

Molecular Weight: 367.44809 ALogP: 3.681 Rotatable Bonds: 5 Acceptors: 5 Donors: 3

Model Prediction

Prediction: Non-Toxic

Probability: 0.384

Enrichment: 0.73

Bayesian Score: -4.99

Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 3.77e-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	Ochratoxin a	Amsacrine	Chenodiol
Structure	OH OH HO WH	N NH OF SIGNAL	OH THE TOH THE TOH
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.552	0.588	0.612
Reference	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 7(2):214-20; 1986	Arch Int Pharm 246:149- 158; 1980

Model Applicability

Unknown features are fingerprint features in 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	1256786467	[*]NC(=O)N[*]	0.12	9 out of 15	

SCFP_6	9	$ \underbrace{ \left(\begin{array}{c} \begin{array}{c} \begin{array}{c} \\ \\ \end{array} \right) \\ \\ \end{array} \right) \\ \\ \end{array} \\ \\ \\ \\ \end{array} \\ \\ \\ \\ \\ \end{array} \\ \\ \\ \\ \\ $	0.0928	45 out of 78
SCFP_6	-1272798659		0.0708	44 out of 78
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	[*]:n1:[*]:[*]:* [*]:n1:[*]:[*]:* [-0.945	0 out of 3
SCFP_6	149212520	[*][c](:[*]):n:n(:[*]	-0.448	5 out of 16
SCFP_6	1424144789	[*]C[c]1:n:[*]:[*]:n:	-0.438	1 out of 4



Molecular Weight: 361.40046 ALogP: 3.399 Rotatable Bonds: 5 Acceptors: 5 Donors: 3

Model Prediction

Prediction: Non-Toxic

Probability: 0.4

Enrichment: 0.76

Bayesian Score: -4.41

Mahalanobis Distance: 10

Mahalanobis Distance p-value: 0.0194

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

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

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

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

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

Structural Similar Compounds

Name	Ochratoxin a	Amsacrine	Benomyl
Structure	OH OH HOW WILL		
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.544	0.587	0.618
Reference	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 7(2):214-20; 1986	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.

Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Toxic in training set SCFP_6 1631845520 Image: Contribution 0.21 8 out of 12 Image: Contribution Image: Contribution Image: Contribution Image: Contribution SCFP_6 1631845520 Image: Contribution Image: Contribution</

SCFP_6	1205586762	[*]N[c](:[cH]):[c H]:[*]	0.139	11 out of 18
SCFP_6	1256786467	[*]NC(=O)N[*]	0.12	9 out of 15
	Top Feat	ures for negative of	contributior	ı
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	[*]:n1:[*]:[*]:[*]	-0.945	0 out of 3
SCFP_6	149212520	[*][c](:[*]):n:n(:[*]):[*]	-0.448	5 out of 16
SCFP_6	1424144789	[*]C[c]1:n:[*]:[*]:n:	-0.438	1 out of 4



Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Non-Toxic

Probability: 0.451

Enrichment: 0.857

Bayesian Score: -2.74

Mahalanobis Distance: 9.32

Mahalanobis Distance p-value: 0.0982

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

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

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

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

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

Structural Similar Compounds

Name		I Y171883	Benomyl
Structure	HO NO MA		
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.641	0.645	0.671
Reference	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 10(4):672-81; 1988	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	18058322	[*]CNC(=[*])[*]	0.478	4 out of 4	

SCFP_6	1435188938		0.478	4 out of 4
SCFP_6	382734644	[*]NC(=S)N[*]	0.478	4 out of 4
	Top Fea	tures for negative of	contributior	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	[*]:n1:[*]:[*]:n:[c]: 1:[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	[*][c](:[*]):n:n(:[*]):[*]	-0.448	5 out of 16
SCFP_6	1424144789	[*]C[c]1:n:[*]:[*]	-0.438	1 out of 4



Molecular Weight: 383.51369 ALogP: 4.898 Rotatable Bonds: 7 Acceptors: 5 Donors: 3

Model Prediction

Prediction: Non-Toxic

Probability: 0.46

Enrichment: 0.874

Bayesian Score: -2.47

Mahalanobis Distance: 11

Mahalanobis Distance p-value: 0.000957

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

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

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

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

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

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

Name	Ochratoxin a	Amsacrine	Dobutamine .HCI (Free base form)
Structure	OH OH HOW OF OUT	NH O	HO HO HN HN HN HN HN HN HN HN HN HN
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.671	0.674	0.698
Reference	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 7(2):214-20; 1986	Yakuri to Chiryo 7:1707- 1730; 1979

Model Applicability

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

Feature Co	Feature Contribution						
	Top fea	atures for positive o	ontribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set			
SCFP_6	1435188938	[*]C(=S)[*]	0.478	4 out of 4			
	1		1	1			

SCFP_6	18058322	[*]CNC(=[*])[*]	0.478	4 out of 4
SCFP_6	382734644	[*]NC(=S)N[*]	0.478	4 out of 4
	Top Featu	ires for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	[*]:n1:[*]:*]:n:[c]: 1:[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	[*][c](:[*]):n:n(:[*]):[*]	-0.448	5 out of 16
SCFP_6	1424144789	[*]C[c]1:n:[*]:[*]:n:	-0.438	1 out of 4

C₁₉H₁₇CIN₆ Molecular Weight: 364.83147 ALogP: 4.967

Rotatable Bonds: 5

Acceptors: 5

Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.371

Enrichment: 0.706

Bayesian Score: -5.47

Mahalanobis Distance: 9.86

Mahalanobis Distance p-value: 0.0301

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

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

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

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

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

Structural Similar Compounds

Name	Perphenazine	Acemetacin	Amsacrine		
Structure	S S C I S S C I S S C I S S C I S S C I S S C I S S C I S S C I S S S S	of o	N NH NH OSSONH		
Actual Endpoint	Toxic	Non-Toxic	Toxic		
Predicted Endpoint	Toxic	Non-Toxic	Toxic		
Distance	0.579	0.618	0.640		
Reference	Toxicol Appl Pharmacol 21(2):230-6; 1972	Oyo Yakuri 22(6):777-786; 1981	Fundam Appl Toxicol 7(2):214-20; 1986		

Model Applicability

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

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	1390366442	CI N N N N N N N N N N N N N	0.381	2 out of 2

SCFP_6	-650738059	CI NN 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	1905487031	CI NNN NFN [*][c]1:[cH]:[cH]:[c] (CI):[cH]:[cH]:1	0.153	2 out of 3
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	CI NNN NHN [*]:n1:[*]:[*]:n:[c]: 1:[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	C' N N N N N N N N N N N N N	-0.448	5 out of 16
SCFP_6	1424144789	C' N N N N N N N N N N N N N	-0.438	1 out of 4

 $C_{19}H_{17}FN_{6}$ Molecular Weight: 348.37688
ALogP: 4.508

Rotatable Bonds: 5

Acceptors: 5

Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.393

Enrichment: 0.748

Bayesian Score: -4.65

Mahalanobis Distance: 9.95

Mahalanobis Distance p-value: 0.0235

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

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

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

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

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

Structural Similar Compounds

Name	Perphenazine	Amsacrine	Budralazine	
Structure	C C C C C C C C C C C C C C C C C C C	N H H H H H H H H H H H H H H H H H H H	N N N N N N N N N N N N N N N N N N N	
Actual Endpoint	Toxic	Toxic	Non-Toxic	
Predicted Endpoint	Toxic	Toxic	Non-Toxic	
Distance	0.601	0.622	0.623	
Reference	Toxicol Appl Pharmacol 21(2):230-6; 1972	Fundam Appl Toxicol 7(2):214-20; 1986	Oyo Yakuri 21(2):321-330; 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.

SCFP_6	1390366442	[*]\N=C\[c](:[*]):[*]	0.381	2 out of 2
SCFP_6	-783770208	F[c]1:[cH]:[cH]:[*]:[0.322	4 out of 5
	Top Featur	es for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	[*]:n1:[*]:[*]:n1:[c]: 1:[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	[*][c](:[*]):n:n(:[*]):[*]	-0.448	5 out of 16
SCFP_6	1424144789	[*]C[c]1:n:[*]:[*]:n: 1:[*]	-0.438	1 out of 4

Molecular Weight: 346.38582 ALogP: 4.061 Rotatable Bonds: 5 Acceptors: 6 Donors: 2

Model Prediction

Prediction: Non-Toxic

Probability: 0.367

Enrichment: 0.698

Bayesian Score: -5.65

Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.0115

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

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

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

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

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

Structural Similar Compounds

Name	Amsacrine	Ochratoxin a	Benomyl	
Structure	N WNH NH O	OH HO WH HO		
Actual Endpoint	Toxic	Toxic	Тохіс	
Predicted Endpoint	Toxic	Toxic	Toxic	
Distance	0.546	0.625	0.640	
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	1390366442	HO N N N N N N N N N N N N N N N N N N N	0.381	2 out of 2

SCFP_6	-650738059	HO N N N N N N N N N N N N N	0.271	1 out of 1
SCFP_6	-1374800107	H ^O N N [*][c]1:[*]:[cH]:[c](O):[cH]:[cH]:1	0.205	26 out of 40
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	HO NN [*]:n1:[*]:(*]:n:[c]: 1:[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	HO N N N N N N N N N N N N N N N N N N N	-0.448	5 out of 16
SCFP_6	1424144789	HO N N N N N N N N N N N N N N N N N N N	-0.438	1 out of 4



C₁₉H₁₇N₇O₂ Molecular Weight: 375.38397 ALogP: 4.197 Rotatable Bonds: 6 Acceptors: 7 Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.331

Enrichment: 0.629

Bayesian Score: -7.19

Mahalanobis Distance: 10.1

Mahalanobis Distance p-value: 0.0164

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

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

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

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

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

Structural Similar Compounds

Name	Acemetacin	Amsacrine	Bunazosin .HCI (Free base form)
Structure	O OH	NH NH O	On NH 2 On NH
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.608	0.665	0.671
Reference	Oyo Yakuri 22(6):777-786; 1981	Fundam Appl Toxicol 7(2):214-20; 1986	Kiso to Rinsho 17:914- 924; 1983

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	1390366442	^{0,2N} ↓ ↓ N N N N N N N N N N N N N N N N N N	0.381	2 out of 2	

SCFP_6	-650738059	C ₂ N C N N (*]N/N=C\[c](:[cH]:[*]):[cH]:[*]	0.271	1 out of 1
SCFP_6	-2056718782	^{Q₂N} ^N ^N ^N ^N ^N ^N ^N 	0.0786	33 out of 58
	Top Fea	tures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	C ₂ N N N N [*]:n1:[*]:[*]:n:[c]: 1:[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	-1380909229	^{0,2N} ^N ^N [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)[N+](=[*])[*]	-0.449	6 out of 19
SCFP_6	149212520	^{02N} ^N ^N ^N ^N ^N ^N ^N 	-0.448	5 out of 16



Acceptors: 8

Donors: 2

Model Prediction

Prediction: Non-Toxic

Probability: 0.368

Enrichment: 0.699

Bayesian Score: -5.62

Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 0.00292

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

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

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

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

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

Structural Similar Compounds

Name	Ochratoxin a	LY171883	Lenampicillin .HCl (Free base form)
Structure	OH OH HOW HOW HOW HOW HOW		HN H2 HN O S HN H2 O O O O O O O O O O O O O O O O O O O
Actual Endpoint	Тохіс	Non-Toxic	Non-Toxic
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic
Distance	0.682	0.698	0.703
Reference	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 10(4):672-81; 1988	Chemotherapy 32:130- 145; 1984

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	Toxic in training set
SCFP_6	1623050981	NHO CC(=0)[c]1:[cH]:[cH]: [*]:[cH]:[cH]:1	0.271	1 out of 1

SCFP_6	1631845520	[*]C(=[*])R[t](:[*]): [*]	0.21	8 out of 12
SCFP_6	1205586762	[*]N[c](?[eH]:[*]):[c H]:[*]	0.139	11 out of 18
	Top Featu	res for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	[*]:n1:[0]:[*]	-0.945	0 out of 3
SCFP_6	-299471172	[*]C(=[*])CS[c]1:n:[*]:[*]:n:1:[*]	-0.594	1 out of 5
SCFP_6	149212520	[*][c](<u>9</u>) ^H n:n(:[*]):[*]	-0.448	5 out of 16



Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.344

Enrichment: 0.654

Bayesian Score: -6.6

Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.000823

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

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

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

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

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

Structural Similar Compounds

Name	Acemetacin	Beclomethasone Dipropionate	Hydrocortisone-17- butyrate-21-propionate
Structure			HO the second se
Actual Endpoint	Non-Toxic	Тохіс	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.701	0.730	0.731
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021- 1038; 1979	Oyo Yakuri 21:441-466; 1981

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	Toxic in training set
SCFP_6	1631845520	$[*]C(=[^{A}]) \mathbb{N}[c](:[^{*}]):$	0.21	8 out of 12

SCFP_6	1205586762	[*]N[c][^k [cH]:[*]):[c H]:[*]	0.139	11 out of 18
SCFP_6	9		0.0928	45 out of 78
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	[*]:n1 ^N [^N]:[*]:n:[c]: 1:[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	-299471172	[*]C(=[*])CS[c]1:n:[*]:[*]:n:1:[*]	-0.594	1 out of 5
SCFP_6	-1380909229	[*][c]1:[N];[cH]:[c](:[cH]:[cH]:1)[N+](=[*]][*]	-0.449	6 out of 19



Acceptors: 8

Donors: 2

Model Prediction

Prediction: Non-Toxic

Probability: 0.367

Enrichment: 0.698

Bayesian Score: -5.64

Mahalanobis Distance: 9.96

Mahalanobis Distance p-value: 0.0234

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

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

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

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

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

Structural Similar Compounds

Name	Lenampicillin .HCl (Free base form)	Bacampicillin .HCI (Free base form)	Cyclic AMP Bucladesine
Structure	HN HN SH HN SH HN SH SH SH SH SH SH SH SH SH SH SH SH SH	HN H2 HN O SH N O C C C C C C C C C C C C C C C C C C	Ho Ho Ho H
Actual Endpoint	Non-Toxic	Toxic	Non-Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Non-Toxic
Distance	0.640	0.686	0.711
Reference	Chemotherapy 32:130- 145; 1984	Chemotherapy 27:30-35; 1979	Oyo Yakuri 27(3):585-597; 1984

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	Toxic in training set
SCFP_6	-1247518081	HNO,S NS(=O)(=O)(c)1:[cH]:[cH]:[']:[cH]:[cH]:1	0.271	1 out of 1

SCFP_6	-1380395165	[*][c]1:[5b[ght]:[c](:[cH]:[cH]:1)S(=[*]) (=[*])[*]	0.255	3 out of 4
SCFP_6	1631845520	$[{}^{kj} \overset{N, k}{\leftarrow} ({}^{kj} ({}^{kj}$	0.21	8 out of 12
	Top Feat	ures for negative of	contributior	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	^N →	-0.945	0 out of 3
SCFP_6	-1463646519	[^h]:[c](:[cH]:[*])S(=O)(=O)N	-0.718	0 out of 2
SCFP_6	-299471172	$[^{t}]^{\mathbb{C}}(\underline{z}^{[*]})^{C$	-0.594	1 out of 5

Toxic

0.657

Arch Environ Contam

Toxicol 15(4):377-84; 1986

N N N H₂N S N N N N N

C₁₅H₁₅N₇OS Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.382

Enrichment: 0.725

Bayesian Score: -5.08

Mahalanobis Distance: 9.46

Mahalanobis Distance p-value: 0.0743

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

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

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

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

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

Structural Similar Compounds				
Name	Guthion	Prazosin .HCI (Free base form)	Dinoseb	
Structure		N N N H ₂ N ^M N H ₂ N ^M		
Actual Endpoint	Non-Toxic	Toxic	Toxic	

Model Applicability

Predicted Endpoint

Distance

Reference

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

Toxic

0.653

1979

Oyo Yakuri 17:57-62;

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

Arch Toxicol 43:177-186;

Non-Toxic

0.630

1980

Feature Co	ntribution			
	Top fea	atures for positive c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1357949052		0.453	8 out of 9
		[*]C(=[*])N		

SCFP_6	9		0.0928	45 out of 78
SCFP_6	-1272798659	^O H ₂ N S N N N N N N N N N N N N N N N N N N	0.0708	44 out of 78
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-572674910	⁰ H ₂ N (*]:n1:[*]:[*]:n:[c]: 1:[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	-299471172	[*]C(=[*])CS[c]1:n:[*]:[*]:n:1:[*]	-0.594	1 out of 5
SCFP_6	149212520	O H₂N S N N N N N N N N N N N N N	-0.448	5 out of 16

Doxorubicin



Molecular Weight: 543.51925 ALogP: -4.4e-002 Rotatable Bonds: 5 Acceptors: 12 Donors: 6

Model Prediction

Prediction: Toxic

Probability: 0.882

Enrichment: 1.68

Bayesian Score: 9.61

Mahalanobis Distance: 11.3

Mahalanobis Distance p-value: 0.000349

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

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

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

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

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

Structural Similar Compounds

	Compounds		
Name	Cefpiramide Sodium (Free acid form)	Latomoxef	Cefoperazone
Structure			N H H H H H H H H H H H H H H H H H H H
Actual Endpoint	Non-Toxic	Toxic	Non-Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Non-Toxic
Distance	0.760	0.821	0.850
Reference	Kiso to Rinsho 17:1000- 1004; 1983	Chemotherapy 28:1119- 1141; 1980	Chemotherapy 28:268- 291; 1980

Model Applicability

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

Feature Co	Feature Contribution						
	Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set			
SCFP_6	1702724181	$[^*]CC(O[^*])[c](:[^*]):$	0.558	9 out of 9			

SCFP_6	-2031220028		0.538	7 out of 7
SCFP_6	-37268149	$HO \rightarrow OH O OH O OH O OH O OH O OH O OH O$	0.523	6 out of 6
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1379591900	HO HO OH OH OH OH OH OH OH OH OH OH OH O	-0.282	33 out of 84
SCFP_6	834949365		-0.252	1 out of 3
SCFP_6	1505318543	$HO \xrightarrow{NH_2} OH \xrightarrow{O} O OH \xrightarrow{O} OH \xrightarrow{O}$	-0.189	2 out of 5



Molecular Weight: 367.44809 ALogP: 3.681 Rotatable Bonds: 5 Acceptors: 5 Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.785

Enrichment: 1.14

Bayesian Score: -1.52

Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.0511

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

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

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

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

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

Structural Similar Compounds

Suluciulai Sinniai Compounds					
Name	ANTHRAQUINONE; 1-((2- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	METHANE;TRIS(4- AMINOPHENYL)-		
Structure	OH HN TH O HN TH H H	HO shows where the shows where	NH ₂ H ₂ N NH ₂		
Actual Endpoint	Mild	Mild	Moderate_Severe		
Predicted Endpoint	Mild	Mild	Moderate_Severe		
Distance	0.695	0.733	0.733		
Reference	28ZPAK 245;72	28ZPAK 239;72	28ZPAK-;73;72		

Model Applicability

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

- 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: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. 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 -796673622 Image: Provide the second sec

[*]C(=[*])NC1CC[*]CC1

FCFP_10	-1539162406	[*]C[c]1:n:[*]:[*]	0.294	3 out of 3
FCFP_10	654530535	[*]C(=[*])NC1CCCCC1	0.294	3 out of 3
	Top Fea	tures 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	-332197802	[*][c]1:[*]:[c](: [*]):n:1:n:[*]	-0.507	0 out of 1
FCFP_10	1294344583	[*]NN[c](:[*]):[*]	-0.507	0 out of 1

$C_{19}H_{19}N_7O$

Molecular Weight: 361.40046 ALogP: 3.399 Rotatable Bonds: 5 Acceptors: 5 Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.643

Enrichment: 0.933

Bayesian Score: -4.44

Mahalanobis Distance: 6.86

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- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	METHANE;TRIS(4- AMINOPHENYL)-	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-
Structure	OH HN W N HN W H H H	H ₂ N NH ₂	HO shows which have a state of the shows where
Actual Endpoint	Mild	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.653	0.709	0.719
Reference	28ZPAK 245;72	28ZPAK-;73;72	28ZPAK 239;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: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. 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 -1539162406 Image: Colspan="2">Image: Colspan="2">Output: Colspan="2">Output: Colspan="2">Colspan="2">Contribution FCFP_10 -1539162406 Image: Colspan="2">Output: Colspan="2">Output: Colspan="2">Moderate_Severe in training set FCFP_10 -1539162406 Image: Colspan="2">Output: Colspan="2" Image: Colspan="2">Output

1:[*]

FCFP_10	-402549409	[']NC(=0)N[c](:[cH]:[']):[cH]:[']	0.186	1 out of 1
FCFP_10	3		0.165	383 out of 491
	Top Featur	es for negative c	ontribution	
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	-332197802	[*][c]1:[*]:[*]	-0.507	0 out of 1



Acceptors: 5

Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.734

Enrichment: 1.07

Bayesian Score: -2.85

Mahalanobis Distance: 7.55

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 Similar Compounds

Structural Sinniar Compounds					
Name	ANTHRAQUINONE; 1-((2- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	METHANE;TRIS(4- AMINOPHENYL)-	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-		
Structure	OH HN W N H H	NH ₂ H ₂ N NH ₂	OH OSSN ^M OH		
Actual Endpoint	Mild	Moderate_Severe	Mild		
Predicted Endpoint	Mild	Moderate_Severe	Mild		
Distance	0.673	0.737	0.747		
Reference	28ZPAK 245;72	28ZPAK-;73;72	28ZPAK 190;72		

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-1939253119	[*]C(=[*])NCC	0.332	5 out of 5	
			-	-	

FCFP_10	-1539162406	[*]C[c]1:n:[*]:[*]:n: 1:[*]	0.294	3 out of 3
FCFP_10	-1272709286	[*]NCC	0.285	234 out of 266
	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	-332197802	[*][c]1:[*]:[c](: [*]):n:1:n:[*]	-0.507	0 out of 1
FCFP_10	1934560662	[*]NC(=S)NCC	-0.507	0 out of 1

$C_{19}H_{25}N_7S$ Molecular Weight: 383.51369

ALogP: 4.898 Rotatable Bonds: 7

Acceptors: 5

Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.79

Enrichment: 1.15

Bayesian Score: -1.36

Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.0543

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

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

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

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

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

Name	4;4'-DIAMINO-1;1'- DIANTHRIMIDE	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	ANTHRAQUINONE; 1-((2- HYDROXYETHYL)AMINO)- 4-(METHYLAMINO)-
Structure	NH 2 NH 2	HO the set of the set	OH HN TH O THE N H H
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.764	0.800	0.828
Reference	28ZPAK-;125;72	28ZPAK 239;72	28ZPAK 245;72

Model Applicability

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

- 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: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

Feature Co	Feature Contribution						
	Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set			
FCFP_10	-796673622	I'IC(=['])NC1CC[']CC1	0.317	4 out of 4			

FCFP_10	-1539162406	[*]C[c]1:n: 1:[*]	0.294	3 out of 3	
FCFP_10	654530535	I''JC(=[''])NC1CCCCC1	0.294	3 out of 3	
Top Features for negative 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	-332197802	[*][c]1:[*]:[*]	-0.507	0 out of 1	

C₁₉H₁₇CIN₆ Molecular Weight: 364.83147 ALogP: 4.967 Rotatable Bonds: 5

Acceptors: 5

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.802

Enrichment: 1.16

Bayesian Score: -0.913

Mahalanobis Distance: 7.84

Mahalanobis Distance p-value: 0.951

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	1-BENZOYLAMINO-4- METHOXY-5- CHLORANTHRAQUINONE	ANTHRAQUINONE; 1;1'- IMINODI-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER		
Structure	HN AND THE CITY OF CITY		CI O O O CI		
Actual Endpoint	Mild	Mild	Moderate_Severe		
Predicted Endpoint	Mild	Mild	Moderate_Severe		
Distance	0.601	0.616	0.637		
Reference	28ZPAK-;90;72	28ZPAK-;125;72	CIGET* -;-;77		

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

Feature Contribution Top features for positive contribution					
FCFP_10	-149636017	C' N-N [*]=C[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.352	7 out of 7	

FCFP_10	-1508180856	CI NNN NH NH NH NH NH NH NH NH NH NH NH NH	0.329	16 out of 17		
FCFP_10	-745491832	CI CI CI CI CI CI CI CI CI CI CI CI CI C	0.304	29 out of 32		
	Top Features for negative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set		
FCFP_10	4427049	C' NNN NNN [*][c](:[*]):n:n(:[*]):[*]	-1.29	0 out of 4		
FCFP_10	-332197802	C' N N N N N N N N N N N N N	-0.507	0 out of 1		
FCFP_10	-1320007763	C' N N N N N N N N N N N N N N N	-0.316	19 out of 40		

Acceptors: 5

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.802

Enrichment: 1.16

Bayesian Score: -0.913

Mahalanobis Distance: 7.8

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.

