Electronic Supplementary Material (ESI) for RSC Advances.

Rationale Design and Synthesis of Novel Apoptotic Thiadiazole Analogues Targeting VEGFR-2: Computational and *In vitro* studies

Walid E. Elgammal[a], Hazem Elkady[b], Hazem A. Mahdy[b], Dalal Z. Husein[c], Aisha A. Alsfouk[d], Bshra A. Alsfouk[d], Ibrahim M. Ibrahim[e], Eslam B. Elkaeed[f], Ahmed M. Metwaly[g,h*], Ibrahim H. Eissa[b*]

a Department of Chemistry, Faculty of Science, Al-Azhar University, Nasr City, Cairo, Egypt
[b] Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo, 11884, Egypt.

[c] Chemistry Department, Faculty of Science, New Valley University, El-Kharja 72511, Egypt.

[d] Department of Pharmaceutical Sciences, College of Pharmacy, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia.

[e] Biophysics Department, Faculty of Science, Cairo University. Cairo 12613, Egypt.

[f] Department of Pharmaceutical Sciences, College of Pharmacy, AlMaarefa University, Riyadh 13713, Saudi Arabia.

[g] Pharmacognosy and Medicinal Plants Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt.
[h] Biopharmaceutical Products Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications (SRTA-City), Alexandria, Egypt.

***Corresponding authors:**
Ibrahim H. Eissa
Medicinal Chemistry Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt**.** **Email:** Ibrahimeissa@azhar.edu.eg
Ahmed M. Metwaly
Pharmacognosy and Medicinal Plants Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt**.** **Email:** ametwaly@azhar.edu.eg
Eslam B. Elkaeed: Department of Pharmaceutical Sciences, College of Pharmacy, AlMaarefa University, Riyadh 13713, Saudi Arabia. **Email:** ekaeed@um.edu.sa

Content
S1. Chemistry
S2. Biological testing
S.3. *In silico* studies
S.4. Spectral data

Lab code	Paper code
WA1	**14**
PH1	**18a**
PH4	**18b**
WC1	**18c**
WC4	**18d**
WM1	**18e**
WM4	**18f**

S.1. Chemistry

The chemicals, reagents, and reaction solvents used in this study were gained from Sigma-Aldrich, Alpha Chem, Fluka, and Loba and were not purified further. The melting points (mp), which have not been corrected, were measured using the SMP50 Digital Melting Point App provided by Bibby Scientific in Staffordshire. A Thermo Fisher Nicolet IS10 spectrophotometer was used to detect infrared spectra as solids on the potassium bromide disc (ν_{max} in cm^{-1}) with a resolution of 4.0 cm^{-1}, covering 4000-400 cm^{-1}.^{1}H-NMR and ^{13}C-NMR spectra (400 and 101 MHz) were recorded at the JNM-ECA 500 II Made by JEOL-JAPAN instrument through a solution of deuterated dimethyl sulfoxide. Proton chemical shifts are labeled in part per million (ppm), downfield from tetramethyl silane (TMS, δ=0) as an internal standard, and the following abbreviations (or a combination thereof) are used to describe splitting patterns: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet, and br, broad. The internal references used were the residual protons (2.50, 3.33 ppm for ^{1}H NMR, and 39.9 ppm for ^{13}C NMR). Mass spectra were measured with a Thermo Scientific

GCMS model (Isq Lt) using the Thermo X-Calibur software (Shimadzu, Kyoto, Japan) at the Regional Center for Mycology and Biotechnology (RCMB), Al-Azhar University, Nasr City, Cairo, Egypt. Elemental studies were conducted at the Regional Center for Microbiology and Biotechnology, Al-Azhar University, Cairo, Egypt, with results accurate to within 0.4%. Thin-layer chromatography (TLC) was carried out on silica gel plates by using DCM: MeOH (95:5%), as the eluting system. The progress of the reaction and evaluation of product purity was determined using a UV indicator at 254 nm.

S2. Biological testing

S.2.1. *In vitro* anti-proliferative activity

S.2.1.1. Determination of sample cytotoxicity on cells (MTT protocol)

1-The 96 well tissue culture plate was inoculated with 1 X 10^5 cells / ml (100 ul / well) and incubated at 37°C for 24 hours to develop a complete monolayer sheet.

2- Growth medium was decanted from 96 well micro titer plates after confluent sheet of cells were formed, cell monolayer was washed twice with wash media.

