahalanobis Distance p-value: 2.75e-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 Sim	ilar Compounds
-----------------------	----------------

Name	796	Omeprazole	Bemitradine
Structure	H ₂ N N N N N N N N N N N N N N N N N N N	- O - H N N N N N N N N N N N N N N N N N N	H ₂ N N O
Actual Endpoint (-log C)	2.71505	3.4628	2.71351
Predicted Endpoint (-log C)	4.45918	4.7324	4.65043
Distance	0.566	0.594	0.599
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 PC11 out of range. Value: 5.1391. Training min, max, SD, explained variance: -5.9328, 4.6461, 1.629, 0.0255.

Feature Contribution

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	(*) = O	0.234	

FCFP_6	32	0 0 2 H 0 0 2 H 2 2 C [*]CI	0.154
FCFP_6	203677720	[*]S(=[*])(=[*])[c](: [cH]:[*]):[cH]:[*]	0.137
		or negative contribution	
Fingerprint	Bit/Smiles 991735244	Feature Structure	Score -0.422
FCFP_6	991733244	[*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	
FCFP_6	-2093839777	[*][c]1:[cH]:[cH]:[cH]:1	-0.378
FCFP_6	16	© 0 0 × H 2 2 C [*]:[cH]:[*]	-0.354