Name	1-BENZOYLAMINO-4- METHOXY-5- CHLORANTHRAQUINONE	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	ANTHRAQUINONE; 1;1'- IMINODI-
Structure	HN ru HN ru U U U U U U U U U U	CI OH	H H H H H H H H H H H H H H H H H H H
Actual Endpoint	Mild	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.595	0.634	0.653
Reference	28ZPAK-;90;72	CIGET* -;-;77	28ZPAK-;125;72

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

FCFP_10	-1508180856	[*][C]1:[CH]:[CH]:[C] (CI):[CH]:[CH]:1	0.329	16 out of 17
FCFP_10	-745491832	CI[c]1:[cH]:[cH]:1	0.304	29 out of 32
	Top Feat	ures for negative of	contributio	n
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	-332197802	[*][c]1:[*]:[*]:[c](: [*]):n:1:n:[*]	-0.507	0 out of 1
FCFP_10	-1320007763	[*][c](:[*]):[c] ^N ,N N N N N N N N N N N N N N N N N N N	-0.316	19 out of 40

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.811

Enrichment: 1.18

Bayesian Score: -0.525

Mahalanobis Distance: 8.13

Mahalanobis Distance p-value: 0.893

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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-	s-TRIAZINE; 2;4- BIS(ISOPROPYLAMINO0- 6-(METHYLTHIO)-	s-TRIAZINE; 2-(tert- BUTYLAMINO)-4- (ETHYLAMINO)-6- (METHYLTHIO)-		
Structure	HO to HO TO HO	N N N N N N N N N N N N N N N N N N N	Z Z Z Z Z Z		
Actual Endpoint	Mild	Mild	Moderate_Severe		
Predicted Endpoint	Mild	Mild	Moderate_Severe		
Distance	0.615	0.691	0.696		
Reference	28ZPAK 239;72	CIGET* -;-;77	CIGET* -;-;77		

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1066794953	HO N N N N N N N N N N N N N N N N N N N	0.378	13 out of 13

FCFP_10	-158888774	HO N N N N N N N N N N N N N	0.356	24 out of 25
FCFP_10	-1539162406	H ^O N N N N N N N N N N N N N	0.294	3 out of 3
	Top Feat	tures for negative of	contributior	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	H ^O N N N N N N N N N N N N N	-1.29	0 out of 4
FCFP_10	-332197802	HO NN NN NN NN NN NN NN NN NN N	-0.507	0 out of 1
FCFP_10	-1320007763	H ^O [*][c](:[*]):[c] ^N N [*][cH]:[cH]:[c]: 1:[*]	-0.316	19 out of 40

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe

Acceptors: 7

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.772

Enrichment: 1.12

Bayesian Score: -1.94

Mahalanobis Distance: 7.74

Mahalanobis Distance p-value: 0.965

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

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

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

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

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

Structural Similar Compounds

Name	COLCHICINE	1;4-PENTADIENE-3- ONE;1;5-BIS(4-(2;3- DIDEHYDROTRIAZIRIDIN YL)PHENYL)-	ANTHRAQUINONE; 1;1'- IMINODI-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.709	0.741	0.766
Reference	AJOPAA 31;837;48	28ZPAK-;123;72	28ZPAK-;125;72

Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]
- 6. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 7. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 8. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 9. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

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

FCFP_10	-1539162406	O2N C	0.294	3 out of 3
		(*]C[c]1:n:[*]:[*]:n: 1:[*]		
FCFP_10	-2100785893	°'N C	0.256	2 out of 2
		[*1\N=C\[c](:[*]);[*]		
FCFP_10	3	0 ₂ N 0 ₂ N	0.165	383 out of 491
	Top Fea	tures for negative	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	O'N CO	-1.29	0 out of 4
		[*][c](:[*]):n:n(:[*]):[*]		
FCFP_10	-332197802	02N C	-0.507	0 out of 1
		[*][c]1:[*]:[*]:[c](: [*]):n:1:n:[*]		

FCFP_10	-1320007763	⁰ 2N	-0.316	19 out of 40
		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~		
		N ^H		
		[*][c](:[*]):[c] [¶] .[cH		
		]:[cH]:[cH]:[*]:[c]: 1·[*]		
		1		

# TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe

# $C_{22}H_{19}N_7O_3S$ Molecular Weight: 461.49635 ALogP: 3.42 Rotatable Bonds: 7 Acceptors: 8

Donors: 2

# **Model Prediction**

#### Prediction: Mild

Probability: 0.715

Enrichment: 1.04

Bayesian Score: -3.24

Mahalanobis Distance: 12.2

#### Mahalanobis Distance p-value: 2.02e-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	4;4'-DIAMINO-1;1'- DIANTHRIMIDE	COLCHICINE	2;7-NAPHTHALENE DISULFONIC ACID;4- AMINO-5-HYDROXY-;P- TOLUENE SULFONATE (ESTER)		
Structure	NH 2 NH 2		HO NH 2 OH		
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe		
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe		
Distance	0.753	0.829	0.852		
Reference	28ZPAK-;125;72	AJOPAA 31;837;48	28ZPAK-;194;72		

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-548632217	$ \begin{array}{c} & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & & \\ & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & $	0.319	54 out of 59	

FCFP_10	-1539162406	[*]C[c ^{qq} ?ft ^t [*]:[*]:n: 1:[*]	0.294	3 out of 3
FCFP_10	7	COOH [*]O	0.219	117 out of 142
	Top Featu	res for negative o	ontribution	_
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	[*][c](:[ ^a ]) th n:n(:[*] ):[*]	-1.29	0 out of 4
FCFP_10	-332197802	[*][c] f:[0]![*]:[c](: [*]):n:1:n:[*]	-0.507	0 out of 1
FCFP_10	-790336137	[*]C(=[*])V[6]1:[CH]: [CH]:[C](:[CH]:[CH])[*]	-0.507	0 out of 1



# **Model Prediction**

#### Prediction: Mild

Probability: 0.726

Enrichment: 1.05

Bayesian Score: -3.03

Mahalanobis Distance: 8.28

#### Mahalanobis Distance p-value: 0.85

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

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

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

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

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

# Structural Similar Compounds

Name	COLCHICINE	1;8;9- ANTHRACENETRIOL; TRIACETATE	1;4-PENTADIENE-3- ONE;1;5-BIS(4-(2;3- DIDEHYDROTRIAZIRIDINY L)PHENYL)-	
Structure				
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild	
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild	
Distance	0.780	0.872	0.881	
Reference	AJOPAA 31;837;48	BJOPAL 53;819;69	28ZPAK-;123;72	

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 4. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 5. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 6. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set		
		•	-			

FCFP_10	-1539162406		0.294	3 out of 3
FCFP_10	3	[*]C[c ['] ]%:n:[*]:[*]:n: 1:[*]	0.165	383 out of 491
FCFP_10	907007053	[*]N[*]	0.138	32 out of 42
<b>.</b>	Top Fea	atures for negative	contributior	י ו
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	NH _O ^S N [*][c]( ^N [ ⁹ ]):n:n(:[*]	-1.29	0 out of 4
FCFP_10	-332197802	).1 J       NHO       NHO       [*][c] ^{N:Q*} ]:[*]:[c](:       [*][:n:1:n:[*]	-0.507	0 out of 1

FCFP_10	-1320007763		-0.316	19 out of 40
		NHO° SN		
		[*][c](:[ <mark>x</mark> ]):[c]1:[cH		
		]:[cH]:[cH]:[*]:[c]: 1 · [*]		
		··[ ]		

# TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe

# $C_{21}H_{20}N_8O_3S_2$ Molecular Weight: 496.56529 ALogP: 2.496 Rotatable Bonds: 7 Acceptors: 8

Donors: 2

# **Model Prediction**

#### Prediction: Mild

Probability: 0.659

Enrichment: 0.956

Bayesian Score: -4.2

Mahalanobis Distance: 8.56

#### Mahalanobis Distance p-value: 0.737

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

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

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

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

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

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

Name	2;2'-Stilbenedisulfonic acid; 4;4'-dinitro-	2;7-NAPHTHALENE DISULFONIC ACID;4- AMINO-5-HYDROXY-;P- TOLUENE SULFONATE (ESTER)	4;4'-DIAMINO-1;1'- DIANTHRIMIDE	
Structure	OH OF OF	HO ILO HO	NH 2 HN 40 HN 40 HN 40 HN 40 HN 2 HN 40 HN 4	
Actual Endpoint	Mild	Moderate_Severe	Mild	
Predicted Endpoint	Mild	Moderate_Severe	Mild	
Distance	0.783	0.787	0.803	
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1062;86	28ZPAK-;194;72	28ZPAK-;125;72	

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set	
FCFP_10	-1539162406	[*]@fc]1:n:[*]:[*]	0.294	3 out of 3	

FCFP_10	1070061035	NH ⁶ (*]S(=[*])(=[*])N	0.239	284 out of 338
FCFP_10	1763397430	[*]N[c]1:6H]:[cH]:[c] ](:[cH]:[cH]:1)S(=O) (=O)N	0.186	1 out of 1
	Top Featu	ures for negative of	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	[*][te](:[*]):n:n(:[*] ):[*]	-1.29	0 out of 4
FCFP_10	-332197802	^{NH} O ^S ^{NH} N ^N	-0.507	0 out of 1
FCFP_10	-790336137	[*]C(=[*])N(G)4;[cH]: [cH]:[c](:[cH]:[cH]: 1)C(=[*])[*]	-0.507	0 out of 1



Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

# Model Prediction

Prediction: Mild

Probability: 0.777

Enrichment: 1.13

Bayesian Score: -1.78

Mahalanobis Distance: 7.91

#### Mahalanobis Distance p-value: 0.941

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

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

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

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

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

# Structural Similar Compounds

Structural Similar Compounds						
Name	2;5-DICHLORO-4(3'- METHYL-5' PYRAZOLON- 1'-YL)BENZENE SULFONIC ACID	Benzenesulfonic acid; 3;5-bis(methoxycarbonyl)- ; sodium salt	Benzenesulfonic acid; 2- anilino-5-nitro-			
Structure		O SHUT O O O O O O O O O O O O O O O O O O O				
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe			
Predicted Endpoint	Mild	Mild	Mild			
Distance	0.607	0.662	0.678			
Reference	28ZPAK-;186;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J ;1059;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J ;1061;86			

# Model Applicability

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

- All properties and OPS components are within expected ranges. 1.
- Unknown FCFP 2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*] 2.

#### Feature Contribution Top features for positive contribution **Bit/Smiles** Feature Structure Score Fingerprint Moderate Severe in training set FCFP 10 1539162406 0.294 B out of 3 [*]C[c]1:n:[*]:[*]:n: 1:[*]

FCFP_10	1070061035	$\begin{bmatrix} 0 & & & N & N \\ H_2 N & & & & N \\ & & & & & N \\ & & & & & N \\ & & & &$	0.239	284 out of 338
FCFP_10	3		0.165	383 out of 491
	Top Featu	res for negative o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	°, s, N=N H₂N s, N=N N=N [*][c](:[*]):n:n(:[*] ):[*]	-1.29	0 out of 4
FCFP_10	-332197802	O H₂N S N=N N=N N=N N=N N=N N=N N=N	-0.507	0 out of 1
FCFP_10	-1320007763	[*][c](:[*]):[c]1:[cH ]:[cH]:[cH]:[*][c]: 1:[*]	-0.316	19 out of 40



Donors: 6

# **Model Prediction**

#### Prediction: Mild

Probability: 0.00731

Enrichment: 0.0106

Bayesian Score: -15

Mahalanobis Distance: 16.1

#### Mahalanobis Distance p-value: 1.56e-017

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

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

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

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

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

# TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe

# **Structural Similar Compounds**

Name	Methanol; (s-triazine- 2;4;6-triyltrinitrilo)hexa-	2-Naphthalenesulfonic acid; 5;6'-iminobis(1- hydroxy-	1;3;6-NAPHTHALENE TRISULFONIC ACID;7- AMINO-
Structure	HO HO HO	HO - DH HO - OH HO - OH O OH	HO SHIT HO NH 2 0 - II OH O 0 - II OH O 0 - II OH O
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.947	0.991	1.168
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;876;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1065;86	28ZPAK-;190;72

# Model Applicability

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

- 1. OPS PC18 out of range. Value: -4.3032. Training min, max, SD, explained variance: -4.1566, 3.8733, 1.069, 0.0139.
- 2. OPS PC20 out of range. Value: -3.1531. Training min, max, SD, explained variance: -3.1297, 4.2414, 1.03, 0.0129.

### **Feature Contribution**

#### Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set

FCFP_10	-1601875224		0.352	7 out of 7
		H ^O → → → → → → → → → → → → → → → → → → →		
FCFP_10	-1272709286		0.285	234 out of 266
FCFP 10	1070061035	[*]NCC	0.239	284 out of 338
	1070001033		0.239	204 001 01 330
		[*]S(=[*])(=[*])N		
	Top Fea	tures for negative of	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	979221427	HO HO HO (*)[C]:[*]EE[([)]: [c]2C(=[*])[C]3:[C]( [*]):[*]:[CH]:[CH]:[ C]:3C(=O)[C]:2:[C]:1 O	-1.6	0 out of 6
FCFP_10	-22686407	[*][c](:[*])):[c]100 H O (Finite Constraints) = 0 (Finite Constraints	-1.5	7 out of 53
	·		·	

FCFP_10	-1549669478	NH	-1.5	7 out of 53
		HO HO HO HO HO HO HO HO HO HO HO HO HO H		



# Acceptors: 5

#### Donors: 3

# **Model Prediction**

#### Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.39

Mahalanobis Distance: 9.4

#### 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	Disperse Black 9	ANTHRAQUINONE; 1-((2- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	METHANE;TRIS(4- AMINOPHENYL)-	
Structure	HO NH2	OH HN TH O THE NY	H ₂ N NH ₂	
Actual Endpoint	Non-Irritant	Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Irritant	Irritant	
Distance	0.668	0.673	0.720	
Reference	J. Am. Coll. Toxicol. 5(3):205;1986	28ZPAK 245;72	28ZPAK-;73;72	

# Model Applicability

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

- 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: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

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	-885550502	[*]C([*])NC(=[*])[*]	0.18	64 out of 66
FCFP_12	4427049	[*][c](:[*])::::(:[*]	0.167	4 out of 4
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175638033	[*]C1[*]CCCC1	-0.133	207 out of 293
FCFP_12	-1525101452		-0.127	108 out of 152
FCFP_12	1		0	872 out of 1051



Rotatable Bonds: 5

Acceptors: 5 Donors: 3

# **Model Prediction**

#### Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.51

Mahalanobis Distance: 6.14

#### 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- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	Disperse Black 9	METHANE;TRIS(4- AMINOPHENYL)-	
Structure			H ₂ NH ₂	
Actual Endpoint	Irritant	Non-Irritant	Irritant	
Predicted Endpoint	Irritant	Non-Irritant	Irritant	
Distance	0.636	0.639	0.699	
Reference	28ZPAK 245;72	J. Am. Coll. Toxicol. 5(3):205;1986	28ZPAK-;73;72	

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

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	4427049	[*][c](:[*]):n:n(:[*] ):[*]	0.167	4 out of 4
FCFP_12	-1539162406	[*]C[c]1:n:[*]:[*]:n:	0.156	3 out of 3
	Top Feat	ures for negative of	contributior	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-2093839777	[*][c]1:[cH]:[cH]:1	0	94 out of 121
FCFP_12	-792685140	[*]C(=[*])N[c] ¹ :[cH]: [cH]:[cH]:[cH]:[cH]:	0	5 out of 6
FCFP_12	0	C	0	1184 out of 1397



# **Model Prediction**

#### Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.45

Mahalanobis Distance: 6.86

#### 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	Disperse Black 9	ANTHRAQUINONE; 1-((2- HYDROXYETHYL)AMINO) -4-(METHYLAMINO)-	2- NAPHTHALENESULFONI C ACID; 4-HYDROXY-7- (METHYLAMINO)-	
Structure	HO NR NR NH 2	OH HN TH OH HN TH H H	OH OH OH	
Actual Endpoint	Non-Irritant	Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Irritant	Irritant	
Distance	0.616	0.656	0.724	
Reference	J. Am. Coll. Toxicol. 5(3):205;1986	28ZPAK 245;72	28ZPAK 190;72	

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

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	-885550502	[*]C([*])NC(=[*])[*]	0.18	64 out of 66
FCFP_12	-1939253119	[*]C(=[*])NCC	0.175	5 out of 5
	Top Feat	ures for negative of	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1272798659	[*]CCC	0	517 out of 643
FCFP_12	307419094	[*][c](:[*]):[c](:[*]):[*]	0	43 out of 52
FCFP_12	1	$\begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & $	0	872 out of 1051



Molecular Weight: 383.51369 ALogP: 4.898 Rotatable Bonds: 7 Acceptors: 5 Donors: 3

# **Model Prediction**

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.32

Mahalanobis Distance: 9.32

#### Mahalanobis Distance p-value: 0.342

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

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

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

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

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

# **Structural Similar Compounds**

Structural Similar Compounds					
Name	4;4'-DIAMINO-1;1'- DIANTHRIMIDE	Disperse Black 9	BENZANILIDE;2';2'''- DITHIOBIS-		
Structure	NH 2 HN mm VNH 2	HO N I N N N N N N N N N N N N N N N N N	H N N H		
Actual Endpoint	Irritant	Non-Irritant	Non-Irritant		
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant		
Distance	0.747	0.748	0.774		
Reference	28ZPAK-;125;72	J. Am. Coll. Toxicol. 5(3):205;1986	28ZPAK-;173;72		

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

#### 

FCFP_12	-885550502	[*]C([*])NC(=[*])[*]	0.18	64 out of 66
FCFP_12	4427049	[*][c](:[*]):n:n(:[*] ):[*]	0.167	4 out of 4
	Top Feat	ures for negative of	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175638033		-0.133	207 out of 293
FCFP_12	-1525101452		-0.127	108 out of 152
FCFP_12	1		0	872 out of 1051



#### Donors: 1

# **Model Prediction**

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.81

Mahalanobis Distance: 6.68

#### 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 Compoun	ds
----------------------------	----

Name	1-BENZOYLAMINO-4- METHOXY-5- CHLORANTHRAQUINONE	ANTHRAQUINONE; 1;1'- IMINODI-	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER		
Structure	HN PAR		CI OH		
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Irritant	Irritant	Irritant		
Distance	0.593	0.607	0.633		
Reference	28ZPAK-;90;72	28ZPAK-;125;72	CIGET* -;-;77		

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

# Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Irritant in training set FCFP_12 1508180856 0.2 17 out of 17 Image: Structure (CI):[cH]:[cH]:[cH]:[cI]:[cH]:[cI]] 0.2 17 out of 17

FCFP_12	17		0.189	48 out of 49
FCFP_12	-149636017	[*]=C[c]1:[cH]:[cH]:1	0.184	7 out of 7
	Top Fea	tures for negative of	contributior	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	136597326		0	612 out of 753
FCFP_12	203677720	C' N-N N-N (*]=C[c](:[cH]:[*]):[ cH]:[*]	0	319 out of 382
FCFP_12	-1272798659		0	517 out of 643



#### Donors: 1

# **Model Prediction**

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 2.2

Mahalanobis Distance: 6.6

#### 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.

|--|

Name	1-BENZOYLAMINO-4- METHOXY-5- CHLORANTHRAQUINONE	BENZILIC ACID; 4;4'- DICHLORO-; ISOPROPYL ESTER	ANTHRAQUINONE; 1;1'- IMINODI-		
Structure	HN AND THO THOMAS AND THE AND	CI O O CI			
Actual Endpoint	Irritant	Irritant	Irritant		
Predicted Endpoint	Irritant	Irritant	Irritant		
Distance	0.590	0.631	0.641		
Reference	28ZPAK-;90;72	CIGET* -;-;77	28ZPAK-;125;72		

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

# Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score Irritant in training set FCFP_12 1508180856 $\int_{V_{n}} \int_{V_{n}} \int_{V_{$

FCFP_12	17	[*]:n:[*]	0.189	48 out of 49
FCFP_12	-149636017	[*]=C[c]1:[cH]:[cH]:1	0.184	7 out of 7
	Top Fea	atures for negative of	contribution	า
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-453677277	[*][c]1:[*]:[cH]:[c]( C=[*]):[cH]:[cH]:1	0	264 out of 323
FCFP_12	991735244	[*]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	136597326		0	612 out of 753



Acceptors: 6

Donors: 2

# **Model Prediction**

#### Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.27

Mahalanobis Distance: 6.78

#### Mahalanobis Distance p-value: 0.999

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

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

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

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

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

Structural	Similar	Compounde
Juuciulai	Jiiiiai	Compounds

Name	ANTHRAQUINONE; 1- AMINO-4-HYDROXY-2- PHENOXY-	2;2';-Dihydroxy-4;4'- dimethoxybenzophenone	s-TRIAZINE; 2;4- BIS(ISOPROPYLAMINO0- 6-(METHYLTHIO)-		
Structure	HO the HO	ON THE HOWEN THE O	NH N HN HN HN N N N N N N N N N N N N N		
Actual Endpoint	Irritant	Non-Irritant	Irritant		
Predicted Endpoint	Irritant	Non-Irritant	Irritant		
Distance	0.607	0.655	0.679		
Reference	28ZPAK 239;72	J. Am. Coll. Toxicol. 2(5):35;1983	CIGET* -;-;77		

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
FCFP_12	17	HO NN NH [*]:n:[*]	0.189	48 out of 49	

FCFP_12	4427049	HO NN NN NN NN NN NN NN NN NN N	0.167	4 out of 4
FCFP_12	-1539162406	HO NN NN [*]C[c]1:n:[*]:[*]:n: 1:[*]	0.156	3 out of 3
	Top Featu	ures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1066794953	HO NN N-N [*][c]1:[cH]:[cH]:[c] (O):[cH]:[cH]:1	-0.0509	13 out of 17
FCFP_12	-158888774	HO N N N N N N N N N N N N N N N N N N N	0	25 out of 30
FCFP_12	-453677277	H ^O N N N N N N N N N N N N N	0	264 out of 323



Acceptors: 7

#### Donors: 1

# **Model Prediction**

#### Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.09

Mahalanobis Distance: 7.36

#### Mahalanobis Distance p-value: 0.991

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

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

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

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

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

Structural Similar Compounds				
Name	COLCHICINE	1;4-PENTADIENE-3- ONE;1;5-BIS(4-(2;3- DIDEHYDROTRIAZIRIDIN YL)PHENYL)-	ANTHRAQUINONE; 1;1'- IMINODI-	
Structure				
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.692	0.741	0.753	
Reference	AJOPAA 31;837;48	28ZPAK-;123;72	28ZPAK-;125;72	

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]
- 6. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 7. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 8. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 9. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

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

FCFP_12	17	O ₂ N	0.189	48 out of 49
		[*]:n:[*]		
FCFP_12	4427049		0.167	4 out of 4
		[*][c](:[*]):n:n(:[*] ):[*]		
FCFP_12	-1539162406		0.156	3 out of 3
		[*]C[c]1:n:[*]:[*]:n: 1:[*]		
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	8		-0.0561	3 out of 4
		[*1[N+1](=[*1)[*1])		
FCFP_12	991735244		0	237 out of 291
		[*]:[c]1:[*]:[cH]:[cH] ]:[cH]:[cH]:1		

FCFP_12	136597326	0 ₂ N	0	612 out of 753
		N-N		
		[*]CC		


## **Model Prediction**

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 2.93

Mahalanobis Distance: 7.78

#### Mahalanobis Distance p-value: 0.961

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

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

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

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

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

|--|

Name	4;4'-DIAMINO-1;1'- DIANTHRIMIDE	;4'-DIAMINO-1;1'- COLCHICINE JIANTHRIMIDE	
Structure	NH 2 NH 2		HO she with NH 2
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.748	0.825	0.843
Reference	28ZPAK-;125;72	AJOPAA 31;837;48	28ZPAK 239;72

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
FCFP_12	-1549222613	[*]:[c](:[*])C(=0)0	0.198	14 out of 14		

FCFP_12	1175665944	COOH [']CC(=0)N[c](:[cH];[ ']);[cH]:[']	0.198	14 out of 14
FCFP_12	937923569	[*]:[cH]:[cH]:[* ])C(=O)O	0.198	14 out of 14
	Top Fea	atures for negative of	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-773983804	[*]N[c]f:[cH]:[cH]:1	0	102 out of 121
FCFP_12	-453677277	[*][c]1:[*]?[cH]:[c]( C=[*]):[cH]:[cH]:1	0	264 out of 323
FCFP_12	1872154524	$ \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ & & & &$	0	563 out of 690



## **Model Prediction**

#### Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.94

Mahalanobis Distance: 8.14

#### Mahalanobis Distance p-value: 0.892

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

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

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

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

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

## Structural Similar Compounds

Name	COLCHICINE	1;8;9- ANTHRACENETRIOL; TRIACETATE	1;4-PENTADIENE-3- ONE;1;5-BIS(4-(2;3- DIDEHYDROTRIAZIRIDINY L)PHENYL)-	
Structure				
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.772	0.866	0.872	
Reference	AJOPAA 31;837;48	BJOPAL 53;819;69	28ZPAK-;123;72	

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 4. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 5. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 6. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

Top features for positive contribution						
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set		
			1			

FCFP_12	1175665944	N N N N N N N N N N N N N N N N N N N	0.198	14 out of 14
FCFP_12	17	NHOS NN	0.189	48 out of 49
FCFP_12	-1410079687	[*]S[c] [@] :n:[*]:[*]:n: 1:[*]	0.187	8 out of 8
	Top Feat	ures for negative	contribution	1
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	8	[*][N+](=[*])[*]	-0.0561	3 out of 4
FCFP_12	1		0	872 out of 1051

FCFP_12	0	. 🗘	0	1184 out of 1397
		$\varphi$		
		[`] ^{NO} ²[*]C[*]		



## **Model Prediction**

#### Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 2.72

Mahalanobis Distance: 8.46

#### Mahalanobis Distance p-value: 0.781

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

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

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

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

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

## Structural Similar Compounds

Name	2;2'-Stilbenedisulfonic acid; 4;4'-dinitro-	2;7-NAPHTHALENE DISULFONIC ACID;4- AMINO-5-HYDROXY-;P- TOLUENE SULFONATE (ESTER)	4;4'-DIAMINO-1;1'- DIANTHRIMIDE
Structure	OH OF OF OF	HO NH 2	HIN rts of the second s
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.783	0.786	0.793
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1062;86	28ZPAK-;194;72	28ZPAK-;125;72

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]

Top features for positive contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set			
FCFP_12	1175665944	HINO,S ["JCC(=0)N[c](:[cH]:[ ")):[cH]:["]	0.198	14 out of 14			

FCFP_12	17	$\sum_{\substack{N \in \mathcal{O} \\ N \in \mathcal{O} \\ SO_{2}, NH_{2} \\ [*]:n:[*]}} N = N$	0.189	48 out of 49
FCFP_12	-1410079687	[*] <b>S</b> [c]1:n:[*]:[*]	0.187	8 out of 8
	Top Feat	ures for negative of	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1872154524	NH ^O SO 3NH [*]C(=O)[*]	0	563 out of 690
FCFP_12	-1272768868	[*]SCC(=[*])[*]	0	396 out of 514
FCFP_12	-1272798659	NH ^N O SO_NH [*]CCC	0	517 out of 643

Ο  $H_{2}N$  $C_{15}H_{15}N_{7}OS$ 

Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

## Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bavesian Score: 1.62

Mahalanobis Distance: 7.73

#### Mahalanobis Distance p-value: 0.966

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

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

Enrichment: An estimate of enrichment, that is, the increased Bayesian Score: The standard Laplacian-modified Bayesian score.