3- two-fold dilutions of tested sample was made in RPMI medium with 2% serum (maintenance medium).

4- 0.1 ml of each dilution was tested in different wells leaving 3 wells as control, receiving only maintenance medium.

5- Plate was incubated at 37°C and examined. Cells were checked for any physical signs of toxicity, e.g. partial or complete loss of the monolayer, rounding, shrinkage, or cell granulation.

6- MTT solution was prepared (5mg/ml in PBS) (BIO BASIC CANADA INC).

8- 20ul MTT solution were added to each well. Place on a shaking table, 150rpm for 5 minutes, to thoroughly mix the MTT into the media.

9) Incubate (37C, 5% CO2) for 4 hours to allow the MTT to be metabolized.

10) Dump off the media. (dry plate on paper towels to remove residue if necessary.

11) Resuspend formazan (MTT metabolic product) in 200ul DMSO. Place on a shaking table, 150rpm for 5 minutes, to thoroughly mix the formazan into the solvent.

12) Read optical density at 560nm and subtract background at 620nm. Optical density should be directly correlated with cell quantity.

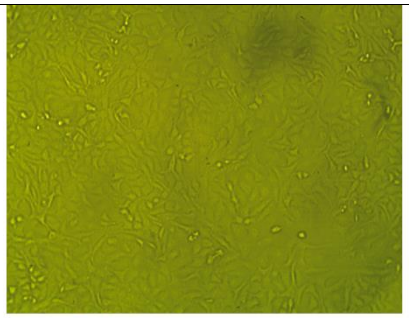
S.2.1.2. Morphological assay

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

❖ Damage can be identified by large decreases in volume secondary to losses in protein and intracellular ions of due to altered permeability to sodium or potassium.

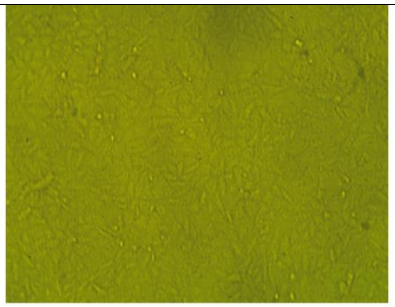
❖ Necrotic cells: nuclear swelling, chromatin flocculation, loss of nuclear basophilia

❖ Apoptotic cells: cell shrinkage, nuclear condansation, nuclear fragmentation



control
Mcf7 cells

Organism : *Homo sapiens*, human
Tissue : mammary gland, breast; derived from metastatic site: pleural effusion
Cell Type : epithelial
Culture Properties : adherent
Disease : adenocarcinoma
ATCC : HTB-22