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

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

- 1. All properties and OPS components are within expected ranges.
- Unknown FCFP 2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*] 2.

#### Feature Contribution Top features for positive contribution **Bit/Smiles** Feature Structure Score Fingerprint Irritant in training set FCFP 12 17 0.189 48 out of 49 [*]:n:[*]

FCFP_12	-1410079687	[*]S[c]1:n:[*]:[*]	0.187	8 out of 8
FCFP_12	4427049	⁰ H ₂ N S N N N N N N N N N N N N N N N N N N	0.167	4 out of 4
	Top Featur	es for negative c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1272798659		0	517 out of 643
FCFP_12	-1272768868	(*)SCC(=[*])[*]	0	396 out of 514
FCFP_12	307419094	^O H ₂ N S N N N N N N N N N N N N N N N N N N	0	43 out of 52

## Doxorubicin



Acceptors: 12

Donors: 6

## **Model Prediction**

#### Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.88

Mahalanobis Distance: 14.4

#### Mahalanobis Distance p-value: 1.33e-011

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

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

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

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

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

## www.stumel.Classiles.Common.com

Name	Methanol; (s-triazine- 2;4;6-triyltrinitrilo)hexa-	2-Naphthalenesulfonic acid; 5;6'-iminobis(1- hydroxy-	1;3;6-NAPHTHALENE TRISULFONIC ACID;7- AMINO-	
Structure		HO HO OH	O HO NH ₂ O S NH ₂ O S NH ₂ O S NH ₂	
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Irritant	Irritant	
Distance	0.923	0.958	1.164	
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J;876;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1065;86	28ZPAK-;190;72	

## Model Applicability

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

OPS PC15 out of range. Value: 4.2823. Training min, max, SD, explained variance: -4.3422. 1. 4.1885. 1.08. 0.0142.

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-415156552	$HO \xrightarrow{\text{NH}_2} OH \xrightarrow{\text{O}} OH \xrightarrow{O} OH$	0.184	7 out of 7

FCFP_12	-1601875224	HO HO HO OH (*]C([*])[c]1:[C]([*]) [c]([*]):[c](C (=[*])[*]):[c]:10	0.184	7 out of 7
FCFP_12	-1099193755	$ \begin{array}{c} HO \longrightarrow H_{2} \\ HO \longrightarrow HO \longrightarrow HO \\ HO \longrightarrow HO \\ OH \\ OH$	0.175	5 out of 5
	Top Feat	ures for negative of	contribution	I
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1186333723	HO + O + O + O + O + O + O + O + O + O +	-0.344	2 out of 4
FCFP_12	1312974458	$\begin{array}{c} HO & NH_2 \\ HO & OH & OH \\ HO & OH & OH \\ HO & OH & OH$	-0.268	1 out of 2
FCFP_12	-880844584	$\begin{matrix} HO & H_2 \\ HO & HO \\ HO & HO \\ OH & OH \\ O$	-0.132	2 out of 3

## TOPKAT_Skin_Irritancy_Mild_vs_Moderate_Severe

# C₁₉H₁₇CIN₆ Molecular Weight: 364.83147 ALogP: 4.967 Rotatable Bonds: 5

Acceptors: 5

Donors: 1

## **Model Prediction**

#### Prediction: Mild

Probability: 0.0319

Enrichment: 0.0866

Bayesian Score: -12.4

Mahalanobis Distance: 8.52

#### Mahalanobis Distance p-value: 0.527

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

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

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

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

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

Structural Similar Compounds				
Name	Anthraquinone, 1,1'- iminodi-	Phosphorothioic acid, O- ethyl S-propyl O-(2,4,6- trichlorophenyl) es ter	Aniline, 2,4-bis(o- methylphenoxy)-	
Structure	H H H H H H H H H H H H H H H H H H H		H ₂ N ⁴	
Actual Endpoint	Mild	Moderate_Severe	Mild	
Predicted Endpoint	Mild	Mild	Mild	
Distance	0.669	0.711	0.733	
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: OTS0535844	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986	

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]

itribution			
Top fe	atures for positive c	ontribution	
Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
	tribution Top fe Bit/Smiles	tribution Top features for positive c Bit/Smiles Feature Structure	tribution Top features for positive contribution Bit/Smiles Feature Structure Score

FCFP_12	-577707014		0.717	4 out of 4
		N−N [*]\N=C\[c]1:[cH]:[cH ]:[*]:[cH]:[cH]:1		
FCFP_12	-2100785893		0.552	7 out of 10
		[*]\N=C\[c](:[*]):[*]		
FCFP_12	-1151884458		0.385	1 out of 1
		[*]N[c](:n[*]):[c](: [*]).[t]		
	Ton Fea	atures for negative	contributio	 n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe
	267002002		1.24	In training set
FOFF_12	507 990000		-1.54	4 001 01 50
		 [*]:[cH]:[c](CI):[cH] :[*]		
FCFP_12	551850122		-1.27	2 out of 27
		[*][c]1:[*]:[cH]:[c](		

FCFP_12	-1508180856		-1.26	0 out of 7
		N		
		[*][c]1:[cH]:[cH]:[c] (CI):[cH]:[cH]:1		

# $C_{19}H_{17}N_7O_2$ Molecular Weight: 375.38397 ALogP: 4.197

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

## **Model Prediction**

#### Prediction: Mild

Probability: 0.00857

Enrichment: 0.0233

Bayesian Score: -16.3

Mahalanobis Distance: 7.75

### Mahalanobis Distance p-value: 0.871

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

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

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

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

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

## **Structural Similar Compounds**

Name	Pregna-1,4-diene-3,20- dione, 21-(acetyloxy)-11- hydroxy-6-methyl-17- (1- oxopropoxy)-, (6-alpha,11- beta)-	1,4-Pentadien-3-one, 1,5- bis(p-azidophenyl)-	Acetic acid, 2-(sec-butyl)- 4,6-dinitrophenyl ester
Structure	of of the		
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.730	0.808	0.809
Reference	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2- 5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1- 1972- Volume(issue) /page/year: 19,3103,1991	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,733,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,750,1986

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]

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

FCFP_12	-577707014	0 ₂ N	0.717	4 out of 4
		]:[*]:[cH]:[cH]:1		
FCFP_12	-2100785893	O_N C	0.552	7 out of 10
		[*]\N=C\[c](:[*]):[*]		
FCFP_12	-1151884458	O,N C	0.385	1 out of 1
		(*)N[c](:n:[*]):[c](: [*]):[*]		
	Top Fea	tures for negative	contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_12	8		-1.6	0 out of 11
		[*][N+](=[*])[*]		
FCFP_12	-828984032		-1.36	0 out of 8
		~ ~ _{N−N} [*][N+](=[*])[c](:[cH ]:[*]):[cH];[*]		

FCFP_12	1872392852	O ₂ N	-1.26	0 out of 7
		N _N		
		[*][N+](=O)[*]		



Rotatable Bonds: 5

Acceptors: 5

Donors: 3

## **Model Prediction**

Prediction: Non-Irritant

Probability: 0.968

Enrichment: 1.05

Bayesian Score: -1.11

Mahalanobis Distance: 10

#### Mahalanobis Distance p-value: 0.0583

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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-	1-Amino-2-bromo-4- hydroxyanthraquinone	8-Methylamino-4-hydroxy- 2-naphthalene sulfonic acid	
Structure	OHCI CI OHCI CI OHCI CI CI CI OH	HO W NH 2	HO WINH HO WINH OF OH O OH	
Actual Endpoint	Irritant	Non-Irritant	Non-Irritant	
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant	
Distance	0.796	0.840	0.844	
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72	28ZPAK -,190,72	

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

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

FCFP_12	4427049	Q	0.0734	5 out of 5
		[*][c](:[*]):n:n(:[*]		
FCFP_12	-796673622	);["]	0.0658	3 out of 3
		[*]C(=[*])NC1CC[*]CC1		
FCFP_12	1499521844	$\phi$	0.0658	3 out of 3
		[*]NC(=0)N[*]		
	Top Feat	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	$\bigcirc$	-0.65	0 out of 1
		[*]NN[c](:[*]):[*]		
FCFP_12	-1320007763		-0.0893	20 out of 24
		[*][c](:[*]):[ø]1:[cH ]:[cH]:[cH]:[c]:		
		1:[^]		

FCFP_12	1618154665	$\bigcirc$	-0.0845	412 out of 490
		]		



Molecular Weight: 361.40046 ALogP: 3.399 Rotatable Bonds: 5 Acceptors: 5 Donors: 3

## **Model Prediction**

Prediction: Non-Irritant

Probability: 0.576

Enrichment: 0.625

Bayesian Score: -4.34

Mahalanobis Distance: 7.03

#### Mahalanobis Distance p-value: 0.996

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

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

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

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

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

Structural Similar Compounds					
Name	5-Norbornene-2,3- dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	8-Methylamino-4-hydroxy- 2-naphthalene sulfonic acid	p-Acetophenetidide, 3'- (bis(2- hydroxyethyl)amino)-		
Structure	OHCI CI OHCI CI OHCI CI OH		HO OH		
Actual Endpoint	Irritant	Non-Irritant	Non-Irritant		
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant		
Distance	0.790	0.805	0.817		
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,190,72	28ZPAK -,100,72		

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

	Ton fe	atures for positive c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
				I

FCFP_12	4427049	Ç ,⊻+°	0.0734	5 out of 5
		[*][c](:[*]):n:n(:[*] ):[*]		
FCFP_12	1499521844		0.0658	3 out of 3
FCFP_12	-1578026142	[*]NC(=O)N[*]	0.0583	2 out of 2
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1838187238		-0.692	5 out of 12
ECEP 12	1204344583	[^]C(=[^])N[C]1:[CH]: [CH]:[*]:[CH]:[CH]:1	-0.65	0 out of 1
	1234344303	[*]NN[c](:[*]):[*]	0.00	

FCFP_12	-792685140	0	-0.65	0 out of 1
		[*]C(=[*])N[c] ¹ :[cH]: [cH]:[cH]:[cH]:[cH]:		



## **Model Prediction**

#### Prediction: Non-Irritant

Probability: 0.967

Enrichment: 1.05

Bayesian Score: -1.2

Mahalanobis Distance: 8.05

#### Mahalanobis Distance p-value: 0.87

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

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

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

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

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

Structural Similar Compounds				
Name	Benzenesulfonic acid, 2- anilino-5-nitro-	8-Methylamino-4-hydroxy- 2-naphthalene sulfonic acid	C.I. Fluorescent Brightening Agent 24	
Structure	ON OH OF STO		OH OH OH OH OH OH	
Actual Endpoint	Irritant	Non-Irritant	Irritant	
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant	
Distance	0.793	0.802	0.817	
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1061,1986	28ZPAK -,190,72	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

realure Co	nunbulion			
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
		÷	•	÷

FCFP_12	4427049	N _H S	0.0734	5 out of 5
		[*][c](:[*]):n:n(:[*] ):[*]		
FCFP_12	-1939253119	s z t	0.0658	3 out of 3
		[*]C(=[*])NCC		
FCFP_12	1499521844		0.0658	3 out of 3
		^{N→N} [*]NC(=O)N[*]		
	Top Feat	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583		-0.65	0 out of 1
		[*]NN[c](:[*]):[*]		
FCFP_12	-1320007763	[*][c](:[*]):[b] ^h :[cH ]:[cH]:[cH]:[*]:[c]: 1:[*]	-0.0893	20 out of 24
			1	

FCFP_12	1618154665	)	-0.0845	412 out of 490
		N S S		
		™™ [*]:[cH]:[cH]:[*		
		]		



## **Model Prediction**

Prediction: Non-Irritant

Probability: 0.968

Enrichment: 1.05

Bayesian Score: -1.16

Mahalanobis Distance: 10.1

#### Mahalanobis Distance p-value: 0.0523

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	Benzenesulfonic acid, 2,2'-(4,4'- biphenylylenedivinylene)d i-, disod ium salt	Urea, 1,3-bis(2- benzothiazolylthiomethyl) -	Benzenesulfonic acid, 2- anilino-5-nitro-
Structure		HN N HN N T	
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.858	0.910	0.911
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	AMIHBC AMA Archives of Industrial Hygiene and Occupational Medicine. (Chicago, IL) V.2-10, 1950-54. For publisher information, see AEHLAU. Volume(issue)/pag e/year: 5,311,1952	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1061,1986

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

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

	[*][c](:[*]):n:n(:[*] ):[*]		
-796673622		0.0658	3 out of 3
1499521844		0.0658	3 out of 3
Top Feat	tures for negative	contribution	
Bit/Smiles	Feature Structure	Score	Irritant in training set
1294344583		-0.65	0 out of 1
	[*]NN[c](:[*]):[*]		
-1320007763	[*][c](:[ [*] ]):[c]: ]:[cH]:[cH]:[cH]:[c]:	-0.0893	20 out of 24
-	-796673622 1499521844 1499521844 <b>Top Fea</b> <b>Bit/Smiles</b> 1294344583 1294344583	[*][c](:[*]):n:n:(:['])         -796673622 $\downarrow$	$["][c](:[1]):n:n(:[1]) :["]       0.0658         -796673622       \downarrow \downarrow 0.0658 \downarrow $

FCFP_12	1618154665	$\bigcirc$	-0.0845	412 out of 490
		Y ⊢ S ⊢ Z		
		]		



## **Model Prediction**

#### Prediction: Irritant

Probability: 0.974

Enrichment: 1.06

Bayesian Score: -0.717

Mahalanobis Distance: 8.11

#### Mahalanobis Distance p-value: 0.849

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	1-Piperazineacetic acid, 4- (2-hydroxyethyl)-alpha- phenyl-, 2,6-xyly I ester, monohydrochloride	Anthraquinone, 1,1'- iminodi-	Phosphorothioic acid, O- ethyl S-propyl O-(2,4,6- trichlorophenyl) es ter
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.674	0.675	0.707
Reference	BCFAAI Bollettino Chimico Farmaceutico. (Societa Editoriale Farmaceutica, Vi a Ausonio 12, 20123 Milan, Italy) V.33- 1894- Volume(issue)/page/year: 107,3 10,1968	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: OTS0535844

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

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

FCFP_12	-2100785893		0.081	11 out of 11
		[*]\N=C\[c](:[*]):[*]		
FCFP_12	-577707014		0.0734	5 out of 5
		(*)/N=C\{c]1:[cH]:[cH]:1		
FCFP_12	4427049		0.0734	5 out of 5
		[*][c](:[*]):n:n(:[*] ):[*]		
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	367998008	CI CI	-0.129	61 out of 76
		[*]:[cH]:[c](CI):[cH]		
FCFP 12	-1508180856	:[*]	-0.125	12 out of 15
		(CI):[CH]:[CH]:[CH]:1 (CI):[CH]:[CH]:1		
	<u> </u>	().[-, -]).[-, -], -		I

FCFP_12	71476542		-0.12	64 out of 79
		[*]:[c](:[*])Cl		



## **Model Prediction**

Prediction: Non-Irritant

Probability: 0.973

Enrichment: 1.06

Bayesian Score: -0.774

Mahalanobis Distance: 8.1

#### Mahalanobis Distance p-value: 0.855

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

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

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

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

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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	1-Piperazineacetic acid, 4- (2-hydroxyethyl)-alpha- phenyl-, 2,6-xyly I ester, monohydrochloride	s-Triazine, 2,4-dichloro-6- (o-chloroanilino)-	Anthraquinone, 1,1'- iminodi-
Structure			H H H H H H H H H H H H H H H H H H H
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.660	0.705	0.708
Reference	BCFAAI Bollettino Chimico Farmaceutico. (Societa Editoriale Farmaceutica, Vi a Ausonio 12, 20123 Milan, Italy) V.33- 1894- Volume(issue)/page/year: 107,3 10,1968	34ZIAG* -,235,69	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.

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

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

FCFP_12	-2100785893	F C	0.081	11 out of 11
		[*]\N=C\[c](:[*]):[*]		
FCFP_12	4427049	F CO	0.0734	5 out of 5
		[*][c](:[*]):n:n(:[*] ):[*]		
FCFP_12	-577707014	F C	0.0734	5 out of 5
		[*]\N=C\[c]1:[cH]:[cH ]:[*]:[cH]:[cH]:1		
	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	367998008		-0.129	61 out of 76
		[*]:[cH]:[c](CI):[cH] :[*]		
FCFP_12	-1508180856	ΓC, N.N.	-0.125	12 out of 15
		[*][c]1:[cH]:[cH]:[c] (CI):[cH]:[cH]:1		

FCFP_12	71476542	5	-0.12	64 out of 79
		N		
		[*]:[c](:[*])Cl		



Rotatable Bonds: 5

Acceptors: 6

## Donors: 2

## **Model Prediction**

Prediction: Non-Irritant

Probability: 0.966

Enrichment: 1.05

Bayesian Score: -1.24

Mahalanobis Distance: 8.51

#### Mahalanobis Distance p-value: 0.685

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

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

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

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

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

Structural Similar Compounds				
Name	5-Norbornene-2,3- dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-2-bromo-4- hydroxyanthraquinone	Phenol, 4,4'-sulfonyldi-	
Structure	OHCI CI OHCI CI OHCI CI OH	HO M Br NH 2	HO OFS OH	
Actual Endpoint	Irritant	Non-Irritant	Irritant	
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant	
Distance	0.786	0.802	0.809	
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72	BIOFX* BIOFAX Industrial Bio-Test Laboratories, Inc., Data Sheets. (1810 Fro ntage Rd., Northbrook, IL 60062) Volume(issue)/page/year: 601-05501,1974	

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

Top features for positive contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set			
FCFP_12	-2100785893	HO	0.081	11 out of 11			
----------------------------------------	-------------	--------------------------------------------	--------	-----------------------------	--	--	--
		N─N [*]\N=C\[c](:[*]):[*]					
FCFP_12	4427049	HO	0.0734	5 out of 5			
		[*][c](:[*]):n:n(:[*]					
FCFP 12	-577707014	):[*]	0.0734	5 out of 5			
1011_12	011101014	HONNN	0.0754				
		[*]\N=C\[c]1:[cH]:[cH ]:[*]:[cH]:[cH]:1					
Top Features for negative contribution							
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set			
FCFP_12	7	HO	-0.118	104 out of 128			
		[*]O					
FCFP_12	-549108873	HOLINN	-0.11	54 out of 66			
		[*]:[c](:[*])O					

FCFP_12	74595001	но	-0.11	54 out of 66
		∧ <del>-</del> ∧ [*]:[cH]:[c](O):[cH]:		
		[*]		



Rotatable Bonds: 6

Acceptors: 7

#### Donors: 1

# **Model Prediction**

#### **Prediction: Irritant**

Probability: 0.996

Enrichment: 1.08

Bayesian Score: 0.207

Mahalanobis Distance: 7.65

#### Mahalanobis Distance p-value: 0.955

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

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

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

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

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

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

Name	Pregna-1,4-diene-3,20- dione, 21-(acetyloxy)-11- hydroxy-6-methyl-17- (1- oxopropoxy)-, (6-alpha,11- beta)-	Acetic acid, 2-(sec-butyl)- 4,6-dinitrophenyl ester	1,4-Pentadien-3-one, 1,5- bis(p-azidophenyl)-	
Structure	Of Of Hand Hand			
Actual Endpoint	Irritant	Irritant	Irritant	
Predicted Endpoint	Irritant	Non-Irritant	Irritant	
Distance	0.735	0.809	0.810	
Reference	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2- 5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1- 1972- Volume(issue) /page/year: 19,3103,1991	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,750,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,733,1986	

# Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

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

2100785893	$\begin{array}{c} O_{2}N & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & $	0.0854 0.0843 0.081	27 out of 27 20 out of 20 11 out of 11
2100785893	[*][O-]	0.0843	20 out of 20
2100785893	0,N → N N N N N N [*][N+](=[*])[*]	0.0843	20 out of 20
2100785893	[*][N+](=[*])[*]	0.081	11 out of 11
2100785893	O ₂ N C	0.081	11 out of 11
	[*]\N=C\[c](:[*]):[*]		
Top Featur	res for negative o	contribution	
it/Smiles	Feature Structure	Score	Irritant in training set
320007763	^{0,N} ^N [*][c](:[*]):[c]*I![cH ]:[cH]:[cH]:[*]:[c]: 1:[*]	-0.0893	20 out of 24
618154665		-0.0845	412 out of 490
6	18154665	[*][c](:[*]):[c]*!:[cH       ]:[cH]:[cH]:[cH]:[c]:       1:[*]       18154665       O_N (N) (N) (N) (N) (N) (N) (N) (N) (N) (N	[*][c](:[*]):[c])*[!cH       ]:[cH]:[cH]:[cH]:[c]:       1:[*]       18154665       ON       N=N       N=N       [*]:[cH]:[cH]:[cH]:[*]

FCFP_12	16		-0.0843	423 out of 503
		N _N		
		[*]:[cH]:[*]		



Donors: 2

# **Model Prediction**

Prediction: Non-Irritant

Probability: 0.00137

Enrichment: 0.00149

Bayesian Score: -7.93

Mahalanobis Distance: 13.3

#### Mahalanobis Distance p-value: 8.82e-009

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

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

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

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

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

Structural	Similar	Compounds
	<b>U</b>	oompoundo

Structural Simila								
Name	2-Anthracenesulfonic acid, 1-amino-9,10- dihydro-9,10-dioxo-4- (2,4,6 -trimethylanilino)-, monosodium salt	2,2'-Stilbenedisulfonic acid, 4,4'-dinitro-	Benzenesulfonic acid, 2,2'-(4,4'- biphenylylenedivinylene)d i-, disod ium salt					
Structure		HO PH SEO						
Actual Endpoint	Irritant	Irritant	Irritant					
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant					
Distance	0.724	0.804	0.847					
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1062,1986	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973					

# Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

Feature Co	ntribution			
	Top fe	atures for positive o	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
			-	

-1410079687		0.0756	6 out of 6				
	۲ [*]S[cĵՔ٩ウᠯ[*]:[*]:n: 1:[*]						
4427049	[*][c](9 ⁽⁹ )) ^t n:n(:[*]	0.0734	5 out of 5				
-1539162406	).[ ] ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	0.0583	2 out of 2				
Top Features for negative contribution							
Bit/Smiles	Feature Structure	Score	Irritant in training set				
1175665944	NNNNN NHO S COOH [']CC(=0)N[c]:(:[cH]]:[	-1.02	2 out of 8				
-1838187238	"]):[cH]:[']	-0.692	5 out of 12				
	-1410079687 4427049 -1539162406 -1539162406 <b>Top Fea</b> <b>Bit/Smiles</b> 1175665944 -1838187238	1410079687 $\downarrow \downarrow $	-1410079687 $0.0756$ (1)S[0]FPRt[1]:[1]:n:       1:[1]         4427049 $0.0734$ 4427049 $0.0734$ (1)S09162406 $0.0583$ (1)S09162406 $0.0583$ (1)C[0]FPRt[1]:[1]:n:       0.0583         (1)C[0]FPRt[1]:[1]:n:       0.0583         (1)C[0]FPRt[1]:[1]:n:       0.0583         (1)C[0]FPRt[1]:[1]:n:       1:[1] <b>Top Features for negative contribution Bit/Smiles Feature Structure</b> Score         1175665944 $0.692$ $0.692$ $0.692$ (1838187238) $0.692$				

FCFP_12	-451043714		-0.65	0 out of 1
		cH]:[c]([*]):[cH]:[c H]:1		



# **Model Prediction**

Prediction: Non-Irritant

Probability: 0.717

Enrichment: 0.779

Bayesian Score: -3.88

Mahalanobis Distance: 8.89

#### Mahalanobis Distance p-value: 0.482

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

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

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

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

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

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

	Compounde		
Name	2,2'-Stilbenedisulfonic acid, 4,4'-dinitro-	2-Anthracenesulfonic acid, 1-amino-9,10- dihydro-9,10-dioxo-4- (2,4,6 -trimethylanilino)-, monosodium salt	Pregna-1,4-diene-3,20- dione, 21-(acetyloxy)-11- hydroxy-6-methyl-17- (1- oxopropoxy)-, (6-alpha,11- beta)-
Structure	HO P SZO	H 2 N 44 H 2 N	Of Contractions of the second
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.808	0.824	0.856
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1062,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2- 5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1- 1972- Volume(issue) /page/year: 19,3103,1991

# Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

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

5 8 -828984032	["][N+](c])[c](:[cH]['])[t]]	0.0854	27 out of 27 20 out of 20 9 out of 9
8 -828984032	[][U-]	0.0843	20 out of 20 9 out of 9
-828984032	[*][N+](=[*])[*]	0.0795	9 out of 9
Top Featu	ures for negative	contribution	
Bit/Smiles	Feature Structure	Score	Irritant in training set
1175665944	NO_ [']CC(=0)NIc](:[dH]:[ ]):[cH]:[']	-1.02	2 out of 8
-1838187238		-0.692	5 out of 12
	-1838187238	-1838187238	-1838187238

FCFP_12	-451043714	N-N N	-0.65	0 out of 1
		NH0 ^S , NN (1)CC(=0)N[6]1:[0H]:[ cH]:[0](1]):[cH]:[c		
		H]:1		



# **Model Prediction**

Prediction: Non-Irritant

Probability: 0.0519

Enrichment: 0.0563

Bayesian Score: -6.25

Mahalanobis Distance: 9.4

#### Mahalanobis Distance p-value: 0.229

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

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

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

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

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

# **Structural Similar Compounds**

Name	2,2'-Stilbenedisulfonic acid, 4,4'-dinitro-	Benzenesulfonic acid, 2,2'-(1,4- anthraquinonylenediimino )bis(5-meth yl-, disodium salt	2-Anthracenesulfonic acid, 1-amino-9,10- dihydro-9,10-dioxo-4- (2,4,6 -trimethylanilino)-, monosodium salt
Structure		H H H H H H H H H H H H H H H H H H H	H 2 New MR
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.748	0.764	0.825
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1062,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1330,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986

# Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

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

FCFP_12	-1410079687		0.0756	6 out of 6
		<i>「</i> [梦 <b>]</b> 名[む]1:n:[*]:[*]:n: 1:[*]		
FCFP_12	4427049		0.0734	5 out of 5
		):[*]		
FCFP_12	-1539162406	NHO ^S NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	0.0583	2 out of 2
		[*)℃fc]1:n:[*]:[*]:n: 1:[*]		
-	Top Fea	tures for negative	contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944		-1.02	2 out of 8
		H_NO_S [*]CC(=O)N[c](:[cH]:[ *]):[cH]:[*]		
FCFP_12	-1838187238		-0.692	5 out of 12
		/ ⁺ /℃(={[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1		

FCFP_12	-451043714		-0.65	0 out of 1
		[*]CC(=0)N[6] r ² [cH]:[ cH]:[c]((*]):[cH]:[c H]:1		

C₁₅H₁₅N₇OS Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

# **Model Prediction**

Prediction: Non-Irritant

Probability: 0.972

Enrichment: 1.06

Bayesian Score: -0.862

Mahalanobis Distance: 7.82

#### Mahalanobis Distance p-value: 0.926

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

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

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

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

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

Structural Similar Compounds						
Name	Benzenesulfonic acid, 2- anilino-5-nitro-	Benzenesulfonamide, 4- amino-N-(5,6-dimethoxy- 4-pyrimidinyl)-	Acetic acid, 2-(sec-butyl)- 4,6-dinitrophenyl ester			
Structure		H ₂ N OF N th H				
Actual Endpoint	Irritant	Irritant	Irritant			
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant			
Distance	0.716	0.727	0.728			
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1061,1986	FCTXAV 14,307,76	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague, Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,750,1986			

# Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

	l op te	atures for positive c	ontribution	
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
		•	-	

FCFP_12	-1410079687		0.0756	6 out of 6
		[*]S[c]1:n:[*]:[*]:n: 1:[*]		
FCFP_12	4427049		0.0734	5 out of 5
		[*][c](:[*]):n:n(:[*] ):[*]		
FCFP_12	-1578026142		0.0583	2 out of 2
		[*]:[c](:[*])CCC		
	Top Fea	tures for negative c	ontribution	ł
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	566058135		-0.367	13 out of 21
		[*]CC(=O)N[*]		
FCFP_12	-1320007763	[*][c](:[*]):[c]1:[cH ]:[cH]:[*][c]: 1:[*]	-0.0893	20 out of 24
		]:[cH]:[cH]:[*]:[c]: 1:[*]		

FCFP_12	1618154665		-0.0845	412 out of 490
		[*]:[cH]:[cH]:[* ]		

28ZPAK -,191,72



Acceptors: 12

Donors: 6

# **Model Prediction**

Prediction: Non-Irritant

Probability: 0.464

Enrichment: 0.503

Bayesian Score: -4.65

Mahalanobis Distance: 14.4

#### Mahalanobis Distance p-value: 5.01e-012

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

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

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

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

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

Structural Similar Compounds					
Name	1,3,6- Naphthalenetrisulfonic acid, 7-amino-	Amipurimycin, hydrate	2,2'-Benzidine disulfonic acid		
Structure		H ₂ N ⁴ H ₂ N ⁴ O O H O H O H O H	H ₂ N _{th} H ₂ N _{th} O O O O O H		
Actual Endpoint	Irritant	Irritant	Non-Irritant		
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant		
Distance	1.158	1.186	1.233		

# Model Applicability

Reference

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

JANTAJ Journal of

Assoc., 2-20-8 Ka

30.1.1977

Antibiotics. (Japan Antibiotics Research

miosaki, Shinagawa-ku,

1948-52; V.21- 1968-

Tokyo, 141, Japan) V.2-5,

Volume(i ssue)/page/year:

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

85JCAE "Prehled

Organicke Latky,"

Czechoslovakia,

Avicenum, 1986

-,1058,1986

Marhold, J., Praque,

Prumyslove Toxikologie;

Volume(issue)/page/year:

Top features for positive contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set	
			4		

FCFP_12	-415156552		0.0854	27 out of 27
		[*]CC(O)(C[*])C(=[*]) [*]		
FCFP_12	699500266		0.0852	25 out of 25
		[*]CC(O[*])O[*]		
FCFP_12	-1804899926		0.0795	9 out of 9
		[*]C([*])OC1CC([*])[* ]C([*])O1		
	Top Fea	tures for negative	e contribution	
Fingerprint	Bit/Smiles	Feature Structure	e Score	Irritant in training set
FCFP_12	979221427	HO HO O (*][C]1:[*\$EC[([]): [c]2C(=[*])[C]3:[C]( [*]):[*]:[CH]:[CH]:[ C]:3C(=O)[C]:2:[C]:1 O	-0.846	1 out of 4
FCFP_12	523781771	$ \begin{array}{c} HO \longrightarrow H^{2} \\ HO \longrightarrow H^{2} \\ HO \longrightarrow H^{2} \\ OH \longrightarrow H^{2} \\ OH \longrightarrow H^{2} \\ [^{*}][C]1:[^{*}]:[C]([^{0}]):O \\ [C]2C(=[^{*}])[^{*}]:[C](: \\ [^{*}])C(=O)[C]:2:[C]:1 \\ O \end{array} $	-0.732	3 out of 8
	I	1		1

FCFP_12	1392625818	NH ₂	-0.558	9 out of 18
		HO HO HO HO HO HO HO HO HO HO HO HO HO H		

# TOPKAT_Carcinogenic_Potency_TD50_Mouse



Molecular Weight: 367.4480 ALogP: 3.681 Rotatable Bonds: 5 Acceptors: 5 Donors: 3

# **Model Prediction**

Prediction: 8.1

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.2

#### Mahalanobis Distance p-value: 0.000779

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# Structural Similar Compounds

Name	542	Ochratoxin A	Phenolphthalein
Structure	AND Enuntioner		HO
Actual Endpoint (-log C)	4.79932	4.79932	2.43468
Predicted Endpoint (-log C)	3.6353	3.6353	3.66084
Distance	0.699	0.699	0.703
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: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]

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

ECFP_6	1559650422	(*)C[*]	0.203
ECFP_6	1333660716	[*][c](:[*]):[c](:[*]):[*]	0.0746
Fingernrint	Top Features f	or negative contributio	n Kaara
ECFP_6	1996767644		-0.251
ECFP_6	642810091	[*]:[cH]:[cH]:[c](:[* ]):[*]	-0.247
		[*]:[c](:[*]):[*]	
ECFP_6	-182236392	[*]:[cH]:[*]	-0.232

# C₁₉H₁₉N₇O Molecular Weight: 361.40046

ALogP: 3.399 Rotatable Bonds: 5

Acceptors: 5

Donors: 3

# **Model Prediction**

Prediction: 7.62

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.3

#### Mahalanobis Distance p-value: 0.000436

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, 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	542
Structure	HO		AND Enantioner
Actual Endpoint (-log C)	2.43468	0.937339	4.79932
Predicted Endpoint (-log C)	3.66084	3.26294	3.6353
Distance	0.751	0.813	0.818
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: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]

FingerprintBit/SmilesFeature StructureScoreECFP_6655739385Image: Construction of the structure0.229	Top features for positive contribution					
ECFP_6 655739385 0.229	gerprint	Bit/Smiles	Feature Structure	Score		
N     O       H     N       N     N       N     N       N     N       [*]:n:[*]	FP_6	655739385	[*]:n:[*]	0.229		

ECFP_6	1559650422	[*]C[*]	0.203
	1333660716	[*][c](:[*]):[c](:[*]):[*]	0.0746
Fingerprint	Top Features f	or negative contributio	n Fooro
ECFP_6	1996767644	[*]:[cH]:[c](:[* ]):[*]	-0.251
ECFP_6	642810091	[*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	[*]:[cH]:[*]	-0.232

# TOPKAT_Carcinogenic_Potency_TD50_Mouse

# $C_{15}H_{19}N_7S$ Molecular Weight: 329.42325 ALogP: 3.388 Rotatable Bonds: 7

Acceptors: 5 Donors: 3

# **Model Prediction**

Prediction: 18.5

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.3

#### Mahalanobis Distance p-value: 0.000422

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# **Structural Similar Compounds**

Name	4-Chloro-6-(2,3-xylidino)- 2-pyri-mi-dinylthio(N-b- hydroxy-ethyl) acetamide	542	Ochratoxin A
Structure	CI NH HO	AND Exantomer	
Actual Endpoint (-log C)	3.91517	4.79932	4.79932
Predicted Endpoint (-log C)	3.92186	3.6353	3.6353
Distance	0.669	0.683	0.683
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: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 4. Unknown ECFP_2 feature: 150794520: [*]NC(=S)N[*]

	Top features for positive contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	(*]:n:[*]	0.229	

ECFP_6	1559650422	[*]C[*]	0.203
ECFP_6	1333660716	[*][c](:[*]):[c](:[*]):[*]	0.0746
Eingerprint	Top Features f	or negative contributio	n Fooro
	10067676 <i>11</i>		
		[*]:[cH]:[c](:[* ]):[*]	0.201
ECFP_6	642810091	[*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	[*]:[cH]:[*]	-0.232



Acceptors: 5 Donors: 3

# **Model Prediction**

Prediction: 6.43

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 12.6

#### Mahalanobis Distance p-value: 9.23e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# Structural Similar Compounds

Name	542	Ochratoxin A	4-Chloro-6-(2,3-xylidino)- 2-pyri-mi-dinylthio(N-b- hydroxy-ethyl) acetamide
Structure	AND Exantomer	OH OH HO HO OH OH OH OH OH OH OH OH OH O	CI N N N N HO
Actual Endpoint (-log C)	4.79932	4.79932	3.91517
Predicted Endpoint (-log C)	3.6353	3.6353	3.92186
Distance	0.716	0.716	0.764
Reference	CPDB	CPDB	CPDB

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 4. Unknown ECFP_2 feature: 150794520: [*]NC(=S)N[*]

Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	Image: Constraint of the second se	0.229	

ECFP_6	1559650422	Image: Signature     Image: Signature <th>0.203</th>	0.203
ECFP_6	1333660716	[*][c](:[*]):[c](:[*]):[*]	0.0746
Fingernrint	Top Features f	or negative contributio	n Faara
	1006767611		
		[*]:[cH]:[c](:[*	
ECFP_6	642810091	[*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	[*]:[cH]:[*]	-0.232

NH Rotatable Bonds: 5

Acceptors: 5 Donors: 1

# **Model Prediction**

Prediction: 3.41

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13.3

#### Mahalanobis Distance p-value: 1.93e-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

	Compounds		
Name	646	Phenolphthalein	C.I. pigment red 3
Structure		HO	
Actual Endpoint (-log C)	0.937339	2.43468	0.937339
Predicted Endpoint (-log C)	3.26294	3.66084	3.17837
Distance	0.726	0.754	0.847
Reference	CPDB	CPDB	CPDB

# Model Applicability

- OPS PC24 out of range. Value: -4.6354. Training min, max, SD, explained variance: -4.4826, 1. 3.8729, 1.034, 0.0133.
- OPS PC26 out of range. Value: -3.6943. Training min, max, SD, explained variance: -2.9667, 2. 3.5042, 1.009, 0.0127.
- Unknown ECFP_2 feature: -1507082173: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*] 3.
- Unknown ECFP 2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*] 4.

Feature Contribution				
	Top features for positive contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385		0.229	



ECFP_6	1559650422		0.203
ECFP_6	-1925046727		0.145
	Top Features f	or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	C' NNN NH NH NH NN NH NN NH NH NH NH NH NH	-0.251
ECFP_6	642810091	CI	-0.247
ECFP_6	-182236392		-0.232

F  $N_{NH}$   $N_{NH}$ N

Acceptors: 5

Donors: 1

# **Model Prediction**

Prediction: 4.41

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.7

#### Mahalanobis Distance p-value: 8.38e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# **Structural Similar Compounds**

	Compoundo		
Name	646	Phenolphthalein	C.I. pigment red 3
Structure		HO	
Actual Endpoint (-log C)	0.937339	2.43468	0.937339
Predicted Endpoint (-log C)	3.26294	3.66084	3.17837
Distance	0.726	0.744	0.840
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: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]

			Score
ECFP_6	655739385	F N N N N N N N N N N N N N	0.229
ECFP_6	1559650422		0.203
-------------	----------------	----------------------------	--------
ECFP_6	-1925046727		0.145
	Top Features f	or negative contributio	n b
Fingerprint	Bit/Smiles	Feature Structure	Score
	1990707044	[*]:[cH]:[c](:[* ]):[*]	-0.251
ECFP_6	642810091	[*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	[*]:[cH]:[*]	-0.232

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

# **Model Prediction**

Prediction: 10.5

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 12.6

### Mahalanobis Distance p-value: 1.12e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### **Structural Similar Compounds**

Name	Phenolphthalein	646	C.I. pigment red 3
Structure	HO		
Actual Endpoint (-log C)	2.43468	0.937339	0.937339
Predicted Endpoint (-log C)	3.66084	3.26294	3.17837
Distance	0.706	0.725	0.842
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: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]

ingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	HO N N N N N N N N N N N N N N N N N N N	0.229

ECFP_6	1559650422		0.203
ECFP_6	-1925046727		0.145
	Top Features f	or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_0	2019062761	H ^O N N N N N N N N N N N N N	-0.258
ECFP_6	1996767644	HO NN [*]:[cH]:[cH]:[c](:[* ]):[*]	-0.251
ECFP_6	642810091	HO N N N N N N N N N N N N N N N N N N N	-0.247

 $C_{19}H_{17}N_7O_2$ Molecular Weight: 375.38397

ALogP: 4.197

Rotatable Bonds: 6 Acceptors: 7

Donors: 1

# **Model Prediction**

Prediction: 5.83

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.8

### Mahalanobis Distance p-value: 4.31e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### **Structural Similar Compounds**

Name	646	Azathioprine s	470
Structure			
Actual Endpoint (-log C)	0.937339	4.49253	4.62839
Predicted Endpoint (-log C)	3.26294	4.28607	3.93264
Distance	0.738	0.827	0.842
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: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]

0 1	DIVSIIIIes	Feature Structure	Score
ECFP_6	655739385	C ₂ N N N N N (*]:n:[*]	0.229

ECFP_6	1559650422		0.203
ECFP_6	-1925046727	[*]C[*]	0.145
	Top Features f	or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	C ₂ N N N N N N N N N N N N N N N N N N N	-0.251
ECFP_6	642810091	C ₂ N N N N N N N N N N N N N N N N N N N	-0.247
ECFP_6	-182236392	O_N N N N N N N N N N N N N N	-0.232



Acceptors: 8 Donors: 2

# **Model Prediction**

Prediction: 8.9

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 12.3

### Mahalanobis Distance p-value: 5.52e-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	223	420	Salicylazosulfapyridine		
Structure		HO NA NA	HN HO HO O		
Actual Endpoint (-log C)	5.08368	2.78302	2.5034		
Predicted Endpoint (-log C)	5.08273	3.31546	3.54214		
Distance	0.966	0.993	0.997		
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: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 4. Unknown ECFP_2 feature: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	NHO COOH [*]:n:[*]	0.229	

ECFP_6	1559650422	NHO COOH [*]C[*]	0.203
ECFP_6	-175146122	[*]C(=[ ⁺ ]):[cH]:[ [*] ]	0.107
	Top Features	for negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*]:[cH ^{9:} f ^e H]:[c](:[* ]):[*]	-0.251
ECFP_6	642810091	NHO COOH [*]:[C](:[*]):[*]	-0.247
ECFP_6	-182236392	лн ^о Соон [*]:[cH]:[*]	-0.232



# **Model Prediction**

Prediction: 8.57

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.7

### Mahalanobis Distance p-value: 9.42e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### Structural Similar Compounds

Name	223	Azathioprine s	420
Structure	All Exerting		HO NA
Actual Endpoint (-log C)	5.08368	4.49253	2.78302
Predicted Endpoint (-log C)	5.08273	4.28607	3.31546
Distance	0.968	0.975	0.979
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: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 4. Unknown ECFP_2 feature: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	NH ^O S [*]:n:[*]	0.229	

ECFP_6	1559650422	NH ^O ₀ [*]C[*]	0.203
ECFP_6	781519895		0.0797
	Top Features	for negative contribution	n
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*]:[cH]![c](:[* ]):[*]	-0.251
ECFP_6	642810091	NH ^O [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	NH ^O [*]:[CH]:[*]	-0.232

 $C_{21}H_{20}N_8O_3S_2$ Molecular Weight: 496.56529 ALogP: 2.496 Rotatable Bonds: 7 Acceptors: 8

Donors: 2

# **Model Prediction**

Prediction: 5.54

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 12.4

### Mahalanobis Distance p-value: 2.36e-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	420	Salicylazosulfapyridine	Azathioprine s
Structure	HO NA NA	HN HO HO HO HO HO HO HO HO HO HO HO HO HO	
Actual Endpoint (-log C)	2.78302	2.5034	4.49253
Predicted Endpoint (-log C)	3.31546	3.54214	4.28607
Distance	0.936	1.012	1.033
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: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 4. Unknown ECFP_2 feature: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	NH ^O S SO_NH ₂ [*]:n:[*]	0.229	

ECFP_6	1572579716	NHO S NN SO 2NH2 [*]N	0.225
ECFP_6	1559650422	NHO SO NHO [*]C[*]	0.203
	Top Features f	or negative contributior	1
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[ ^M ]?[cH]:[c](:[* ]):[*]	-0.251
ECFP_6	642810091	NH ^o [*]:[C](:[*]):[*]	-0.247
ECFP_6	-182236392	NH ^N O SO ,NH, [*]:[cH]:[*]	-0.232

 $C_{15}H_{15}N_{7}OS$ 

Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

# **Model Prediction**

Prediction: 14.9

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 10.7

### Mahalanobis Distance p-value: 0.00395

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### **Structural Similar Compounds**

Name	Azathioprine s	646	470
Structure			
Actual Endpoint (-log C)	4.49253	0.937339	4.62839
Predicted Endpoint (-log C)	4.28607	3.26294	3.93264
Distance	0.724	0.810	0.814
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: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 3. Unknown ECFP_2 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 4. Unknown ECFP_2 feature: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	O H ₂ N S N N N N N N N N N N N N N N N N N N	0.229	
		·		

ECFP_6	1572579716		0.225
ECFP_6	1559650422		0.203
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	[*]:[cH]:[cH]:[c](:[* ]):[*]	-0.251
ECFP_6	642810091	°, s, N-N H₂N s, N-N N-N [*]:[c](:[*]):[*]	-0.247
ECFP_6	-182236392	°, , , , , , , , , , , , , , , , , , ,	-0.232

# Doxorubicin



Donors: 6

# **Model Prediction**

Prediction: 11.2

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 16.6

### Mahalanobis Distance p-value: 3.86e-018

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### Structural Similar Compounds

Name	532	338	380
Structure	All Environment of the second secon	H H H H H H H H H H H H H H H H H H H	
Actual Endpoint (-log C)	4.38903	4.39533	4.48977
Predicted Endpoint (-log C)	5.60554	4.31268	6.22716
Distance	0.909	0.937	0.952
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: 140080459: [*]CC(O)(C[*])C(=[*])[*]
- 3. Unknown ECFP_2 feature: 456242574: [*]C([*])OC([*])[*]
- 4. Unknown ECFP_2 feature: -1783281539: [*]CC(=O)C([*])([*])[*]
- 5. Unknown ECFP_2 feature: -1907393688: [*]C(=[*])CO
- 6. Unknown ECFP_2 feature: 1535429263: [*]OC(C)C([*])[*]
- 7. Unknown ECFP_2 feature: 1201786014: [*]CC(N)C([*])[*]
- 8. Unknown ECFP_2 feature: -932844120: [*]C([*])N
- 9. Unknown ECFP_2 feature: -2061744983: [*]CC(O[*])O[*]

# Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

ECFP_6	1572579716		0.225
ECFP_6	1559650422		0.203
ECFP_6	683445015		0.136
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	$HO \rightarrow H^{NH_2}$ $HO \rightarrow H^{O} $	-0.275
ECFP_6	2019062761		-0.258

ECFP_6	1996767644	NHA	-0.251
		HO HO OH OH OH OH OH OH OH OH OH OH OH O	
		1)+[ ]	

# TOPKAT_Carcinogenic_Potency_TD50_Rat



Molecular Weight: 367.44809 ALogP: 3.681 Rotatable Bonds: 5 Acceptors: 5 Donors: 3

### **Model Prediction**

Prediction: 0.651

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 14.7

### Mahalanobis Distance p-value: 2.78e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### Structural Similar Compounds

Name	542	Ochratoxin A	Fluvastatin
Structure	AND Exantomer		
Actual Endpoint (-log C)	6.59334	6.47264	3.51742
Predicted Endpoint (-log C)	5.06501	5.06501	5.41573
Distance	0.659	0.659	0.675
Reference	CPDB	CPDB	CPDB

### Model Applicability

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

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

	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
-CFP_6	-1043250487	[*]CC(C[*])N[*]	1.15	

FCFP_6 FCFP_6	1 -885550502	[*] = O	0.234 0.229
	Ton Features for n	$\begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ $	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1175638033	[*]C1[*]CCCC1	-0.512
FCFP_6	991735244	[*]:[c]1:[']:[cH]:[cH]:1	-0.422
FCFP_6	16	[*]:[cH]:[*]	-0.354

# TOPKAT_Carcinogenic_Potency_TD50_Rat



Rotatable Bonds: 5

Acceptors: 5

Donors: 3

# **Model Prediction**

Prediction: 19.9

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 14.7

### Mahalanobis Distance p-value: 2.53e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Name	Fluvastatin	913	4-(2-Hydroxyethylamino)- 2-(5-nitro-2- thienyl)quinazoline
Structure	F HO HO	C C C C C C C C C C C C C C C C C C C	
Actual Endpoint (-log C)	3.51742	3.51742	5.22831
Predicted Endpoint (-log C)	5.41573	5.41573	4.31976
Distance	0.716	0.716	0.726
Reference	CPDB	CPDB	CPDB

# Model Applicability

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

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

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	[*]=O	0.234	

FCFP_6	307419094	[*][c](:[*]):[c](:[*]):[*]	0.121
FCFP_6	-1272798659	CCC	0.11
	Top Features fo	r negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.422
FCFP_6	-2093839777	[*][c]1:[cH]:[cH]:1	-0.378
FCFP_6	16	[*]:[cH]:[*]	-0.354

# TOPKAT_Carcinogenic_Potency_TD50_Rat



### Donors: 3

### **Model Prediction**

Prediction: 34.6

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 12.8

### Mahalanobis Distance p-value: 1.55e-005

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### **Structural Similar Compounds**

Name	4-(2-Hydroxyethylamino)- 2-(5-nitro-2- thienyl)quinazoline	4-Chloro-6-(2,3-xylidino)- 2-pyri-mi-dinylthio(N-b- hydroxy-ethyl) acetamide	Fluvastatin	
Structure	CO-R S N HO	S NH HO		
Actual Endpoint (-log C)	5.22831	4.75226	3.51742	
Predicted Endpoint (-log C)	4.31976	3.29421	5.41573	
Distance	0.623	0.629	0.638	
Reference	CPDB	CPDB	CPDB	

### Model Applicability

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

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

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	⁰ ² ² [⊥] ² ^{−²^{−²^{−² = 0 [*]=0}}}	0.234	

FCFP_6	-885550502	[*]C([*])NC(=[*])[*]	0.229
FCFP_6	307419094	[*][c](:[*]):[c](:[*])	0.121
Fingerprint	Top Features fo	or negative contributio	n Scoro
FCFP_6	-1272709286		-0.526
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.422
FCFP_6	16	[*]:[cH]:[*]	-0.354



ALogP: 4.898 Rotatable Bonds: 7 Acceptors: 5 Donors: 3

### **Model Prediction**

Prediction: 0.873

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 14.4

### Mahalanobis Distance p-value: 1.34e-009

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### Structural Similar Compounds

Name	913	Fluvastatin	542
Structure	C C C C C C C C C C C C C C C C C C C		AND Exantomer
Actual Endpoint (-log C)	3.51742	3.51742	6.59334
Predicted Endpoint (-log C)	5.41573	5.41573	5.06501
Distance	0.618	0.618	0.671
Reference	CPDB	CPDB	CPDB

### Model Applicability

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

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

Feature Cont	ribution			
	Top features f	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
=CFP_6	-1043250487	[*]CC(C[*])N[*]	1.15	
		[*]CC(C[*])N[*]		

FCFP_6 FCFP_6	-885550502	[*]=O	0.234
		[*]C([*])NC(=[*])[*]	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1175638033	[*]C1[*]CCCC1	-0.512
FCFP_6	991735244	[*]:[c]1:[']:[cH]:[cH]:1	-0.422
FCFP_6	16	[*]:[cH]:[*]	-0.354



Acceptors: 5

Donors: 1

# **Model Prediction**

Prediction: 2.51

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13.1

### Mahalanobis Distance p-value: 3.12e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### **Structural Similar Compounds**

Name	Indomethacin	646	Omeprazole	
Structure				
Actual Endpoint (-log C)	5.49293	2.41938	3.4628	
Predicted Endpoint (-log C)	4.9569	3.77987	4.7324	
Distance	0.621	0.680	0.711	
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.