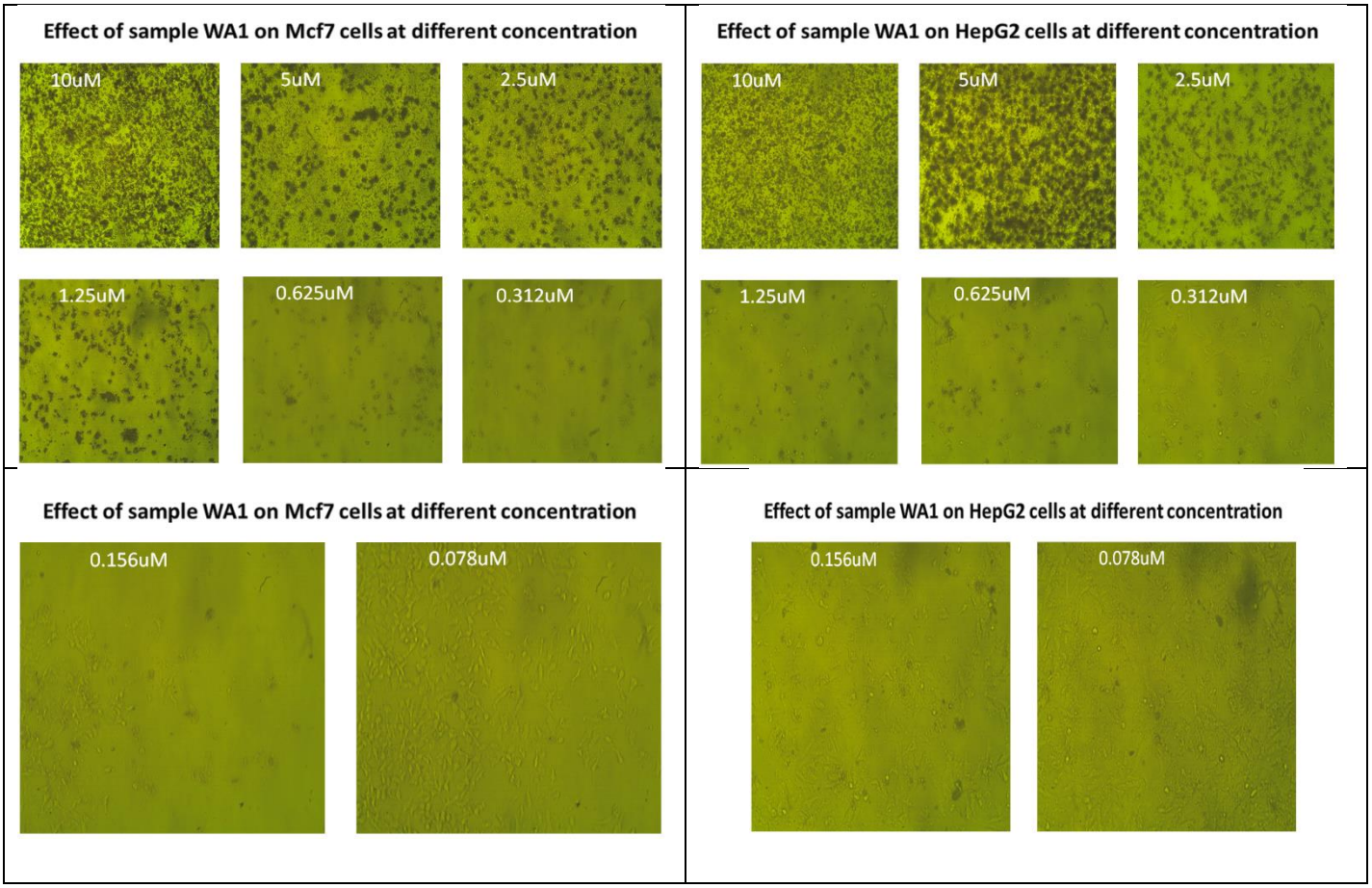
control
HepG2 cells

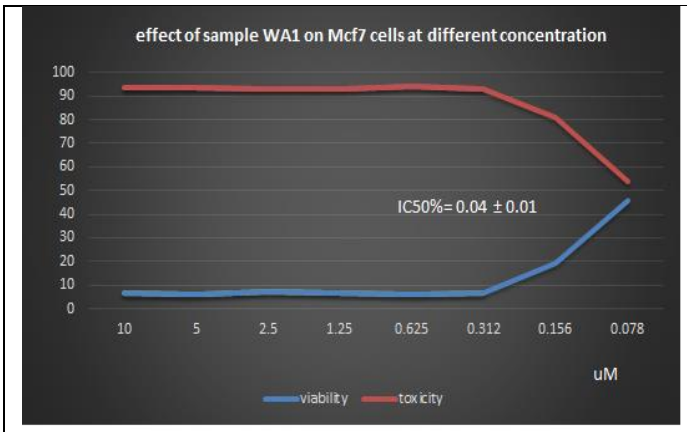
Organism : *Homo sapiens*, human
Tissue : liver
Cell Type : epithelial
Culture Properties : adherent
Disease : hepatocellular carcinoma
ATCC : HB-8065

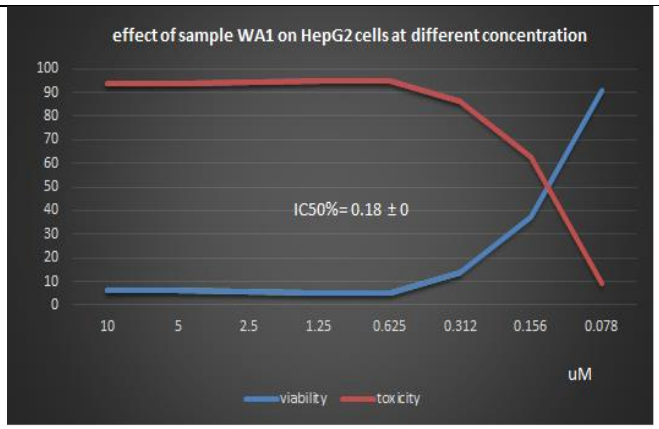
***** Cytotoxicity of WA1 (compound 14) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
MCF-7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
WA1	10	0.053	0.048	0.055	0.052	0.002082	6.727037516	93.27296248	0.04 ± 0.01
	5	0.05	0.05	0.047	0.049	0.001	6.338939198	93.6610608	
	2.5	0.056	0.055	0.051	0.054	0.001528	6.985769728	93.01423027	
	1.25	0.056	0.049	0.055	0.053333	0.002186	6.899525658	93.10047434	
	0.625	0.048	0.042	0.048	0.046	0.002	5.95084088	94.04915912	
	0.312	0.049	0.055	0.053	0.052333	0.001764	6.770159552	93.22984045	
	0.156	0.163	0.152	0.133	0.149333	0.008762	19.31867184	80.68132816	
	0.078	0.367	0.342	0.359	0.356	0.007371	46.05433376	53.94566624	
	5	0.173	0.156	0.162	0.163667	0.004978	21.17291936	78.82708064	
	2.5	0.263	0.269	0.271	0.267667	0.002404	34.62699439	65.37300561	
	1.25	0.388	0.367	0.393	0.382667	0.007965	49.50409659	50.49590341	
	0.625	0.659	0.678	0.666	0.667667	0.005548	86.37343683	13.62656317	
	0.312	0.771	0.764	0.778	0.771	0.004041	99.74126779	0.258732212	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
WA1	10	0.043	0.047	0.044	0.044667	0.001202	6.08537693	93.91462307	0.18 ± 0
	5	0.045	0.046	0.045	0.045333	0.000333	6.176203451	93.82379655	
	2.5	0.042	0.039	0.044	0.041667	0.001453	5.676657584	94.32334242	
	1.25	0.037	0.041	0.036	0.038	0.001528	5.177111717	94.82288828	
	0.625	0.037	0.039	0.038	0.038	0.000577	5.177111717	94.82288828	
	0.312	0.11	0.093	0.106	0.103	0.005132	14.03269755	85.96730245	
	0.156	0.289	0.261	0.277	0.275667	0.00811	37.55676658	62.44323342	
	0.078	0.683	0.651	0.669	0.667667	0.009262	90.96276113	9.037238874	