#
FCFP_6	32		0.154
		C' NN NN NN NN NN NN NN NN NN N	0.107
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	C' N N N N N N N N N N N N N	-0.422
FCFP_6	16	CI N, N N, N N	-0.354
FCFP_6	17	CI NNN NNN NNN NNN NNN NNN NNN NNN NNN N	-0.149



Acceptors: 5

Donors: 1

# **Model Prediction**

Prediction: 2.67

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13.2

### Mahalanobis Distance p-value: 2.07e-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**

Namo		646	Pomitradina
Structure			
Actual Endpoint (-log C)	5.49293	2.41938	2.71351
Predicted Endpoint (-log C)	4.9569	3.77987	4.65043
Distance	0.616	0.681	0.688
Reference	CPDB	CPDB	CPDB

# Model Applicability

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

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

	Top features	for positive contributio	n	
[;] ingerprint	Bit/Smiles	Feature Structure	Score	
CFP_6	1		0.234	

FCFP_6	32		0.154
FCFP_6	203677720	[*]=C[c](:[cH]:[*]):[ cH]:[*]	0.137
Eingorprint	Top Features fo	or negative contributio	n Scoro
FCFP_6	991735244		-0.422
FCFP_6	16	[]][CH]:[CH]:1	-0.354
FCFP_6	17	[^]:[CH]:[^]	-0.149

ALogP: 4.061 Rotatable Bonds: 5

Acceptors: 6

Donors: 2

# **Model Prediction**

Prediction: 20.5

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13.1

### Mahalanobis Distance p-value: 2.81e-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	Phenolphthalein	646	Omeprazole
Structure	HO		
Actual Endpoint (-log C)	2.54766	2.41938	3.4628
Predicted Endpoint (-log C)	3.7508	3.77987	4.7324
Distance	0.671	0.680	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. OPS PC7 out of range. Value: 6.5394. Training min, max, SD, explained variance: -5.0422, 6.1749, 1.868, 0.0335.

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1		0.234

FCFP_6	203677720	HO NN N [*]=C[c](:[cH]:[*]):[ cH]:[*]	0.137
FCFP_6	307419094	HO N N N N N N N N N N N N N	0.121
Eingorprint	Top Features	for negative contributio	n Scoro
FCFP_6	991735244	^{HO} ^N ^N ^N ^N ^N ^N ^N ^N	-0.422
FCFP_6	7		-0.372
FCFP_6	16	HO N N N N N N N N N N N N N N N N N N N	-0.354



Rotatable Bonds: 6

Acceptors: 7

### Donors: 1

# **Model Prediction**

Prediction: 1.99

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13.6

### Mahalanobis Distance p-value: 1.67e-007

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# **Structural Similar Compounds**

Name	C.I. direct brown 95	4-Morpholino-2-(5-nitro-2- thi-enyl)quinazoline	4-Bis(2- hydroxyethyl)amino-2-(5- nitro-2-thienyl)quinazoline
Structure			
Actual Endpoint (-log C)	5.31387	4.83293	5.05984
Predicted Endpoint (-log C)	4.30266	4.97658	4.23808
Distance	0.677	0.691	0.707
Reference	CPDB	CPDB	СРОВ

# Model Applicability

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

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

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	5		0.431	

FCFP_6	8	0 ₂ N	0.336
		N H	
		[*][N+](=[*])[*]	
FCFP_6	1		0.234
		→ 1 ^N N	
		N N	
		[*]=O	
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244		-0.422
		ĭ _{№−Ň} [*]:[c]1:[*]:[cH]:[cH	
		]:[cH]:[cH]:1	
FCFP_6	16		-0.354
		[*]:[cH]:[*]	
FCFP_6	1872392852	0 ₂ N	-0.155
		[*][N+](=O)[*]	

# $C_{22}H_{19}N_7O_3S$ Molecular Weight: 461.49635 ALogP: 3.42 Rotatable Bonds: 7 Acceptors: 8

# Donors: 2

# **Model Prediction**

Prediction: 25.2

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 17.2

### Mahalanobis Distance p-value: 2.58e-018

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# **Structural Similar Compounds**

Name	C.I. direct brown 95	4-Bis(2- hydroxyethyl)amino-2-(5- nitro-2-thienyl)quinazoline	623
Structure			Na "
Actual Endpoint (-log C)	5.31387	5.05984	2.39985
Predicted Endpoint (-log C)	4.30266	4.23808	3.4177
Distance	0.827	0.833	0.840
Reference	CPDB	CPDB	СРОВ

# Model Applicability

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

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

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1	NHOSSIN COOH [*]=O	0.234	

FCFP_6	203677720	[*]=C[c](![tH]:[*]):[ cH]:[*]	0.137
FCFP_6	-1272768868	NHO SCC(=[*])[*]	0.127
	Top Features	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c]199![cH]:[cH]:1	-0.422
FCFP_6	7	NH ^S COOH [*]O	-0.372
FCFP_6	16	NH ^O COOH [*]:[cH]:[*]	-0.354



# **Model Prediction**

Prediction: 5.43

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 17

### Mahalanobis Distance p-value: 1.16e-017

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

## Structural Similar Compounds

Name	623	C.I. direct brown 95	4-Bis(2- hydroxyethyl)amino-2-(5- nitro-2-thienyl)quinazoline
Structure	NH HN HN HN HN HN HN HN HN HN		
Actual Endpoint (-log C)	2.39985	5.31387	5.05984
Predicted Endpoint (-log C)	3.4177	4.30266	4.23808
Distance	0.803	0.829	0.843
Reference	CPDB	CPDB	CPDB

# Model Applicability

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

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

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	5		0.431	

FCFP_6	8	[*][N+](=[*])[*]	0.336
FCFP_6	1	$\sum_{\substack{N \in \mathcal{O}_{2}}^{N \in \mathcal{O}_{2}}} \sum_{\substack{N \in \mathcal{O}_{2}}^{N \in \mathcal{O}_{2}}} \sum_{\substack{N \in \mathcal{O}_{2}^{N} \in \mathcal{O}_{2}^{N}}} \sum_{\substack{N \in \mathcal{O}_{2}^{N}}} \sum_{\substack{N \in \mathcal{O}_{2}^{N} \in \mathcal{O}_{2}^{N}}} $	0.234
	Top Features for ne	egative contribution	<b>b</b>
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c] ^{1:P^x} ]:[cH]:[cH]:1	-0.422
FCFP_6	16	NHO [*]:[cH]:[*]	-0.354
FCFP_6	590925877	[*]N[c][%[cH]:[*]):[c H]:[*]	-0.323



# **Model Prediction**

Prediction: 22

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 15.7

### Mahalanobis Distance p-value: 1.78e-013

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# Structural Similar Compounds

Name	623	4-Bis(2- hydroxyethyl)amino-2-(5- nitro-2-thienyl)quinazoline	418
Structure	NH HN HN HN HN HN HN HN HN HN		HO NA NA NA NA NA NA NA NA NA
Actual Endpoint (-log C)	2.39985	5.05984	2.9349
Predicted Endpoint (-log C)	3.4177	4.23808	3.45907
Distance	0.800	0.870	0.870
Reference	CPDB	CPDB	СРДВ

# Model Applicability

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

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

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1		0.234	

FCFP_6	203677720	[*]≝C[c](:[cH]:[*]):[ cH]:[*]	0.137
FCFP_6	-1272768868	[*]SCC(=[*])[*]	0.127
	Top Features f	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[ ^M ]:[cH]:[cH]:1	-0.422
FCFP_6	16		-0.354
FCFP_6	590925877	[*]N[ê](:[cH]:[*]):[c H]:[*]	-0.323

 $\begin{array}{c|c}
 O & N \\
 H_2 N \\
 S & N \\
 N \\$ 

Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

# **Model Prediction**

Prediction: 23.3

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 14.5

### Mahalanobis Distance p-value: 1.12e-009

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# **Structural Similar Compounds**

Name	4-(2-Hydroxyethylamino)- 2-(5-nitro-2- thienyl)quinazoline	4-Morpholino-2-(5-nitro-2- thi-enyl)quinazoline	4-Bis(2- hydroxyethyl)amino-2-(5- nitro-2-thienyl)quinazoline
Structure			
Actual Endpoint (-log C)	5.22831	4.83293	5.05984
Predicted Endpoint (-log C)	4.31976	4.97658	4.23808
Distance	0.662	0.675	0.678
Reference	CPDB	CPDB	CPDB

# Model Applicability

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

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

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_6	1		0.234	

FCFP_6	-1272768868	[*]SCC(=[*])[*]	0.127
FCFP_6	307419094	[*][c](:[*]):[c](:[cH ]:[*]):[c](:[*]):[*]	0.121
	Top Features f	or negative contribution	n baan
Fingerprint	Bit/Smiles	Feature Structure	Score
г <b>сгг_</b> б	991733244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.422
FCFP_6	16	⁰ H ₂ N s N N N N N N N N N N	-0.354
FCFP_6	566058135	[*]CC(=O)N[*]	-0.182

# Doxorubicin



Rotatable Bonds: 5

Acceptors: 12

Donors: 6

# **Model Prediction**

Prediction: 0.861

Unit: mg/kg_body_weight/day

Mahalanobis Distance: 17.4

### Mahalanobis Distance p-value: 6.65e-019

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# Structural Similar Compounds

Name	424	D & C red no. 5	Salicylazosulfapyridine
Structure		HO HO HN Na	HN N HN N HO HO HO HO
Actual Endpoint (-log C)	4.70022	3.04419	2.39891
Predicted Endpoint (-log C)	5.67571	4.52866	3.17598
Distance	1.039	1.056	1.057
Reference	CPDB	CPDB	CPDB

# Model Applicability

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

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

Top features for positive contribution					
-ingerprint	Bit/Smiles	Feature Structure	Score		
-CFP_6	-1043250487	$HO \rightarrow OH O OH O OH O OH O OH O OH O OH O$	1.15		

FCFP_6	136627117	HO HO O OH OH OH OH OH OH OH OH OH OH OH	0.69
FCFP_6	565968762	$HO \rightarrow OH $	0.266
	Top Features for ne	egative contribution	<b>b</b>
Fingerprint	Bit/Smiles	Feature Structure	Score
	-1272709286		-0.526
FCFP_6	991735244	HO HO OH OH OH OH OH OH OH OH OH OH OH O	-0.422
FCFP_6	7		-0.372



Prediction: 0.118

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

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: -5.5208. Training min, max, SD, explained variance: -4.7991, 6.1674, 1.831, 0.0147.
- 2. OPS PC31 out of range. Value: 6.0969. Training min, max, SD, explained variance: -4.0208, 4.2529, 1.246, 0.0068.
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 5. 6.

7. 8.

- Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 9. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 10. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 12. Unknown ECFP_6 feature: 835630791: [*][c](:[*]):n:n(:[*]):[*]
- 13. Unknown ECFP_6 feature: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 14. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -1238415266: [*]NN[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 17. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:(*]:n:n:1
- 18. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 19. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 20. Unknown ECFP_6 feature: 1635339976: [*]NNC(=[*])[*]
- 21. Unknown ECFP_6 feature: -649580166: [*]NC(=O)N[*]
- 22. Unknown ECFP_6 feature: -2091181441: [*]C([*])NC(=[*])[*]
- 23. Unknown ECFP_6 feature: -859078569: [*]CC(C[*])N[*]
- 24. Unknown ECFP_6 feature: -1332781180: [*]CCC[*]

I op features for positive contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	-167460056	(*)C([*])[*]	0.136
ECFP_6	1559650422		0.129
FCFP_6	3	$ \bigvee_{\substack{Z \in \mathbb{Z}^{n} \\ Z \in \mathbb{Z}^{n} $	0.0924
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	(*]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	[*]:[cH]:[cH]:1	-0.133

FCFP_6	1	$\square \square$	-0.102
		[*]=O	



Prediction: 0.142

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

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

- 1. OPS PC31 out of range. Value: 5.2395. Training min, max, SD, explained variance: -4.0208, 4.2529, 1.246, 0.0068.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 4. 5.

6.

- Unknown ECFP 6 feature: 672362763: [*]:n(:[*]):[*]
- 9. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 10. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 12. Unknown ECFP_6 feature: 835630791: [*][c](:[*]):n:n(:[*]):[*]
- 13. Unknown ECFP_6 feature: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 14. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -1238415266: [*]NN[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 17. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- 18. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 19. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 20. Unknown ECFP_6 feature: 1635339976: [*]NNC(=[*])[*]
- 21. Unknown ECFP_6 feature: -649580166: [*]NC(=O)N[*]
- 22. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]

Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	1559650422	$ \begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & & \\ & & & \\ & & & & \\ & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\$	0.129	

FCFP_6 FCFP_6	3 -2093839777	<pre></pre>	0.0924 0.078
	Top Features for ne	C ¹ [cH]:[cH]:1 contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	(*):[cH]:[cH]:1	-0.134
ECFP_6	1564392544	[*]:[cH]:[cH]:1	-0.133
FCFP_6	1		-0.102



Prediction: 0.0717

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

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

- 1. OPS PC31 out of range. Value: 4.987. Training min, max, SD, explained variance: -4.0208, 4.2529, 1.246, 0.0068.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 4. 5.

6. 7. 8.

- Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 9. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 10. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 12. Unknown ECFP_6 feature: 835630791: [*][c](:[*]):n:n(:[*]):[*]
- 13. Unknown ECFP_6 feature: -1507082173: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 14. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -1238415266: [*]NN[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 17. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- 18. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 19. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 20. Unknown ECFP_6 feature: 1635339976: [*]NNC(=[*])[*]
- 21. Unknown ECFP_6 feature: 150794520: [*]NC(=S)N[*]
- 22. Unknown ECFP_6 feature: 1979182050: [*]C(=S)[*]
- 23. Unknown ECFP_6 feature: 497523368: [*]CNC(=[*])[*]
- 24. Unknown ECFP_6 feature: -950223878: [*]NCC

Feature Cont	ribution			
	Top features for positive contribution			
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	1559650422	[*]C[*]	0.129	

ECFP_6	-845108448	S	0.105
FCFP_6	3	[*]N[*]	0.0924
Fingerprint	Top Features	for negative contributio	n Scoro
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	[*]:[c]1:[*]:[cH]:[cH]:1	-0.133
FCFP_6	1	[*]=O	-0.102



Prediction: 0.0838

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

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

- 1. OPS PC20 out of range. Value: 6.397. Training min, max, SD, explained variance: -5.771, 6.0322, 1.721, 0.0130.
- 2. OPS PC31 out of range. Value: 5.4043. Training min, max, SD, explained variance: -4.0208, 4.2529, 1.246, 0.0068.
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
   5.
- 5. 6.

7. 8.

- Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 9. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 10. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 12. Unknown ECFP_6 feature: 835630791: [*][c](:[*]):n:n(:[*]):[*]
- 13. Unknown ECFP_6 feature: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 14. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -1238415266: [*]NN[c](:[*]):[*]
- 16. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 17. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:(*]:n:n:1
- 18. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 19. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 20. Unknown ECFP_6 feature: 1635339976: [*]NNC(=[*])[*]
- 21. Unknown ECFP_6 feature: 150794520: [*]NC(=S)N[*]
- 22. Unknown ECFP_6 feature: 1979182050: [*]C(=S)[*]
- 23. Unknown ECFP_6 feature: -2091181441: [*]C([*])NC(=[*])[*]
- 24. Unknown ECFP_6 feature: -859078569: [*]CC(C[*])N[*]
- 25. Unknown ECFP_6 feature: -1332781180: [*]CCC[*]

	Top features	s for positive contributio	Top features for positive contribution				
ingerprint	Bit/Smiles	es Feature Structure	Score				
	I						
ECFP_6	-167460056	[*]C([*])[*]	0.136				
-------------	---------------------	--------------------------	--------				
ECFP_6	1559650422		0.129				
ECFP_6	-845108448		0.105				
	Top Features for ne	gative contribution					
Fingerprint	Bit/Smiles	Feature Structure	Score				
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.134				
ECFP_6	1564392544	[*]:[cH]:[cH]:1	-0.133				

FCFP_6	1	$\square \cap$	-0.102
		[*]=O	



Prediction: 0.059

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

Unknown features are fingerprint features in 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: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 4. 5.

6. 7.

- 8. Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 9. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 10. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 12. Unknown ECFP_6 feature: 835630791: [*][c](:[*]):n:n(:[*]):[*]
- 13. Unknown ECFP_6 feature: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 14. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 16. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- 17. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 18. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 19. Unknown ECFP_6 feature: -1236714312: [*]=NN[c](:[*]):[*]
- 20. Unknown ECFP_6 feature: 1814278164: [*]N\N=C\[*]
- 21. Unknown ECFP_6 feature: -1832102709: [*]\N=C\[c](:[*]):[*]
- 22. Unknown ECFP_6 feature: -176483725: [*]=C[c](:[cH]:[*]):[cH]:[*]
- 23. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](CI):[cH]:[*]
- 24. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	1559650422		0.129	

FCFP_6	32		0.101
FCFP_6	3		0.0924
Eingernrint	Top Features for ne	egative contribution	Saara
	001725244	reature Structure	0 124
		CI N N N N N N N N N N N N N N N N N N N	
ECFP_6	1564392544	CI NN NH NH NH NH NH NH NH NH NH NH NH NH	-0.133
FCFP_6	1		-0.102



Prediction: 0.0634

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

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: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*] 2.
- Unknown FCFP 2 feature: 1294285001: [*]=NN[c](:[*]):[*] 3.
- 4. 5.

6.

- 7.
- 8. Unknown ECFP 6 feature: 672362763: [*]:n(:[*]):[*]
- 9. Unknown ECFP 6 feature: -1046436026: [*]F
- 10. Unknown ECFP 6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[r]
- Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*] 11.
- Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*] 12.
- Unknown ECFP 6 feature: 835630791: [*][c](:[*]):n:n(:[*]):[*] 13.
- Unknown ECFP_6 feature: -1507082173: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*] 14.
- Unknown ECFP 6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*] 15.
- Unknown ECFP 6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*] 16.
- 17. Unknown ECFP 6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*] 18.
- 19. Unknown ECFP 6 feature: -1793471910: [*]CCC
- Unknown ECFP 6 feature: -1236714312: [*]=NN[c](:[*]):[*] 20.
- 21. Unknown ECFP_6 feature: 1814278164: [*]N\N=C\[*]
- 22. Unknown ECFP 6 feature: -1832102709: [*]\N=C\[c](:[*]):[*]
- 23. Unknown ECFP_6 feature: -176483725: [*]=C[c](:[cH]:[*]):[cH]:[*]
- 24. Unknown ECFP_6 feature: -176686665: [*]:[cH]:[c](F):[cH]:[*]
- 25. Unknown ECFP 6 feature: 220735655: [*]:[c](:[*])F

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	1559650422		0.129	

FCFP_6	32		0.101
FCFP_6	3		0.0924
Eingernrint	Top Features for ne	egative contribution	Saara
FCFP_6	991735244	[*]:[c]1:[*]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	[*]:[cH]:[cH]:1	-0.133
FCFP_6	1		-0.102



Prediction: 0.148

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

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

- 1. OPS PC31 out of range. Value: 4.9251. Training min, max, SD, explained variance: -4.0208, 4.2529, 1.246, 0.0068.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 4. 5.

6. 7.

- 8. Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 9. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 10. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 12. Unknown ECFP_6 feature: 835630791: [*][c](:[*]):n:n(:[*]):[*]
- 13. Unknown ECFP_6 feature: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 14. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 16. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- 17. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 18. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 19. Unknown ECFP_6 feature: -1236714312: [*]=NN[c](:[*]):[*]
- 20. Unknown ECFP_6 feature: 1814278164: [*]N\N=C\[*]
- 21. Unknown ECFP_6 feature: -1832102709: [*]\N=C\[c](:[*]):[*]
- 22. Unknown ECFP_6 feature: -176483725: [*]=C[c](:[cH]:[*]):[cH]:[*]
- 23. Unknown ECFP_6 feature: -177786161: [*]:[cH]:[c](O):[cH]:[*]
- 24. Unknown ECFP_6 feature: 2019062761: [*]:[c](:[*])O

Feature Contribution				
	Top features	for positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	1559650422		0.129	

FCFP_6	3		0.0924
ECFP_6	-1925046727		0.0915
Fingerprint	Top Features for ne	egative contribution	Score
FCFP_6	991735244	H ^O N N N N N N N N N N N N N	-0.134
ECFP_6	1564392544	HO N N N N N N N N N N N N N N N N N N N	-0.133
FCFP_6	1		-0.102



Prediction: 0.13

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

Unknown features are fingerprint features in 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: 5: [*][O-]
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 6. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 7. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 8. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]
- 9. Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 10. Unknown ECFP_6 feature: 1043790491: [*][N+](=[*])[*]
- 11. Unknown ECFP_6 feature: 781519895: [*][O-]
- 12. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 13. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 14. Unknown ECFP_6 feature: 1049768340: [*]N[c](:n:[*]):[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: 835630791: [*][c](:[*]):n:n(:[*]):[*]
- 16. Unknown ECFP_6 feature: -1507082173: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 17. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 18. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 19. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 21. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 22. Unknown ECFP_6 feature: -1236714312: [*]=NN[c](:[*]):[*]
- 23. Unknown ECFP_6 feature: 1814278164: [*]N\N=C\[*]
- 24. Unknown ECFP_6 feature: -1832102709: [*]\N=C\[c](:[*]):[*]
- 25. Unknown ECFP_6 feature: -176483725: [*]=C[c](:[cH]:[*]):[cH]:[*]
- 26. Unknown ECFP_6 feature: -179073144: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 27. Unknown ECFP_6 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
- 28. Unknown ECFP_6 feature: 2104376220: [*][N+](=O)[*]
- 29. Unknown ECFP_6 feature: -659271057: [*][N+](=[*])[O-]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
	·			

ECFP_6	1559650422		0.129
FCFP_6	3		0.0924
ECFP_6	-1925046727	O ₂ N → → → → → → → → → → → → → → → → → → →	0.0915
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	^{O_N}	-0.134
ECFP_6	1564392544	^{02N} ^N ^N ^N ^N ^N ^N ^N 	-0.133

FCFP_6	1	0 ₂ N	-0.102
		N. N H	
		[*]=O	



Prediction: 0.212

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

Unknown features are fingerprint features in 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: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*] 3.
- 4.
- 5. 6.