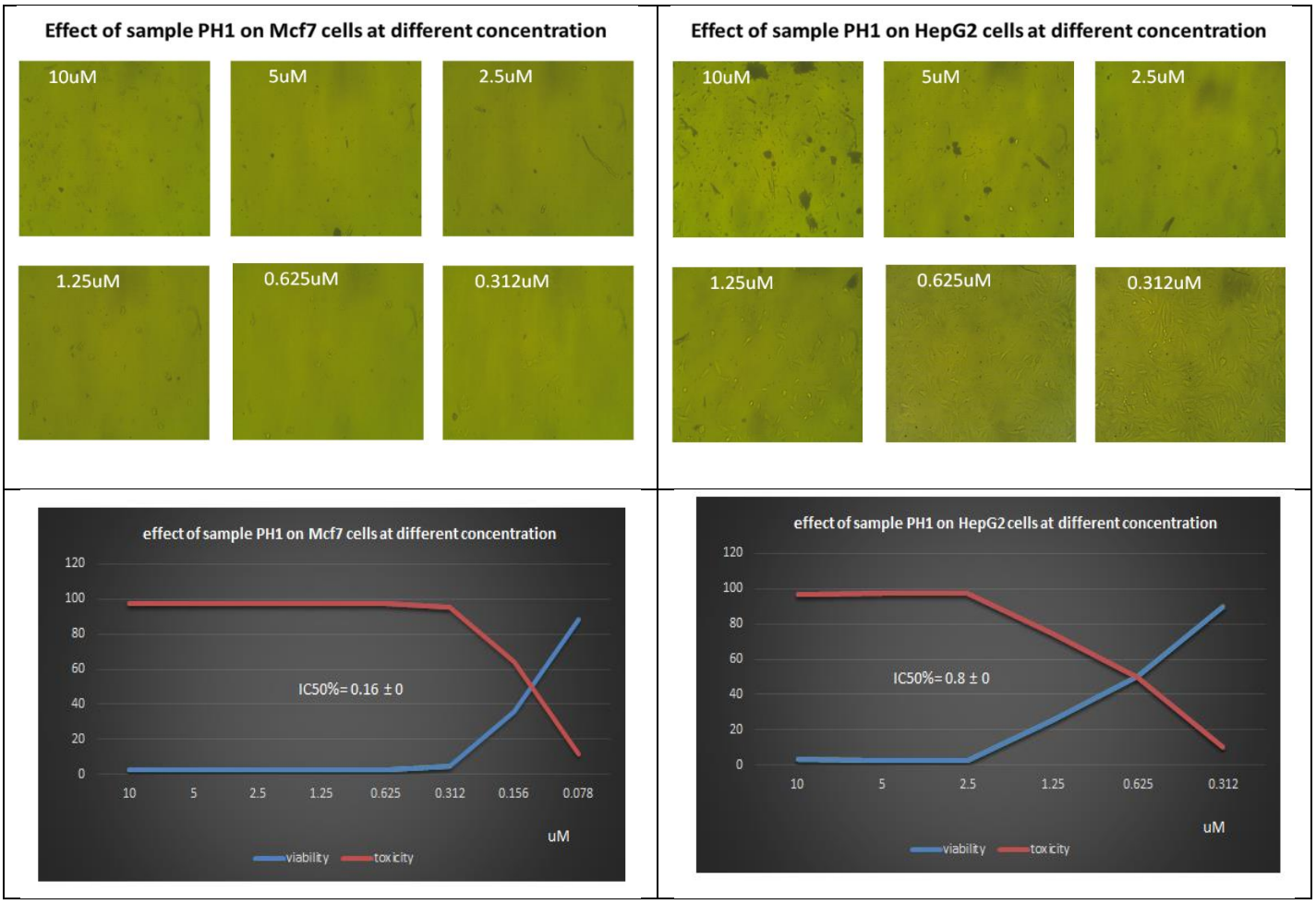




***** Cytotoxicity of PH-1 (compound 18a) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
MCF-7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
PH1	10	0.02	0.018	0.019	0.019	0.000577	2.457956016	97.54204398	0.16 ± 0
	5	0.019	0.018	0.021	0.019333	0.000882	2.501078051	97.49892195	
	2.5	0.019	0.022	0.018	0.019667	0.001202	2.544200086	97.45579991	
	1.25	0.02	0.024	0.019	0.021	0.001528	2.716688228	97.28331177	
	0.625	0.018	0.023	0.022	0.021	0.001528	2.716688228	97.28331177	
	0.312	0.035	0.042	0.039	0.038667	0.002028	5.002156102	94.9978439	
	0.156	0.264	0.277	0.293	0.278	0.008386	35.96377749	64.03622251	
	0.078	0.673	0.7	0.682	0.685	0.007937	88.61578266	11.38421734	