7.

- 8. Unknown ECFP_6 feature: -830332112: [*]S[*]
- 9. Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 10. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
- 11. Unknown ECFP_6 feature: -955816473: [*]SCC(=[*])[*]
- 12. Unknown ECFP_6 feature: 1427820655: [*]CS[c](:[*]):[*]
- 13. Unknown ECFP_6 feature: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 14. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- 15. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 16. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 17. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 18. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 19. Unknown ECFP_6 feature: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 20. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 21. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 22. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
- 23. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
- 24. Unknown ECFP_6 feature: 1429461619: [*]:[c](:[*])C(=O)O

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422		0.129

FCFP_6	3		0.0924
ECFP_6	2099970318	NHO COOH [*]C(=O)[*]	0.0766
	Top Features for ne	egative contribution	<b>b</b>
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c]1 ^{:[cH]} :[cH]:[cH]:1	-0.134
ECFP_6	1564392544	[*]:[c]1ºP9![cH]:[cH] ]:[cH]:[cH]:1	-0.133
FCFP_6	1		-0.102



Prediction: 0.0948

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

Unknown features are fingerprint features in 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: 5: [*][O-]
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 5. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 6. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 7. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]
- 8. Unknown ECFP_6 feature: -830332112: [*]S[*]
- 9. Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 10. Unknown ECFP_6 feature: 1043790491: [*][N+](=[*])[*]
- 11. Unknown ECFP_6 feature: 781519895: [*][O-]
- 12. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
- 13. Unknown ECFP_6 feature: -955816473: [*]SCC(=[*])[*]
- 14. Unknown ECFP_6 feature: 1427820655: [*]CS[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 16. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- 17. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 18. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 19. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 20. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 21. Unknown ECFP_6 feature: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 22. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 23. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 24. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
- 25. Unknown ECFP_6 feature: -179073144: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 26. Unknown ECFP_6 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
- 27. Unknown ECFP_6 feature: 2104376220: [*][N+](=0)[*]
- 28. Unknown ECFP_6 feature: -659271057: [*][N+](=[*])[O-]

Top features for positive contribution				
Bit/Smiles	Feature Structure	Score		
		•		
	Top features Bit/Smiles	Top features for positive contribution           Bit/Smiles         Feature Structure	Top features for positive contribution         Bit/Smiles       Feature Structure       Score	

ECFP_6	1559650422		0.129
FCFP_6	3	$ \sum_{\substack{N \in \mathcal{N}_{2} \\ N \in \mathcal{N}_{2}}}^{N} \sum_{\substack{N \in \mathcal{N}_{2}}}^{N} \sum_{\substack{N \in \mathcal{N}_{2} \\ N \in \mathcal{N}_{2}}}^{N} \sum_{\substack{N \in \mathcal{N}_{2}}}^{N} \sum_{N \in \mathcal{N$	0.0924
ECFP_6	2099970318	NHO [*]Č(=O)[*]	0.0766
	Top Features for n	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[*]:[c] [↑] [ ⁶ ]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	[*]:[c]↑![¢]:[cH]:[cH]:1	-0.133

FCFP_6	1		-0.102
		N N N N	
		S NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	
		¹¹⁰ ² [*]=O	

 $C_{21}H_{20}N_8O_3S_2$ Molecular Weight: 496.56529
ALogP: 2.496
Rotatable Bonds: 7
Acceptors: 8
Donors: 2

# **Model Prediction**

Prediction: 0.17

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

Unknown features are fingerprint features in 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: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*] 3.
- 4.
- -. 5. 6.

7.

- 8. Unknown ECFP_6 feature: -830332112: [*]S[*]
- 9. Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 10. Unknown ECFP_6 feature: -797085356: [*]S(=[*])(=[*])[*]
- 11. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
- 12. Unknown ECFP_6 feature: -955816473: [*]SCC(=[*])[*]
- 13. Unknown ECFP_6 feature: 1427820655: [*]CS[c](:[*]):[*]
- 14. Unknown ECFP_6 feature: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 15. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- 16. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 17. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 18. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 19. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 20. Unknown ECFP_6 feature: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 21. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 22. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 23. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
- 24. Unknown ECFP_6 feature: -177264675: [*]S(=[*])(=[*])[c](:[cH]:[*]):[cH]:[*]
- 25. Unknown ECFP_6 feature: -2121766239: [*]:[c](:[*])S(=O)(=O)N
- 26. Unknown ECFP_6 feature: 2102150379: [*]S(=[*])(=O)[*]
- 27. Unknown ECFP_6 feature: -934226723: [*]S(=[*])(=[*])N

Feature Contribution	e Contribution
----------------------	----------------

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

ECFP_6	1559650422		0.129
FCFP_6	3		0.0924
ECFP_6	2099970318	NHO SO 2NH2 [*]C(=O)[*]	0.0766
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	[ ^M ]?[6 ^s ]1:[ [*] ]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	[ ^M ]NFo ^S [ ^M ]NFo ^S [ ^M ]NFc ^S ]1:[*]:[cH]:[cH] ]:[cH]:[cH]:1	-0.133

FCFP_6	1	$\bigcirc$	-0.102
		N N N N N N N N N N N N N N N N N N N	
		s s	
		[*]=O	

# TOPKAT_Chronic_LOAEL



# **Model Prediction**

Prediction: 0.0849

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

Unknown features are fingerprint features in 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: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. 4. 5.

6. 7.

- 8. Unknown ECFP_6 feature: 672362763: [*]:n(:[*]):[*]
- 9. Unknown ECFP_6 feature: -830332112: [*]S[*]
- 10. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
- 11. Unknown ECFP_6 feature: 1333660716: [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]
- 12. Unknown ECFP_6 feature: 1049075205: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 13. Unknown ECFP_6 feature: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 14. Unknown ECFP_6 feature: -435188506: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 15. Unknown ECFP_6 feature: 911256832: [*][c]1:[*]:[*]:n:n:1
- 16. Unknown ECFP_6 feature: -1795525632: [*]CC[c](:[*]):[*]
- 17. Unknown ECFP_6 feature: -1793471910: [*]CCC
- 18. Unknown ECFP_6 feature: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 19. Unknown ECFP_6 feature: 1427820655: [*]CS[c](:[*]):[*]
- 20. Unknown ECFP_6 feature: -955816473: [*]SCC(=[*])[*]
- 21. Unknown ECFP_6 feature: -1708545601: [*]CC(=O)N
- 22. Unknown ECFP_6 feature: -932108170: [*]C(=[*])N

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422		0.129
		[*]C[*]	

FCFP_6	3		0.0924
ECFP_6	2099970318	[*]C(=O)[*]	0.0766
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	⁰ H ₂ N s → 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	-0.134
ECFP_6	1564392544	[*]:[c]1:[*]:[cH]:[cH]:1	-0.133
FCFP_6	1		-0.102

# Doxorubicin



### **Model Prediction**

Prediction: 0.013

Unit: g/kg_body_weight

Mahalanobis Distance: Not Available

### Mahalanobis Distance p-value: Not Available

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction. Mahalanobis Distance p-value: The p-value gives the fraction of

Manalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

2.

3.

4. 5.

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

- 1. All properties and OPS components are within expected ranges.
- 6.
   7.
   8. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
   9. Unknown ECFP_6 feature: -813997308: [*]C(=[*])[c](:[c]([*]):[*]):[c]([*]):[*]
   10. Unknown ECFP_6 feature: 1717462980: [*]:[c](:[*])C(=O)[c](:[*]):[*]
- 11. Unknown ECFP_6 feature: -1660913849: [*][c](:[*]):[c](O):[c]([*]):[*]
- 12. Unknown ECFP_6 feature: -1659009760: [*]C([*])[c](:[c]([*]):[*]):[c]([*]):[*]
- 13. Unknown ECFP_6 feature: -1660340418: [*]C[c](:[c]([*]):[*]):[c]([*]):[*]
- 14. Unknown ECFP_6 feature: 1123660302: [*]CC(O[*])[c](:[*]):[*]
- 15. Unknown ECFP_6 feature: 413587124: [*]C([*])CC([*])([*])[*]
- 16. Unknown ECFP_6 feature: 140080459: [*]CC(0)(C[*])C(=[*])[*]
- 17. Unknown ECFP_6 feature: 407900312: [*]C([*])([*])C[c](:[*]):[*]
- 18. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC
- 19. Unknown ECFP_6 feature: 2019062761: [*]:[c](:[*])O
- 20. Unknown ECFP_6 feature: 456242574: [*]C([*])OC([*])[*]
- 21. Unknown ECFP_6 feature: -1783281539: [*]CC(=O)C([*])([*])[*]
- 22. Unknown ECFP_6 feature: -1907393688: [*]C(=[*])CO
- 23. Unknown ECFP_6 feature: 2022454958: [*]CO
- 24. Unknown ECFP_6 feature: 2018700401: [*]C([*])([*])O
- 25. Unknown ECFP_6 feature: 1535429263: [*]OC(C)C([*])[*]
- 26. Unknown ECFP_6 feature: 305695353: [*]C([*])C(O)C([*])[*]
- 27. Unknown ECFP_6 feature: 1201786014: [*]CC(N)C([*])[*]
- 28. Unknown ECFP_6 feature: -801490360: [*]C([*])CC([*])[*]
- 29. Unknown ECFP_6 feature: -1409796893: [*]C([*])OC([*])[*]
- 30. Unknown ECFP_6 feature: -932844120: [*]C([*])N
- 31. Unknown ECFP_6 feature: 865482986: [*]C([*])C
- 32. Unknown ECFP_6 feature: 2024749573: [*]C([*])O
- 33. Unknown ECFP_6 feature: -2061744983: [*]CC(O[*])O[*]

Feature Contribution					
Top features for positive contribution					
Fingerprint Bit/Smiles Feature Structure Score					
	• • • • • • • • • • • • • • • • • • •	· · · · · · · · · · · · · · · · · · ·			

ECFP_6	-167460056		0.136
FCFP_6	-1143715940	[*]C([*])[*]	0.13
ECFP_6	1559650422		0.129
	Top Features	for negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	HO HO HO OH OH OH OH OH OH OH OH OH OH O	-0.134
ECFP_6	2106656448	$HO \rightarrow H^{2} \rightarrow H^{2} \rightarrow HO \rightarrow OH \rightarrow OH \rightarrow OH \rightarrow OH \rightarrow OH \rightarrow OH \rightarrow O$	-0.11

FCFP_6	1	NHa	-0.102
		[*]=O	



Rotatable Bonds: 5

Acceptors: 5

Donors: 3

# **Model Prediction**

Prediction: 0.222

Unit: g/kg_body_weight

Mahalanobis Distance: 8.93

### Mahalanobis Distance p-value: 0.00177

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

# Structural Similar Compounds

Name	PHENOLPHTHALEIN	DISPERSE YELLOW 3	C.I.PIGMENT RED 3	
Structure	O O O O O O O O O O O O O O O O O O O	OH NH	OH ONEO	
Actual Endpoint (-log C)	2.20184	2.77703	2.65635	
Predicted Endpoint (-log C)	2.8857	2.80195	2.97957	
Distance	0.695	0.739	0.754	
Reference	NCI/NTP TR-465	NCI/NTP TR-222	NCI/NTP TR-407	

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

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

FCFP_2	3	(*)N[*]	0.0737
FCFP_2	17	[*]:n:[*]	0.0441
Fingerprint	Top Features for I	negative contributio	n Scoro
FCFP_2	-1272798659		-0.111
FCFP_2	1872154524	[*]C(=O)[*]	-0.105
FCFP_2	1	C → C → C → C → C → C → C → C → C →	-0.0796

# TOPKAT_Rat_Maximum_Tolerated_Dose_Feed

# $C_{19}H_{19}N_7O$

Molecular Weight: 361.40046 ALogP: 3.399 Rotatable Bonds: 5 Acceptors: 5 Donors: 3

# **Model Prediction**

Prediction: 0.369

Unit: g/kg_body_weight

Mahalanobis Distance: 10.5

### Mahalanobis Distance p-value: 5.88e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Name	PHENOLPHTHALEIN	C.I.PIGMENT RED 3	SALICYLAZOSULFAPYRI DINE
Structure	о с с с с с с с с с с с с с с с с с с с	N ^{OH} N ^{FO} W _N × N ^W	HN CONTRACTOR
Actual Endpoint (-log C)	2.20184	2.65635	3.375
Predicted Endpoint (-log C)	2.8857	2.97957	2.80292
Distance	0.765	0.818	0.837
Reference	NCI/NTP TR-465	NCI/NTP TR-407	NCI/NTP TR-457

# Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	3	(*)N[*]	0.0737	
<u> </u>	I		I	

FCFP_2 FCFP_2	17 590925877	<pre></pre>	0.0441 0.00762
	Top Features for pe	[*]N[c](:[cH]:[*]):[c H]:[*]	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659		-0.111
FCFP_2	1872154524	[*]C(=O)[*]	-0.105
FCFP_2	1		-0.0796

### TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



Acceptors: 5

### Donors: 3

### **Model Prediction**

Prediction: 0.366

Unit: g/kg_body_weight

Mahalanobis Distance: 8.77

### Mahalanobis Distance p-value: 0.00286

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### **Structural Similar Compounds**

Name	SALICYLAZOSULFAPYRI DINE	FUROSEMIDE	TOLBUTAMIDE	
Structure	HN TO H	HO H N N N N N N N N N N N N N N N N N N	HN HN OF S N H	
Actual Endpoint (-log C)	3.375	4.04236	2.3985	
Predicted Endpoint (-log C)	2.80292	2.8614	3.32272	
Distance	0.737	0.751	0.795	
Reference	NCI/NTP TR-457	NCI/NTP TR-356	NCI/NTP TR-031	

# Model Applicability

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

- 1. OPS PC12 out of range. Value: -2.4923. Training min, max, SD, explained variance: -2.364, 2.9228, 1.079, 0.0263.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

I op features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	-885550502	[*]C([*])NC(=[*])[*]	0.115	
			•	


### TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



Rotatable Bonds: 7

Acceptors: 5

Donors: 3

### **Model Prediction**

Prediction: 0.26

Unit: g/kg_body_weight

Mahalanobis Distance: 8.89

#### Mahalanobis Distance p-value: 0.00199

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### **Structural Similar Compounds**

Name	SALICYLAZOSULFAPYRI DINE	C.I.PIGMENT RED 3	ACETOHEXAMIDE		
Structure	H N N N N N N N N N N N N N N N N N N N	OH ONEO NN/NN	NH ONH OSS OSS		
Actual Endpoint (-log C)	3.375	2.65635	2.55683		
Predicted Endpoint (-log C)	2.80292	2.97957	3.62413		
Distance	0.742	0.802	0.822		
Reference	NCI/NTP TR-457	NCI/NTP TR-407	NCI/NTP TR-050		

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	-885550502	[*]C([*])NC(=[*])[*]	0.115	
		·	·	

FCFP_2	3	S	0.0737
FCFP_2	17	[*]:n:[*]	0.0441
Eingerprint	Top Features for no	egative contribution	6aara
	1272708650		
FCFP_2	1872154524	[*]C(=O)[*]	-0.105
FCFP_2	1	$ \underbrace{ \begin{array}{c} & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ &$	-0.0796

#### Donors: 1

### **Model Prediction**

Prediction: 0.29

Unit: g/kg_body_weight

Mahalanobis Distance: 11.7

#### Mahalanobis Distance p-value: 3.8e-008

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### **Structural Similar Compounds**

Name	C.I.PIGMENT RED 3	PHENOLPHTHALEIN	CHLORBENZILATE		
Structure	OH ONEO NN/NN	O O O O O O O O O O O O O O O O O O O	CI OH OCI		
Actual Endpoint (-log C)	2.65635	2.20184	3.38252		
Predicted Endpoint (-log C)	2.97957	2.8857	3.27894		
Distance	0.716	0.761	0.788		
Reference	NCI/NTP TR-407	NCI/NTP TR-465	NCI/NTP TR-75		

## Model Applicability

Unknown features are fingerprint features in 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: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]

1		Score
FCFP_2	3	0.0737

FCFP_2	17		0.0441
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	C C Z Z Z Z Z Z Z Z Z Z Z Z Z	-0.134
FCFP_2	-1272798659		-0.111
FCFP_2	203677720	CI N N N [*]=C[c](:[cH]:[*])):[ cH]:[*]	-0.0829

## **Model Prediction**

Prediction: 0.311

Donors: 1

Unit: g/kg_body_weight

Mahalanobis Distance: 11.9

#### Mahalanobis Distance p-value: 2.12e-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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

### **Structural Similar Compounds**

PHENOLPHTHALEIN	CHLORBENZILATE
2.20184	3.38252
2.8857	3.27894
0.748	0.781
NCI/NTP TR-465	NCI/NTP TR-75
	PHENOLPHTHALEIN           PHENOLPHTHALEIN

## Model Applicability

Unknown features are fingerprint features in 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: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3		0.0737

FCFP_2	17	F	0.0441	
		" " "		
		[*]:n:[*]		
	Top Features	for negative contributio	on l	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	71476542	F C C C C C C C C C C C C C C C C C C C	-0.134	
		[*]:[c](:[*])Cl		
FCFP_2	-1272798659	F C	-0.111	
		[*]CCC		
FCFP_2	203677720	F CA	-0.0829	
		[*]=C[c](:[cH]:[*]):[ cH]:[*]		

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

## **Model Prediction**

Prediction: 0.779

Unit: g/kg_body_weight

Mahalanobis Distance: 10.6

#### Mahalanobis Distance p-value: 4.05e-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**

	Compounds		
Name	C.I.PIGMENT RED 3	PHENOLPHTHALEIN	DISPERSE YELLOW 3
Structure	N N N N N N N N N N N N N N N N N N N	O O O O O O O O O O O O O O O O O O O	OH NH
Actual Endpoint (-log C)	2.65635	2.20184	2.77703
Predicted Endpoint (-log C)	2.97957	2.8857	2.80195
Distance	0.654	0.675	0.814
Reference	NCI/NTP TR-407	NCI/NTP TR-465	NCI/NTP TR-222

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]

	Bit/Smiles	Feature Structure	Score
FCFP_2	3		0.0737

FCFP_2	17	HO	0.0441	
		N-N		
		["]:n:["]		
<b></b>	Top Features	for negative contributio	on la	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	7	HO	-0.214	
		[*]0		
FCFP_2	-549108873	но	-0.127	
		[*]:[c](:[*])O		
FCFP_2	-1272798659	HO	-0.111	
		N-N		
		[*]CCC		

### TOPKAT_Rat_Maximum_Tolerated_Dose_Feed

 $C_{19}H_{17}N_7O_2$ Molecular Weight: 375.38397

ALogP: 4.197 Rotatable Bonds: 6

Acceptors: 7

Donors: 1

## **Model Prediction**

Prediction: 0.177

Unit: g/kg_body_weight

Mahalanobis Distance: 10.5

#### Mahalanobis Distance p-value: 6.76e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

### **Structural Similar Compounds**

Name	C.I.PIGMENT RED 3	SALICYLAZOSULFAPYRI DINE	C.I.PIGMENT RED 23
Structure	WINN NW WINN	HN HN HN HN HOH	Production of the second secon
Actual Endpoint (-log C)	2.65635	3.375	2.30052
Predicted Endpoint (-log C)	2.97957	2.80292	3.55333
Distance	0.771	0.840	0.905
Reference	NCI/NTP TR-407	NCI/NTP TR-457	NCI/NTP TR-411

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_2 feature: 8: [*][N+](=[*])[*]
- 3. Unknown FCFP_2 feature: 5: [*][O-]
- 4. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 5. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 6. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 7. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 8. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 9. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
<u> </u>				

FCFP_2	3		0.0737
FCFP_2	17	O ₂ N N N N N N N N N N N N N N N	0.0441
Eingernrint	Top Features for ne	egative contribution	Saara
FCFP_2	-1272798659		-0.111
FCFP_2	203677720	^{0_N} N N N N N N N N N N N N N	-0.0829
FCFP_2	1		-0.0796



Acceptors: 8

Donors: 2

## **Model Prediction**

Prediction: 0.429

Unit: g/kg_body_weight

Mahalanobis Distance: 13.2

#### Mahalanobis Distance p-value: 5.35e-011

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

## **Structural Similar Compounds**

Name	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRI DINE	RESERPINE
Structure	Contraction of the second seco	HN CONTRACTOR	
Actual Endpoint (-log C)	2.30052	3.375	6.13118
Predicted Endpoint (-log C)	3.55333	2.80292	4.38304
Distance	0.650	0.779	0.999
Reference	NCI/NTP TR-411	NCI/NTP TR-457	NCI/NTP TR-193

## Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 5. Training min, max, mean, SD: 0, 4, 1.1685, 0.8469.
- 2. Unknown FCFP_2 feature: -1410079687: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	1036089772	[*]CS[c](:[*]):[*]	0.0749	
	•			





## **Model Prediction**

Prediction: 0.0899

Unit: g/kg_body_weight

Mahalanobis Distance: 13

#### Mahalanobis Distance p-value: 1.37e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

## **Structural Similar Compounds**

Structural Sinnial Compounds			
Name	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRI DINE	RESERPINE
Structure		HN 20 HN N N N N N N N N N N N N N N N N N N	
Actual Endpoint (-log C)	2.30052	3.375	6.13118
Predicted Endpoint (-log C)	3.55333	2.80292	4.38304
Distance	0.698	0.924	1.020
Reference	NCI/NTP TR-411	NCI/NTP TR-457	NCI/NTP TR-193

## Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 5. Training min, max, mean, SD: 0, 4, 1.1685, 0.8469.
- 2. OPS PC9 out of range. Value: 4.4579. Training min, max, SD, explained variance: -2.8548, 3.3954, 1.263, 0.0360.
- 3. Unknown FCFP_2 feature: 8: [*][N+](=[*])[*]
- 4. Unknown FCFP_2 feature: 5: [*][O-]
- 5. Unknown FCFP_2 feature: -1410079687: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 6. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 8. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 9. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 10. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

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

FCFP_2	1036089772	[*]CS[c](:[*]):[*]	0.0749
FCFP_2	3		0.0737
FCFP_2	17	NH ^O [*]:n:[*]	0.0441
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659		-0.111
FCFP_2	1872154524	NH S NN N NH S NN N [*]C(=O)[*]	-0.105

FCFP_2	1	$\land$	-0.0796
		۲ NO₂	
		[*] [*]=O	

 $C_{21}H_{20}N_8O_3S_2$ Molecular Weight: 496.56529 ALogP: 2.496 Rotatable Bonds: 7 Acceptors: 8

## **Model Prediction**

Prediction: 0.0954

Donors: 2

Unit: g/kg_body_weight

Mahalanobis Distance: 14.1

#### Mahalanobis Distance p-value: 7.51e-013

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

<b>Structural Similar</b>	^r Compounds
---------------------------	------------------------

Name	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRI DINE	4,4'-DIAMINO-2,2'- STILBENEDISULFONIC ACID.2NaSALT
Structure	Contraction of the second seco	HN CONTROL	HO U HO U NH 2 HO U NH 2 NH 2
Actual Endpoint (-log C)	2.30052	3.375	2.50759
Predicted Endpoint (-log C)	3.55333	2.80292	3.26068
Distance	0.675	0.870	1.078
Reference	NCI/NTP TR-411	NCI/NTP TR-457	NCI/NTP TR-412

#### Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 5. Training min, max, mean, SD: 0, 4, 1.1685, 0.8469.
- 2. OPS PC9 out of range. Value: 4.4353. Training min, max, SD, explained variance: -2.8548, 3.3954, 1.263, 0.0360.
- 3. Unknown FCFP_2 feature: -1410079687: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 4. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

<b>—</b>	I op features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score		
FCFP_2	1036089772	[*]CS[c](:[*]):[*]	0.0749		

FCFP_2	3	NHO S NH SO NH [*]N[*]	0.0737
FCFP_2	17	$ \begin{array}{c}                                     $	0.0441
	Top Features for n	egative contribution	h
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1272798659	SO NHE [*]CCC	-0.111
FCFP_2	1872154524	NHO SO ,NH [*]C(=O)[*]	-0.105
FCFP_2	203677720	[*]≝C[c](:[cH]:[*]):[ cH]:[*]	-0.0829

C₁₅H₁₅N₇OS Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

## **Model Prediction**

Prediction: 0.118

Unit: g/kg_body_weight

Mahalanobis Distance: 12.2

#### Mahalanobis Distance p-value: 4.36e-009

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Name	C.I.PIGMENT RED 3	SALICYLAZOSULFAPYRI DINE	TRIAMTERENE
Structure	WNYNW WK	HN CONTRACTOR	H ₂ N _{th} N N NH ₂
Actual Endpoint (-log C)	2.65635	3.375	4.00564
Predicted Endpoint (-log C)	2.97957	2.80292	3.1992
Distance	0.831	0.846	0.884
Reference	NCI/NTP TR-407	NCI/NTP TR-457	NCI/NTP TR-420

### Model Applicability

Unknown features are fingerprint features in 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: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: -1410079687: [*]S[c]1:n:[*]:[*]:n:1:[*]

Top features	for positive contributio	n
Bit/Smiles	Feature Structure	Score
1036089772		0.0749
	[*]CS[c](:[*]):[*]	
	Top features         Bit/Smiles         1036089772	Top features for positive contributioBit/SmilesFeature Structure1036089772 $\stackrel{\circ}{H_2N} _{\downarrow} \stackrel{N}{\downarrow} \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow} \stackrel{\downarrow}{\downarrow} \stackrel{\downarrow}{\downarrow} \stackrel{\downarrow}{\downarrow} \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow} \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow}  \stackrel{\downarrow}{\downarrow}  $



## Doxorubicin

#### NH 2 HO $\cap$ OH 0 HO 0 ÓН OH 0 C₂₇H₂₉NO₁₁

Molecular Weight: 543.51925 ALogP: -4.4e-002 Rotatable Bonds: 5 Acceptors: 12 Donors: 6

## Model Prediction

Prediction: 0.277

Unit: g/kg_body_weight

Mahalanobis Distance: 11.1

#### Mahalanobis Distance p-value: 5.33e-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.

## ustural Cimilar Compounds

Name	OXYTETRACYCLINE	4,4'-DIAMINO-2,2'- STILBENEDISULFONIC ACID.2NaSALT	ERYTHROMYCIN
Structure	HO H ₂ N O O O O O H O O H O O H O O H O H O H	NH 2 HO II NH 2 NH 2	Ho the
Actual Endpoint (-log C)	2.36214	2.50759	3.29629
Predicted Endpoint (-log C)	2.77834	3.26068	4.83895
Distance	0.706	0.916	1.034
Reference	NCI/NTP TR-315	NCI/NTP TR-412	NCI/NTP TR-338

## Model Applicability

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

- Molecular PolarSurfaceArea out of range. Value: 206.07. Training min, max, mean, SD: 0, 1. 201.84, 63.052, 40.7.
- Unknown FCFP_2 feature: 1186333723: [*]CC(O[*])[c](:[*]):[*] 2.

	Top features	for positive contribution	on	
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	136627117		0.173	
		·		

FCFP_2	-1143715940		0.095
FCFP_2	1036089772	$HO \rightarrow OH $	0.0749
	Top Features f	or negative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	$HO \xrightarrow{NH_2} O \xrightarrow{OH} O \xrightarrow{O} O \xrightarrow{OH} O \xrightarrow{O} O \longrightarrow{O} O \to O O \to O \to O O \to O $	-0.214
FCFP_2	-549108873	$HO \xrightarrow{NH_2} OH \xrightarrow{O} O O O O O O O O O O O O O O O O O O $	-0.127
FCFP_2	-1272798659		-0.111

 $C_{19}H_{25}N_7O$ Molecular Weight: 367.44809 ALogP: 3.681 Rotatable Bonds: 5

Acceptors: 5

#### Donors: 3

## **Model Prediction**

Prediction: 0.00961

Unit: g/kg_body_weight

Mahalanobis Distance: 14

#### Mahalanobis Distance p-value: 1.21e-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

	Compounds		
Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK
Structure	OH OH HO WAY		
Actual Endpoint (-log C)	6.28396	2.82494	2.54455
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702
Distance	0.706	0.886	1.038
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

## Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC9 out of range. Value: -2.7104. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 6. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 7. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]
- 8. Unknown FCFP_2 feature: 1499521844: [*]NC(=O)N[*]

Top features for positive contribution			
ngerprint	Bit/Smiles	Feature Structure	Score

FCFP 2	1		0.511
		ļΥ	
		N P	
		[*]=O	
FCFP_2	3		0.104
		N _V O	
		[*]N[*]	
FCFP_2	-1272798659	$\cap$	0.0703
		N _v o	
		[*]CCC	
	Top Features for ne	gative contribution	·
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	$\bigcirc$	-0.489
		H T	
		[*]CC	
FCFP_2	1872154524	$\cap$	-0.307
		H T N. N.	
		[*]C(=O)[*]	

FCFP_2	0	$\square \square$	-0.29
		H Y NN H N	
		N N	
		[*]C[*]	

C₁₉H₁₉N₇O Molecular Weight: 361.40046 ALogP: 3.399 Rotatable Bonds: 5

Acceptors: 5

Donors: 3

## **Model Prediction**

Prediction: 0.0102

Unit: g/kg_body_weight

Mahalanobis Distance: 16.9

#### Mahalanobis Distance p-value: 1.3e-015

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### Structural Similar Compounds

	Compounds		
Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK
Structure	OH OH HO WY CI		
Actual Endpoint (-log C)	6.28396	2.82494	2.54455
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702
Distance	0.957	0.977	1.275
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

## Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 4. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC9 out of range. Value: -2.7992. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 6. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 7. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]
- 8. Unknown FCFP_2 feature: 1499521844: [*]NC(=O)N[*]

	Top features	for positive contributio	n
gerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	1		0.511
FCFP_2	3	$ \sum_{\substack{N \in \mathbb{Z} \\ H \in \mathbb{Z} \\ N \in$	0.104
FCFP_2	-1272798659	CCC ^o ^z ^z ^z ^z ^z ^z ^z ^z	0.0703
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	CC	-0.489
FCFP_2	1872154524	[*]C(=O)[*]	-0.307

FCFP_2	0	$\bigcirc$	-0.29
		[*]C[*]	

## **Model Prediction**

Prediction: 0.0421

Unit: g/kg_body_weight

Mahalanobis Distance: 12

#### Mahalanobis Distance p-value: 8.55e-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**

	Compounds		
Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK
Structure	OH OH HOW WCI		
Actual Endpoint (-log C)	6.28396	2.82494	2.54455
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702
Distance	0.721	0.867	1.029
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

## Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC9 out of range. Value: -2.8136. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 6. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 7. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]
- 8. Unknown FCFP_2 feature: 1499521844: [*]NC(=O)N[*]

	Top features	for positive contributio	n
ngerprint	Bit/Smiles	Feature Structure	Score
	4		0 511
-------------	---------------------	----------------------	--------
FGFP_2			0.511
		[*]=O	
FCFP_2	3		0.104
	1070709650		0.0702
FGFP_2	- 12/2/98659		0.0703
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326		-0.489
FCFP_2	1872154524	[*]C(=O)[*]	-0.307

FCFP_2	0		-0.29
		, N H <b>F</b> ^S	
		N N N	
		N N	
		[*]C[*]	

 $C_{19}H_{25}N_7S$ Molecular Weight: 383.51369 ALogP: 4.898 Rotatable Bonds: 7 Acceptors: 5

#### **Model Prediction**

Prediction: 0.0189

Donors: 3

Unit: g/kg_body_weight

Mahalanobis Distance: 12.3

#### Mahalanobis Distance p-value: 3.65e-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.

#### Structural Similar Compounds

Structural Simila	Compounds		
Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK
Structure	OH OH HO WAY		
Actual Endpoint (-log C)	6.28396	2.82494	2.54455
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702
Distance	0.713	0.999	1.097
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

#### Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC9 out of range. Value: -2.7796. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: 1294344583: [*]NN[c](:[*]):[*]
- 6. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 7. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]
- 8. Unknown FCFP_2 feature: 1499521844: [*]NC(=O)N[*]

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

FCFP_2	1		0.511
FCFP_2	3	[*]=0	0.104
FCFP_2	-1272798659		0.0703
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326		-0.489
FCFP_2	1872154524	[*]C(=O)[*]	-0.307

FCFP_2	0	$\cap$	-0.29
		N _s s	
		[*]C[*]	

#### TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage

## C₁₉H₁₇CIN₆ Molecular Weight: 364.83147 ALogP: 4.967 Rotatable Bonds: 5 Acceptors: 5 Donors: 1

#### **Model Prediction**

Prediction: 0.000411

Unit: g/kg_body_weight

Mahalanobis Distance: 16

#### Mahalanobis Distance p-value: 2.06e-014

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### **Structural Similar Compounds**

	l'oompounds		
Name	PHENYLBUTAZONE	PROMETHAZINE.HCL	CHLORPHENIRAMINE MALEATE
Structure		N N S	
Actual Endpoint (-log C)	3.48909	3.93152	3.96188
Predicted Endpoint (-log C)	3.17333	4.72433	3.83117
Distance	1.074	1.102	1.116
Reference	NCI/NTP TR-367	NCI/NTP TR-425	NCI/NTP TR-317

#### Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 4. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC7 out of range. Value: -3.0704. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 6. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 7. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

Top features for positive contribution					
Fingerprint Bit/Smiles Feature Structure Score					
			-		

FCFP_2	32		0.526
FCFP_2	1		0.511
FCFP_2	367998008	C' N N N N N N N N N N N N N N N N N N N	0.413
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326		-0.489
FCFP_2	203677720	CI NNN NH (*]=C[c](:[cH]:[*]):[ cH]:[*]	-0.406

FCFP_2	0		-0.29
		[*]C[*]	

#### TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage

### F $rac{19}{H_{17}}FN_6$ Molecular Weight: 348.37688 ALogP: 4.508 Rotatable Bonds: 5 Acceptors: 5 Donors: 1

#### **Model Prediction**

Prediction: 0.000525

Unit: g/kg_body_weight

Mahalanobis Distance: 16.1

#### Mahalanobis Distance p-value: 1.64e-014

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### **Structural Similar Compounds**

Name	PHENYLBUTAZONE	PROMETHAZINE.HCL	CHLORPHENIRAMINE MALEATE
Structure			
Actual Endpoint (-log C)	3.48909	3.93152	3.96188
Predicted Endpoint (-log C)	3.17333	4.72433	3.83117
Distance	1.062	1.092	1.101
Reference	NCI/NTP TR-367	NCI/NTP TR-425	NCI/NTP TR-317

#### Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 4. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. OPS PC7 out of range. Value: -3.0294. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 5. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 6. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 7. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

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

FCFP_2	32		0.526
FCFP_2	1		0.511
FCFP_2	367998008	[*]:[cH]:[c](CI):[cH] :[']	0.413
	Top Features for ne	gative contribution	-
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326		-0.489
FCFP_2	203677720	[*]=C[c](:[cH]:[*]):[ cH]:[*]	-0.406

FCFP_2	0	F C	-0.29
		[*]C[*]	

HO N N N N

C₁₉H₁₈N₆O Molecular Weight: 346.38582 ALogP: 4.061 Rotatable Bonds: 5 Acceptors: 6 Donors: 2

#### **Model Prediction**

Prediction: 0.00197

Unit: g/kg_body_weight

Mahalanobis Distance: 19.1

#### Mahalanobis Distance p-value: 2.18e-018

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### **Structural Similar Compounds**

Name	OCHRATOXIN	SULFISOOXAZOLE	PROBENECID		
Structure	HONN HONN HONN HONN HONN	H ₂ N N N H			
Actual Endpoint (-log C)	6.28396	2.82494	2.85333		
Predicted Endpoint (-log C)	5.12358	3.0705	2.4258		
Distance	0.968	1.064	1.232		
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-395		

#### Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 4. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 3. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 4. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 5. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 6. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

Feature Cont	Feature Contribution					
	Top features	for positive contributio	n			
Fingerprint	Bit/Smiles	Feature Structure	Score			
FCFP_2	1		0.511			

FCFP_2	3		0.104
FCFP_2	-1272798659		0.0703
Eingerprint	Top Features for n	egative contribution	600r0
FCFP_2	136597326		-0.489
FCFP_2	203677720	HO NN NH NH NH NH NH NH NH NH NH NH NH NH	-0.406
FCFP_2	0		-0.29

O2N

 $C_{19}H_{17}N_7O_2$ Molecular Weight: 375.38397 ALogP: 4.197

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

#### **Model Prediction**

Prediction: 0.00156

Unit: g/kg_body_weight

Mahalanobis Distance: 17.6

#### Mahalanobis Distance p-value: 1.56e-016

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### Structural Similar Compounds

Name	OCHRATOXIN	SULFISOOXAZOLE	PROBENECID	
Structure	OH OH HOW HCI			
Actual Endpoint (-log C)	6.28396	2.82494	2.85333	
Predicted Endpoint (-log C)	5.12358	3.0705	2.4258	
Distance	1.096	1.156	1.294	
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-395	

#### Model Applicability

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

- 1. Num_H_Acceptors out of range. Value: 7. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
- 2. Num_AromaticRings out of range. Value: 4. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 3. OPS PC6 out of range. Value: -2.6439. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
- 4. Unknown FCFP_2 feature: 8: [*][N+](=[*])[*]
- 5. Unknown FCFP_2 feature: 5: [*][O-]
- 6. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 9. Unknown FCFP_2 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 10. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]
- 11. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 12. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 13. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 14. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

# Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

FCFP_2	1		0.511
FCFP_2	3	[*]=O	0.104
FCFP_2	-1272798659		0.0703
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326		-0.489
FCFP_2	203677720	^{0,N} ^N ^N ^N ^N ^N ^N ^N 	-0.406

FCFP_2	0	0 ₂ N	-0.29
		N, N	
		[*]C[*]	



#### **Model Prediction**

Prediction: 0.00596

Unit: g/kg_body_weight

Mahalanobis Distance: 18.3

#### Mahalanobis Distance p-value: 2.45e-017

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### Structural Similar Compounds

Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK	
Structure	OH OH HO WAY	H ₂ N OF N H	E the second sec	
Actual Endpoint (-log C)	6.28396	2.82494	2.54455	
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702	
Distance	1.228	1.392	1.462	
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336	

#### Model Applicability

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

- 1. Molecular_Weight out of range. Value: 461.5. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
- 2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
- 3. Num_AromaticRings out of range. Value: 5. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 4. Molecular_PolarSurfaceArea out of range. Value: 152.07. Training min, max, mean, SD: 0, 138.03, 28.978, 32.1.
- 5. OPS PC1 out of range. Value: 8.6119. Training min, max, SD, explained variance: -4.0008, 7.9165, 2.861, 0.2531.
- 6. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 7. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

Top features for positive contribution					
Fingerprint Bit/Smiles Feature Structure Score					
	•				

FCFP_2	1	NHOS SINCE	0.511
FCFP_2	3	N[*]N[*]	0.104
FCFP_2	-1272798659	NHO COOH [*]CCC	0.0703
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	NHO COOH [*]CC	-0.489
FCFP_2	203677720	[*]=C[c](?[čH]:[*]):[ cH]:[*]	-0.406

FCFP_2	1872154524		-0.307
		NHO SN	
		$\mathbf{\mathcal{P}}$	
		соон [*]C(=O)[*]	



#### **Model Prediction**

Prediction: 0.00306

Unit: g/kg_body_weight

Mahalanobis Distance: 18.8

#### Mahalanobis Distance p-value: 4.78e-018

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### Structural Similar Compounds

Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK	
Structure	OH OH HOW WCI			
Actual Endpoint (-log C)	6.28396	2.82494	2.54455	
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702	
Distance	1.372	1.468	1.544	
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336	

#### Model Applicability

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

- 1. Molecular_Weight out of range. Value: 462.48. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
- 2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
- 3. Num_AromaticRings out of range. Value: 5. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 4. Molecular_PolarSurfaceArea out of range. Value: 160.6. Training min, max, mean, SD: 0, 138.03, 28.978, 32.1.
- 5. OPS PC9 out of range. Value: -2.7709. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 6. Unknown FCFP_2 feature: 8: [*][N+](=[*])[*]
- 7. Unknown FCFP_2 feature: 5: [*][O-]
- 8. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 9. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 10. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 11. Unknown FCFP_2 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 12. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 13. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
- 14. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

#### Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1		0.511
FCFP_2	3		0.104
FCFP_2	-1272798659		0.0703
_	Top Features	for negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	$\sum_{\substack{N \neq 0 \\ N \neq 0}}^{N \neq N} \sum_{\substack{N \neq 0 \\ N \neq 0}}^{N \neq N} \sum_{\substack{N \neq 0 \\ N \neq 0}}^{N \neq N}$	-0.489

FCFP_2	1872154524	NHO [*]C(=O)[*]	-0.307
FCFP_2	0		-0.29

 $H_{PNO_{2}S}$   $C_{21}H_{20}N_{8}O_{3}S_{2}$ Molecular Weight: 496.56529
ALogP: 2.496
Rotatable Bonds: 7
Acceptors: 8

Donors: 2

#### **Model Prediction**

Prediction: 0.0275

Unit: g/kg_body_weight

Mahalanobis Distance: 16

#### Mahalanobis Distance p-value: 2.39e-014

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### **Structural Similar Compounds**

Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK	
Structure	OH OH OH OH OH OH			
Actual Endpoint (-log C)	6.28396	2.82494	2.54455	
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702	
Distance	1.363	1.453	1.562	
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336	

#### Model Applicability

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

- 1. Molecular_Weight out of range. Value: 496.57. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
- 2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
- 3. Num_AromaticRings out of range. Value: 5. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 4. Molecular_PolarSASA out of range. Value: 269.26. Training min, max, mean, SD: 0, 223.97, 50.816, 55.15.
- 5. Molecular_PolarSurfaceArea out of range. Value: 183.32. Training min, max, mean, SD: 0, 138.03, 28.978, 32.1.
- 6. OPS PC1 out of range. Value: 8.7116. Training min, max, SD, explained variance: -4.0008, 7.9165, 2.861, 0.2531.
- 7. OPS PC6 out of range. Value: -2.4899. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
- 8. OPS PC9 out of range. Value: -3.1064. Training min, max, SD, explained variance: -2.7086, 2.9267, 1.019, 0.0321.
- 9. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 10. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 11. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

#### Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1		0.511
FCFP_2	3		0.104
FCFP_2	-1272798659	NHO SO NHO [*]CCC	0.0703
	Top Features f	or negative contributio	n
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	NHO SO NHO [*]CC	-0.489

FCFP_2	203677720	[*] [₩] C[c](:[cH]:[*]):[ cH]:[*]	-0.406
FCFP_2	1872154524	$ \begin{array}{c} & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & $	-0.307

Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

#### **Model Prediction**

Prediction: 0.0362

Unit: g/kg_body_weight

Mahalanobis Distance: 15

#### Mahalanobis Distance p-value: 5.46e-013

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### **Structural Similar Compounds**

Name	SULFISOOXAZOLE	OCHRATOXIN	PENICILLIN VK	
Structure		OH OH HO WALL		
Actual Endpoint (-log C)	2.82494	6.28396	2.54455	
Predicted Endpoint (-log C)	3.0705	5.12358	3.9702	
Distance	1.009	1.086	1.153	
Reference	NCI/NTP TR-138	NCI/NTP TR-358	NCI/NTP TR-336	

#### Model Applicability

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

- 1. Num_AromaticRings out of range. Value: 4. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
- 2. Unknown FCFP_2 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 3. Unknown FCFP_2 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 4. Unknown FCFP_2 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
FCFP_2	1	[*]=O	0.511	
	·	·	·	

FCFP_2	3		0.104
FCFP_2	-1272798659		0.0703
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326		-0.489
FCFP_2	1872154524	⁰ H ₂ N S → N N N N N N N N N N N N N N	-0.307
FCFP_2	0		-0.29

#### Doxorubicin

#### NH 2 HO 1.0 $\cap$ OH 0 HO 0 ÓН OH 0 C₂₇H₂₉NO₁₁ Molecular Weight: 543.51925 ALogP: -4.4e-002 Rotatable Bonds: 5 Acceptors: 12 Donors: 6

#### **Model Prediction**

Prediction: 4.1e-005

Unit: g/kg_body_weight

Mahalanobis Distance: 18.3

#### Mahalanobis Distance p-value: 1.97e-017

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### Structural Similar Compounds

Structural Similar Compounds					
Name	AMPICILLIN TRIHYDRATE	OCHRATOXIN	PENICILLIN VK		
Structure	H 2N H	OH OH HO WH	NH NH NH NH NH NH NH NH NH NH		
Actual Endpoint (-log C)	2.36724	6.28396	2.54455		
Predicted Endpoint (-log C)	2.27651	5.12358	3.9702		
Distance	1.486	1.531	1.674		
Reference	NCI/NTP TR-318	NCI/NTP TR-358	NCI/NTP TR-336		

#### Model Applicability

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

- 1. Molecular_Weight out of range. Value: 543.52. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
- 2. Num_H_Donors out of range. Value: 6. Training min, max, mean, SD: 0, 3, 0.4375, 0.8311.
- 3. Num_H_Acceptors out of range. Value: 12. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
- 4. Molecular_PolarSASA out of range. Value: 345.3. Training min, max, mean, SD: 0, 223.97, 50.816, 55.15.
- 5. Molecular_PolarSurfaceArea out of range. Value: 206.07. Training min, max, mean, SD: 0, 138.03, 28.978, 32.1.
- 6. OPS PC1 out of range. Value: 12.738. Training min, max, SD, explained variance: -4.0008, 7.9165, 2.861, 0.2531.
- 7. Unknown FCFP_2 feature: -1549669478: [*]:[c](:[*])C(=O)[c](:[*]):[*]
- 8. Unknown FCFP_2 feature: -415156552: [*]CC(O)(C[*])C(=[*])[*]

Top features for positive contribution					
Fingerprint Bit/Smiles Feature Structure Score					
		•			

FCFP_2	332760439	$HO \longrightarrow OH $	0.672
FCFP_2	1		0.511
FCFP_2	3		0.104
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326		-0.489
FCFP_2	203677720	HO NH ₂ HO OH OH O OH OH O OH OH O OH OH O OH OH OH O OH O O	-0.406

FCFP_2	1872154524	$NH_2$	-0.307
		[*]C(=O)[*]	



#### **Model Prediction**

Prediction: 0.861

Unit: g/kg_body_weight

Mahalanobis Distance: 27.2

#### Mahalanobis Distance p-value: 1.37e-044

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	FLUBENDAZOLE	CARBAMIC ACID; N-(5- BENZOYLBENZIMIDAZOL -2-YL)-; METHYL ESTER	METHYL-1- (BUTYLCARBAMOYL)-2- BENZIMIDAZOLE CARBAMATE	
Structure	F N N H V N H			
Actual Endpoint (-log C)	2.088	2.617	1.463	
Predicted Endpoint (-log C)	2.69288	2.2368	1.78491	
Distance	0.641	0.651	0.701	
Reference	YRTMA6 9;11;78	IYKEDH 19;735;88	JHEMA2 24;295;80	

#### Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 3. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
- 4. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 6. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1294344583: [*]NN[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 10. Unknown FCFP_6 feature: -885461129: [*]NNC(=[*])[*]

# Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

ECFP_6	642810091	[*]:[c](:[*]):[*]	0.281
FCFP_6	1499521844	[*]NC(=0)N[*]	0.258
ECFP_6	-1897341097		0.216
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	<pre></pre>	-0.239
ECFP_6	734603939	$O_{z_{\pi}}^{O} \xrightarrow{z_{\pi}}_{z_{\pi}}^{z_{\pi}} \xrightarrow{z_{\pi}}_{z_{\pi}}^{z_{\pi}}} \xrightarrow{z_{\pi}}_{z_{\pi}}^{z_{\pi}} z_$	-0.201

ECFP_6	-1795525632		-0.176
		[*]CC[c](:[*]):[*]	

 $C_{19}H_{19}N_7O$ Molecular Weight: 361.40046 ALogP: 3.399 Rotatable Bonds: 5

Acceptors: 5

#### Donors: 3

#### **Model Prediction**

Prediction: 2.51

Unit: g/kg_body_weight

Mahalanobis Distance: 27.5

#### Mahalanobis Distance p-value: 2.14e-046

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	CARBAMIC ACID; N-(5- BENZOYLBENZIMIDAZOL -2-YL)-; METHYL ESTER	FLUBENDAZOLE	SULFAQUINOXALINE	
Structure			N N N N N N N N N N N N N N N N N N N	
Actual Endpoint (-log C)	2.617	2.088	2.341	
Predicted Endpoint (-log C)	2.2368	2.69288	2.42674	
Distance	0.705	0.709	0.785	
Reference	IYKEDH 19;735;88	YRTMA6 9;11;78	MahWM# 16NOV82	

#### Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 3. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[cH]:[*]
- 4. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 6. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1294344583: [*]NN[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 10. Unknown FCFP_6 feature: -885461129: [*]NNC(=[*])[*]

Top features for positive contribution				
Bit/Smiles	Feature Structure	Score		
	Top features Bit/Smiles	Top features for positive contributio           Bit/Smiles         Feature Structure	Top features for positive contribution           Bit/Smiles         Feature Structure         Score	
ECFP_6	642810091	[*]:[c](:[*]):[*]	0.281	
-------------	---------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------	
FCFP_6	1499521844	[*]NC(=0)N[*]	0.258	
ECFP_6	-1897341097		0.216	
	Top Features for ne	gative contribution		
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	655739385	<pre></pre>	-0.239	
ECFP_6	734603939	$\sum_{Z_{\perp}}^{O} \sum_{Z_{\perp}}^{Z_{\perp}} \sum_{Z_{\perp}}^{Z_{\perp}} \sum_{Z_{\perp}}^{Z_{\perp}} \sum_{Z_{\perp}}^{Z_{\perp}} \sum_{Z_{\perp}}^{Z_{\perp}} \sum_{Z_{\perp}}^{U_{\perp}} \sum_{Z_{\perp}}^{U_{$	-0.201	