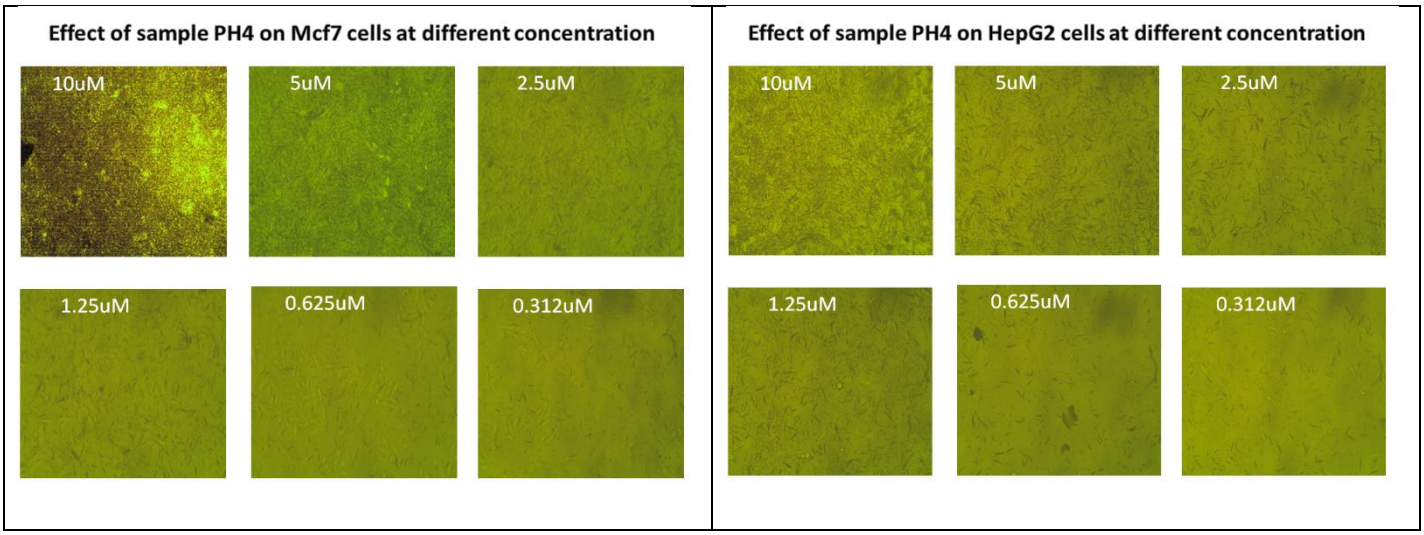
ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
PH1	10	0.024	0.022	0.023	0.023	0.000577	3.133514986	96.86648501	0.8 ± 0
	5	0.022	0.019	0.026	0.022333	0.002028	3.042688465	96.95731153	
	2.5	0.02	0.02	0.021	0.020333	0.000333	2.770208901	97.2297911	
	1.25	0.183	0.189	0.192	0.188	0.002646	25.61307902	74.38692098	
	0.625	0.385	0.362	0.358	0.368333	0.008413	50.18165304	49.81834696	
	0.312	0.668	0.652	0.648	0.656	0.00611	89.373297	10.626703	