ECFP_6	-1795525632		-0.176
		[*]CC[c](:[*]):[*]	



Prediction: 0.742

Unit: g/kg_body_weight

Mahalanobis Distance: 26.7

#### Mahalanobis Distance p-value: 2.43e-041

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	CARBAMIC ACID; [1-[(5- CYANOPENTYL)CARBAM OYL]BENZIMIDAZOL-2- YL]-; METHYL ESTER	OCHRATOXIN A	CARBAMIC ACID; N-(5- BENZOYLBENZIMIDAZOL -2-YL)-; METHYL ESTER	
Structure		OH OH HO WY CI		
Actual Endpoint (-log C)	2.12	4.305	2.617	
Predicted Endpoint (-log C)	1.78415	3.03558	2.2368	
Distance	0.666	0.720	0.725	
Reference	85ARAE 4;118;76/77	FCTXAV 6;479;68	IYKEDH 19;735;88	

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 3. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
- 4. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 6. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1294344583: [*]NN[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 10. Unknown FCFP_6 feature: -885461129: [*]NNC(=[*])[*]

# Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

ECFP_6	642810091		0.281
FCEP 6	1/005218//	[*]:[c](:[*]):[*]	0 258
FOFF_0	1499521844	[*]NC(=O)N[*]	0.200
ECFP_6	-1897341097	[*]N[*]	0.216
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	497523368	[*]CNC(=[*])[*]	-0.301
ECFP_6	655739385	[*]:n:[*]	-0.239

ECFP_6	734603939		-0.201
		N H F ^S	
		[*]C	



Prediction: 0.297

Unit: g/kg_body_weight

Mahalanobis Distance: 27.1

#### Mahalanobis Distance p-value: 1.27e-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	PHOSPHORAMIDOTHIOIC ACID; ACETIMIDOYL-; O;O-bis-(p- CHLOROPHENYL)ESTER	OCHRATOXIN A	[4-CHLORO-6-(2;3- XYLIDINO)2- PYRIMIDINYLTHIO]ACETI C ACID	
Structure		OH OH HOW WILL	HO O S CI N H	
Actual Endpoint (-log C)	5.006	4.305	1.892	
Predicted Endpoint (-log C)	3.23989	3.03558	3.41405	
Distance	0.716	0.730	0.737	
Reference	FMCHA2 -;C149;89	FCTXAV 6;479;68	DRFUD4 4;273;79	

## Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 3. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
- 4. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 6. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: 1294344583: [*]NN[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 10. Unknown FCFP_6 feature: -885461129: [*]NNC(=[*])[*]

## Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

ECFP_6	642810091	[*]:[c](:[*]):[*]	0.281
FCFP_6	1499521844	[*]NC(=O)N[*]	0.258
ECFP_6	-1897341097		0.216
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	() ∠ ^Z ^T ^Z ^T ^Z ^T ^Z ^T ^Z	-0.239
ECFP_6	734603939	$O_{z_{\pm}}^{m} \overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}}{\overset{z_{\pm}}{\overset{z_{\pm}}{\overset{z_{\pm}}}{\overset{z_{\pm}}{\overset{z_{\pm}}}{\overset{z_{\pm}}{\overset{z_{\pm}}}{\overset{z_{\pm}}{\overset{z_{\pm}}}{\overset{z_{\pm}}{\overset{z_{\pm}}}{\overset{z_{\pm}}}{\overset{z_{\pm}}}{\overset{z_{\pm}}}{\overset{z_{\pm}}}}}}}}}}}}}$	-0.201

ECFP_6	-1795525632		-0.176
		[*]CC[c](:[*]):[*]	

C₁₉H₁₇CIN₆ Molecular Weight: 364.83147 ALogP: 4.967 Rotatable Bonds: 5 Acceptors: 5 Donors: 1

## **Model Prediction**

Prediction: 0.447

Unit: g/kg_body_weight

Mahalanobis Distance: 23.9

#### Mahalanobis Distance p-value: 2.66e-025

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Name	TALNIFLUMATE	INDOMETHAZINE	ACRIDINE; 9-[3- (DIMETHYLAMINO)PROPY LAMINO]-1-NITRO-
Structure	or or h h h h h h h	HO	
Actual Endpoint (-log C)	1.538	5.17	4.101
Predicted Endpoint (-log C)	2.82541	3.33605	3.3633
Distance	0.671	0.676	0.687
Reference	FRPSAX 36;372;81	ARZNAD 25;1526;75	MMDPA6 8;252;76

#### Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 3. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[cH]:[*]
- 4. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 6. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 9. Unknown FCFP_6 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 10. Unknown FCFP_6 feature: 581019816: [*]N\N=C\[*]
- 11. Unknown FCFP_6 feature: -2100785893: [*]\N=C\[c](:[*]):[*]
- 12. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

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

ECFP_6	642810091		0.281
ECFP_6	-1897341097		0.216
FCFP_6	-149636017	CI NN NH NH NH NH NH NH NH NH NH NH NH NH	0.193
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-817402818		-0.263
ECFP_6	655739385	CI NNN NH NH NH NH NH NH NH NH NH NH NH NH	-0.239

ECFP_6	734603939		-0.201
		[*]C	



Prediction: 0.229

Unit: g/kg_body_weight

Mahalanobis Distance: 25.2

#### Mahalanobis Distance p-value: 3e-032

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### **Structural Similar Compounds**

Name	ACRIDINE; 9-[3- (DIMETHYLAMINO)PROP YLAMINO]-1-NITRO-	INDOMETHAZINE	TALNIFLUMATE
Structure		HO	
Actual Endpoint (-log C)	4.101	5.17	1.538
Predicted Endpoint (-log C)	3.3633	3.33605	2.82541
Distance	0.657	0.680	0.682
Reference	MMDPA6 8;252;76	ARZNAD 25;1526;75	FRPSAX 36;372;81

#### Model Applicability

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

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 3. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
- 4. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 6. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 9. Unknown FCFP_6 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 10. Unknown FCFP_6 feature: 581019816: [*]N\N=C\[*]
- 11. Unknown FCFP_6 feature: -2100785893: [*]\N=C\[c](:[*]):[*]
- 12. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

#### Feature Contribution

Top features for positive contribution				
Fingerprint	ingerprint Bit/Smiles Feature Structure Score			

ECFP_6	-1046436026		0.349
ECFP_6	642810091	[*]:[c](:[*]):[*]	0.281
ECFP_6	-1897341097	[*]N[*]	0.216
	Top Features for ne	gative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	F → → N N N N N N N N N N N N N	-0.239
ECFP_6	734603939	F	-0.201

ECFP_6	-1795525632	F N N	-0.176
		[*]CC[c](:[*]):[*]	



#### Donors: 2

## **Model Prediction**

Prediction: 0.501

Unit: g/kg_body_weight

Mahalanobis Distance: 24.6

#### Mahalanobis Distance p-value: 9.13e-029

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

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

Name	ACRIDINE; 9-[3- (DIMETHYLAMINO)PROP YLAMINO]-1-NITRO-	FLUBENDAZOLE	CARBAMIC ACID; N-(5- BENZOYLBENZIMIDAZOL -2-YL)-; METHYL ESTER
Structure		P P P P P P P P P P P P P P P P P P P	
Actual Endpoint (-log C)	4.101	2.088	2.617
Predicted Endpoint (-log C)	3.3633	2.69288	2.2368
Distance	0.701	0.709	0.713
Reference	MMDPA6 8;252;76	YRTMA6 9;11;78	IYKEDH 19;735;88

#### Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 3. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
- 4. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 5. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 6. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 7. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 8. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 9. Unknown FCFP_6 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 10. Unknown FCFP_6 feature: 581019816: [*]N\N=C\[*]
- 11. Unknown FCFP_6 feature: -2100785893: [*]\N=C\[c](:[*]):[*]
- 12. Unknown FCFP_6 feature: 74595001: [*]:[cH]:[c](O):[cH]:[*]
- 13. Unknown FCFP_6 feature: -549108873: [*]:[c](:[*])O

Feature Contr	ribution			
	Top features	for positive contribution	n	
Fingerprint Bit/Smiles Feature Structure Score				

ECFP_6	642810091		0.281
		[*]:[c](:[*]):[*]	
ECFP_6	-1897341097	HO	0.216
		[*]N[*]	
ECFP_6	2019062761		0.138
		[*1:[c](:[*1))O	
	Top Features for ne	agative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385		-0.239
	046590555	[*]:n:[*]	0.204
г <b>чг</b> й_0	840009000	H ^O N N N N N N N N N N N N N	-v.2v4

ECFP_6	734603939	HO	-0.201
		[*]C	

 $C_{19}H_{17}N_7O_2$ Molecular Weight: 375.38397

ALogP: 4.197 Rotatable Bonds: 6

Acceptors: 7

Donors: 1

## **Model Prediction**

Prediction: 0.573

Unit: g/kg_body_weight

Mahalanobis Distance: 27.1

#### Mahalanobis Distance p-value: 5.18e-044

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

<b>Structural S</b>	Similar Co	mpounds
---------------------	------------	---------

Name	ACEMETACIN	bis-OXATIN ACETATE	PRASOZIN .HCI (HCI STRIPPED)
Structure	OF OH		N N H ₂ N ^M H ₂ N ^M N H ₂ N ^M N N N N N N N N N N N N N N N N N N N
Actual Endpoint (-log C)	4.235	1.717	2.294
Predicted Endpoint (-log C)	3.39415	2.40947	3.00765
Distance	0.695	0.721	0.728
Reference	ARZNAD 30;1398;80	NIIRDN 6;609;82	NIIRDN 6;688;82

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1043790491: [*][N+](=[*])[*]
- 3. Unknown ECFP_2 feature: 781519895: [*][O-]
- 4. Unknown ECFP_2 feature: -179073144: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 5. Unknown ECFP_2 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
- 6. Unknown ECFP_2 feature: 2104376220: [*][N+](=O)[*]
- 7. Unknown ECFP_2 feature: -659271057: [*][N+](=[*])[O-]
- 8. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 9. Unknown FCFP_6 feature: 8: [*][N+](=[*])[*]
- 10. Unknown FCFP_6 feature: 5: [*][O-]
- 11. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
- 12. Unknown FCFP_6 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
- 13. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 14. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 15. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 16. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 17. Unknown FCFP_6 feature: 1294285001: [*]=NN[c](:[*]):[*]
- 18. Unknown FCFP_6 feature: 581019816: [*]N\N=C\[*]
- 19. Unknown FCFP_6 feature: -2100785893: [*]\N=C\[c](:[*]):[*]

- 20. Unknown FCFP_6 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 21. Unknown FCFP_6 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 22. Unknown FCFP_6 feature: 1872392852: [*][N+](=O)[*]
- 23. Unknown FCFP_6 feature: 260476081: [*][N+](=[*])[O-]



ECFP_6	655739385	O ₂ N N N N N N N N N N N N N N N	-0.239
ECFP_6	734603939		-0.201
ECFP_6	-1795525632	^{O₂N → → → → → → → → → → → → → → → → → →}	-0.176



Prediction: 2.37

Unit: g/kg_body_weight

Mahalanobis Distance: 28.8

#### Mahalanobis Distance p-value: 1.8e-055

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	BENZENESULFONIC ACID; 2;2'-(4;4'- BIPHENYLYLENEDIVINYL ENE)DI-; DISODIUM SALT (Na STRIPPED)	BENZOTHIAZOLE; 6- NITRO-2-(p- NITROBENZOYLAMINO)-	AZOSEMIDE	
Structure	C C C C C C C C C C C C C C C C C C C		HN X N HN X N HN X N HN X N H NH 2 CI	
Actual Endpoint (-log C)	1.968	2.361	2.163	
Predicted Endpoint (-log C)	1.72109	2.96257	2.21052	
Distance	0.839	0.960	0.986	
Reference	MVCRB3 2;193;73	JPETAB 90;260;47	IYKEDH 18;666;87	

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 3. Unknown ECFP_2 feature: 78665610: [*][c]1:[*]:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 4. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 5. Unknown FCFP_6 feature: -1410079687: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 6. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 7. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 8. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 10. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
- 11. Unknown FCFP_6 feature: -1549222613: [*]:[c](:[*])C(=O)O

Feature Contribution					
Top features for positive contribution					
Fingerprint Bit/Smiles Feature Structure Score					
			-		

ECFP_6	642810091	NH [*] O ^N N ^N	0.281
ECFP_6	-1897341097		0.216
ECFP_6	1444581947	[*]C(=[*](G]4:[CH]:[ *]:[c]([*]):[CH]:[CH]:[CH	0.163
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	$ \begin{array}{c}                                     $	-0.239
FCFP_6	566058135	[*]CC(=O)N[*]	-0.216

ECFP_6	734603939	N	-0.201
		Г ^{соон} [*]С	



Prediction: 1.86

Unit: g/kg_body_weight

Mahalanobis Distance: 27.7

#### Mahalanobis Distance p-value: 6.3e-048

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural Similar Compounds				
Name	BENZENESULFONIC ACID; 2;2'-(4;4'- BIPHENYLYLENEDIVINYL ENE)DI-; DISODIUM SALT (Na STRIPPED)	BENZOTHIAZOLE; 6- NITRO-2-(p- NITROBENZOYLAMINO)-	ACEMETACIN	
Structure			OC OH	
Actual Endpoint (-log C)	1.968	2.361	4.235	
Predicted Endpoint (-log C)	1.72109	2.96257	3.39415	
Distance	0.884	0.916	1.015	
Reference	MVCRB3 2;193;73	JPETAB 90;260;47	ARZNAD 30;1398;80	

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown ECFP_2 feature: 1043790491: [*][N+](=[*])[*]
- 3. Unknown ECFP_2 feature: 781519895: [*][O-]
- 4. Unknown ECFP_2 feature: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 5. Unknown ECFP_2 feature: 78665610: [*][c]1:[*]:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 6. Unknown ECFP_2 feature: -179073144: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 7. Unknown ECFP_2 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
- 8. Unknown ECFP_2 feature: 2104376220: [*][N+](=O)[*]
- 9. Unknown ECFP_2 feature: -659271057: [*][N+](=[*])[O-]
- 10. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 11. Unknown FCFP_6 feature: 8: [*][N+](=[*])[*]
- 12. Unknown FCFP_6 feature: 5: [*][O-]
- 13. Unknown FCFP_6 feature: -1410079687: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 14. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 15. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 16. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 17. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]

- 18. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[cH]:[*]
- 19. Unknown FCFP_6 feature: -828984032: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
- 20. Unknown FCFP_6 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
- 21. Unknown FCFP_6 feature: 1872392852: [*][N+](=O)[*]
- 22. Unknown FCFP_6 feature: 260476081: [*][N+](=[*])[O-]

	Top features f	or positive contributio	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	
ECFP_6	642810091		0.281	
ECFP_6	-1897341097		0.216	
ECFP_6	-1074141656	$\sum_{\substack{N \in \mathcal{S} \\ N \in $	0.142	
	Top Features f	or negative contribution	n	
Fingerprint	Bit/Smiles	Feature Structure	Score	

ECFP_6	655739385	NH ^S [*]:n:[*]	-0.239
FCFP_6	566058135	[*]CC(=O)N[*]	-0.216
ECFP_6	734603939		-0.201



Prediction: 12.3

Unit: g/kg_body_weight

Mahalanobis Distance: 29.9

#### Mahalanobis Distance p-value: 8.72e-064

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The 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

Unknown features are fingerprint features in 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: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 3. Unknown ECFP_2 feature: 78665610: [*][c]1:[*]:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 4. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 5. Unknown FCFP_6 feature: -1410079687: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 6. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 7. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 8. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 9. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 10. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[cH]:[*]

## Feature Contribution Top features for positive contribution Fingerprint Bit/Smiles Feature Structure Score

ECFP_6	642810091	NH ^S SO ,NH, [*]:[c](:[*]):[*]	0.281
ECFP_6	-1897341097		0.216
ECFP_6	-1074141656		0.142
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385		-0.239
FCFP_6	-1096219292	HNO.S [']:[c](:')]S(=O)(=O) NN	-0.225

FCFP_6	566058135		-0.216
		[*]CC(=Ó)N[*]	

 $C_{15}H_{15}N_{7}OS$ 

Molecular Weight: 341.39089 ALogP: 2.008 Rotatable Bonds: 5 Acceptors: 6 Donors: 1

#### **Model Prediction**

Prediction: 0.542

Unit: g/kg_body_weight

Mahalanobis Distance: 25.8

#### Mahalanobis Distance p-value: 5.79e-036

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

Structural	Similar	Com	pounds
------------	---------	-----	--------

Name	PRASOZIN .HCI (HCI STRIPPED)	AZATHIOPRINE	SULFAQUINOXALINE
Structure			NH NH ON NH2
Actual Endpoint (-log C)	2.294	2.715	2.341
Predicted Endpoint (-log C)	3.00765	2.77505	2.42674
Distance	0.684	0.700	0.729
Reference	NIIRDN 6;688;82	NIIRDN 6;3;82	MahWM# 16NOV82

## Model Applicability

Unknown features are fingerprint features in 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: 78665610: [*][c]1:[*]:[c](:[*]):n:1:n(:[*]):[*]
- 3. Unknown ECFP_2 feature: 1986731747: [*]S[c]1:n:[*]:[*]:n:1:[*]
- 4. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 5. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[*]
- 6. Unknown FCFP_6 feature: -1564473960: [*]:n1:[*]:[*]:n:[c]:1:[c](:[*]):[*]
- 7. Unknown FCFP_6 feature: -332197802: [*][c]1:[*]:[c](:[*]):n:1:n:[*]
- 8. Unknown FCFP_6 feature: -1539162406: [*]C[c]1:n:[*]:[*]:n:1:[*]
- 9. Unknown FCFP_6 feature: 4427049: [*][c](:[*]):n:n(:[*]):[*]
- 10. Unknown FCFP_6 feature: -1410079687: [*]S[c]1:n:[*]:[*]:n:1:[*]

#### **Feature Contribution**

Top features for positive contribution				
Fingerprint	Bit/Smiles	Feature Structure	Score	
		1		

ECFP_6	642810091	[*]:[c](:[*]):[*]	0.281
ECFP_6	-1074141656		0.142
ECFP_6	-932108170	⁰ H ₂ N S N N N N N N N N N N N	0.126
	Top Features	for negative contribution	on
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	O H ₂ N S N N N N N N N N N N N N N N N N N N	-0.239
FCFP_6	566058135	[*]CC(=0)N[*]	-0.216

ECFP_6	734603939		-0.201
		[*]C	

## Doxorubicin



Donors: 6

## **Model Prediction**

Prediction: 0.227

Unit: g/kg_body_weight

Mahalanobis Distance: 23.8

#### Mahalanobis Distance p-value: 7.02e-025

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.

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

#### **Structural Similar Compounds**

Name	DAUNOMYCIN	VIRIDICATUM TOXIN	SECALONIC ACID
Structure	HO to the total state of total		
Actual Endpoint (-log C)	3.196	3.666	4.463
Predicted Endpoint (-log C)	3.6117	3.0269	3.37815
Distance	0.329	0.669	0.680
Reference	YKYUA6 25;573;74	TXAPA9 24;507;73	TXAPA9 48;A14;79

## Model Applicability

- 1. All properties and OPS components are within expected ranges.
- 2. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
- 3. Unknown FCFP_6 feature: 1618154665: [*]:[cH]:[cH]:[cH]:[cH]:[*]
- 4. Unknown FCFP_6 feature: -1549669478: [*]:[c](:[*])C(=O)[c](:[*]):[*]
- 5. Unknown FCFP_6 feature: 74595001: [*]:[cH]:[c](O):[cH]:[*]
- 6. Unknown FCFP_6 feature: 1186333723: [*]CC(O[*])[c](:[*]):[*]
- 7. Unknown FCFP_6 feature: -549108873: [*]:[c](:[*])O

Score	Feature Structure Sc	Bit/Smiles	
			ngerprint

ECFP_6	642810091	HO HO OH OH OH OH OH OH OH OH OH OH OH O	0.281
FCFP_6	136627117		0.17
ECFP_6	-1074141656	HO NH ₂ HO OH OH O OH OH O OH OH O OH OH O (*]=O	0.142
	Top Features for ne	egative contribution	
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	$HO \rightarrow H^{H_2}$ $HO \rightarrow H^{H_2}$	-0.352
ECFP_6	683445015		-0.266

ECFP_6	734603939	NH ₂	-0.201
		[*]C	

#### In silico ADMET analysis

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

#### **Toxicity studies**

The toxicity parameters of the synthesized compounds were calculated using Discovery studio 4.0. Doxorubicin 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.
## In silico ADMET analysis

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

# **Toxicity studies**

The toxicity parameters of the synthesized compounds were calculated using Discovery studio 4.0. Doxorubicin 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.

#### 4.2.1. In vitro anti-proliferative activity against MCF-7, HepG-2 and HCT-116

The cells were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), penicillin (100 unit /mL) and streptomycin sulphate (100µg/mL) at 37 oC in a 5% CO2 incubator. Then, cells were dropped in 96-well plates at a density of  $3-8 \times 103$  cells/well and incubated for two days in a 5% CO2 incubator at 37 °C. Then, the cells were treated with the synthesized compounds and the cell cultures were continued incubated for 24 h. However, Different concentrations of the compound under test (0.0, 5.0, 12.5, 25.0 and 50 mg/ml) were added to the cell monolayer. Then, (3-[4,5-dimethylthiazol-2-yl]- 2,5-diphenyltetrazolium bromide) MTT solution (20 µl, 5mg/mL) was added to each well and incubated for additional 4 h. The formed purple crystals of MTT-formazan were dissolved in 100 µl DMSO each well; the absorbance of each well was measured at 570 nm using a plate reader (EXL 800, USA). All of the compounds were tested three times. The relative cell viability in percentage was calculated. The results for IC₅₀ values of the active compounds are summarized in Table 1. The data represented the mean of three independent experiments in triplicate and were expressed as means ± SD. The IC₅₀ value was defined as the concentration at which 50% of the cells could survive.

### 4.2.2. DNA intercalation assay (DNA/methyl green colorimetric assay)

Affinity of some compounds for DNA-binding was examined in vitro using DNA/methyl green assay using doxorubicin as a positive control. In this test, a mixture of methyl green (20 mg) and Calf thymus DNA (10 mg) (Sigma-Aldrich) were suspended in 0.05 M TrisHCl buffer (100 mL, pH 7.5) containing 7.5 mM MgSO4. This mixture was stirred continuously for 24 h at 37 °C. Then, ethanolic solution of the test compounds was pitted into the wells of a 96-well microtiter tray at a concentration of 10, 100 and 1000 µM. The excess solvent was removed from each well under vacuum, with subsequent addition of 200 µL of the DNA/methyl green solution. The test samples were incubated for 24 h in dark at ambient temperature. After that, absorbance of each sample was determined at 642.5-645 nm. In this test, the methyl green dye reversibly binds to DNA to form persistent colored complex of DNA/methyl green. This color still stables at neutral pH. When the DNA intercalators were added, the methyl green was displaced from DNA with addition of H2O molecule to the dye resulting in formation of the colorless carbinol leading to a decrease in spectrophotometric absorbance.  $\Delta A$  value (the difference between DNA/methyl green complex and free cabinol) provides the simplest means for detecting the DNA-binding affinity and relative binding strength. IC50's were determined by linear regression of data plotted on a semi-log scale and the data were compared with doxorubicin as standard DNA intercalator.

#### 4.2.3. Measurement of topoisomerase II activity

Compounds (8a, 8b, 9a, 9c, 9d, 12a, and 12b) were further evaluated to assess their Topo II inhibitory activities. In this test, Topo II drug screening kit (TopoGEN, Inc., Columbus) was utilized to determine the Topo II activity. Doxorubicin was used as a reference drug in this test. A typical enzyme reaction contained a mixture of human Topo II (2  $\mu$ L), substrate super coiled pHot1 DNA (0.25  $\mu$ g), 50  $\mu$ g/mL test compound (2  $\mu$ L), and assay buffer (4  $\mu$ L). The reaction started upon incubation of the mixture in 37 °C for 30 min. The reaction was terminated by addition of 10% sodium dodecylsulphate (2  $\mu$ L) and proteinase K (50  $\mu$ g/mL) at 37 °C for 15 min. followed by incubation for 15 min at 37 °C. Then, the DNA was run on 1% agarose gel in BioRad gel electrophoresis system for 1–2 h followed by staining with GelRedTM stain for 2 h and destained for 15 min with TAE buffer. The gel was imaged Via BioRad's Gel DocTMEZ system. Both supercoiled and linear strands DNA were incorporated in the gel as markers for DNA-Topo II intercalators. The results of IC50 values were calculated using the GraphPad Prism version 5.0. Each reaction was performed in duplicate, and at least three independent determinations of each IC₅₀ were made.