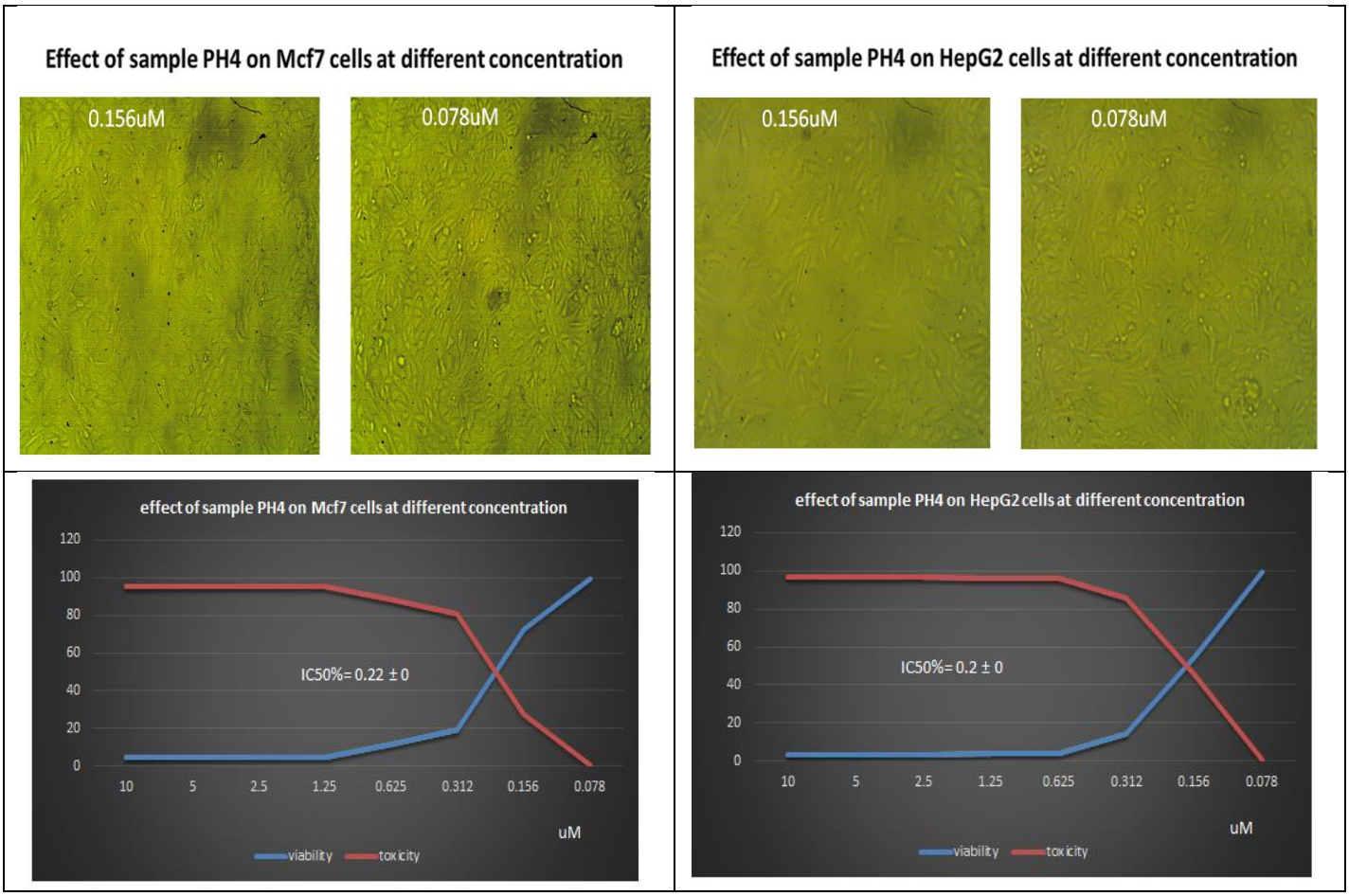


***** Cytotoxicity of PH-4 (compound 18b) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
MCF-7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
PH4	10	0.036	0.038	0.038	0.037433	0.000722	4.842604571	95.15739543	0.22 ± 0
	5	0.039	0.038	0.039	0.038667	0.000333	5.002156102	94.9978439	
	2.5	0.03	0.036	0.038	0.034667	0.002404	4.484691677	95.51530832	
	1.25	0.036	0.035	0.036	0.035667	0.000333	4.614057784	95.38594222	
	0.625	0.085	0.074	0.11	0.089667	0.010651	11.59982751	88.40017249	
	0.312	0.163	0.128	0.154	0.148333	0.010493	19.18930574	80.81069426	
	0.156	0.584	0.542	0.561	0.562333	0.012143	72.74687365	27.25312635	
	0.078	0.77	0.765	0.774	0.769667	0.002603	99.56877965	0.431220354	
	5	0.034	0.035	0.033	0.034	0.000577	4.398447607	95.60155239	
	2.5	0.034	0.034	0.036	0.034667	0.000667	4.484691677	95.51530832	
	1.25	0.032	0.036	0.034	0.034	0.001155	4.398447607	95.60155239	
	0.625	0.03	0.028	0.03	0.029333	0.000667	3.794739112	96.20526089	
	0.312	0.032	0.036	0.028	0.032	0.002309	4.139715395	95.86028461	
	0.156	0.177	0.19	0.164	0.177	0.007506	22.89780078	77.10219922	
	0.078	0.38	0.362	0.389	0.377	0.007937	48.77102199	51.22897801	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
PH4	10	0.025	0.028	0.026	0.026333	0.000882	3.587647593	96.41235241	0.2 ± 0
	5	0.023	0.025	0.021	0.023	0.001155	3.133514986	96.86648501	
	2.5	0.026	0.023	0.025	0.024667	0.000882	3.36058129	96.63941871	
	1.25	0.035	0.032	0.03	0.032333	0.001453	4.405086285	95.59491371	
	0.625	0.028	0.026	0.029	0.027667	0.000882	3.769300636	96.23069936	
	0.312	0.099	0.116	0.105	0.106667	0.004978	14.53224342	85.46775658	
	0.156	0.413	0.386	0.394	0.397667	0.008007	54.17801998	45.82198002	
	0.078	0.736	0.73	0.727	0.731	0.002646	99.59128065	0.408719346	
	5	0.03	0.027	0.027	0.028	0.001	3.814713896	96.1852861	
	2.5	0.021	0.02	0.022	0.021	0.000577	2.861035422	97.13896458	
	1.25	0.02	0.019	0.024	0.021	0.001528	2.861035422	97.13896458	
	0.625	0.023	0.021	0.022	0.022	0.000577	2.997275204	97.0027248	
	0.312	0.022	0.021	0.026	0.023	0.001528	3.133514986	96.86648501	
	0.156	0.264	0.289	0.291	0.281333	0.008686	38.32879201	61.67120799	
	0.078	0.673	0.642	0.668	0.661	0.009609	90.05449591	9.945504087	

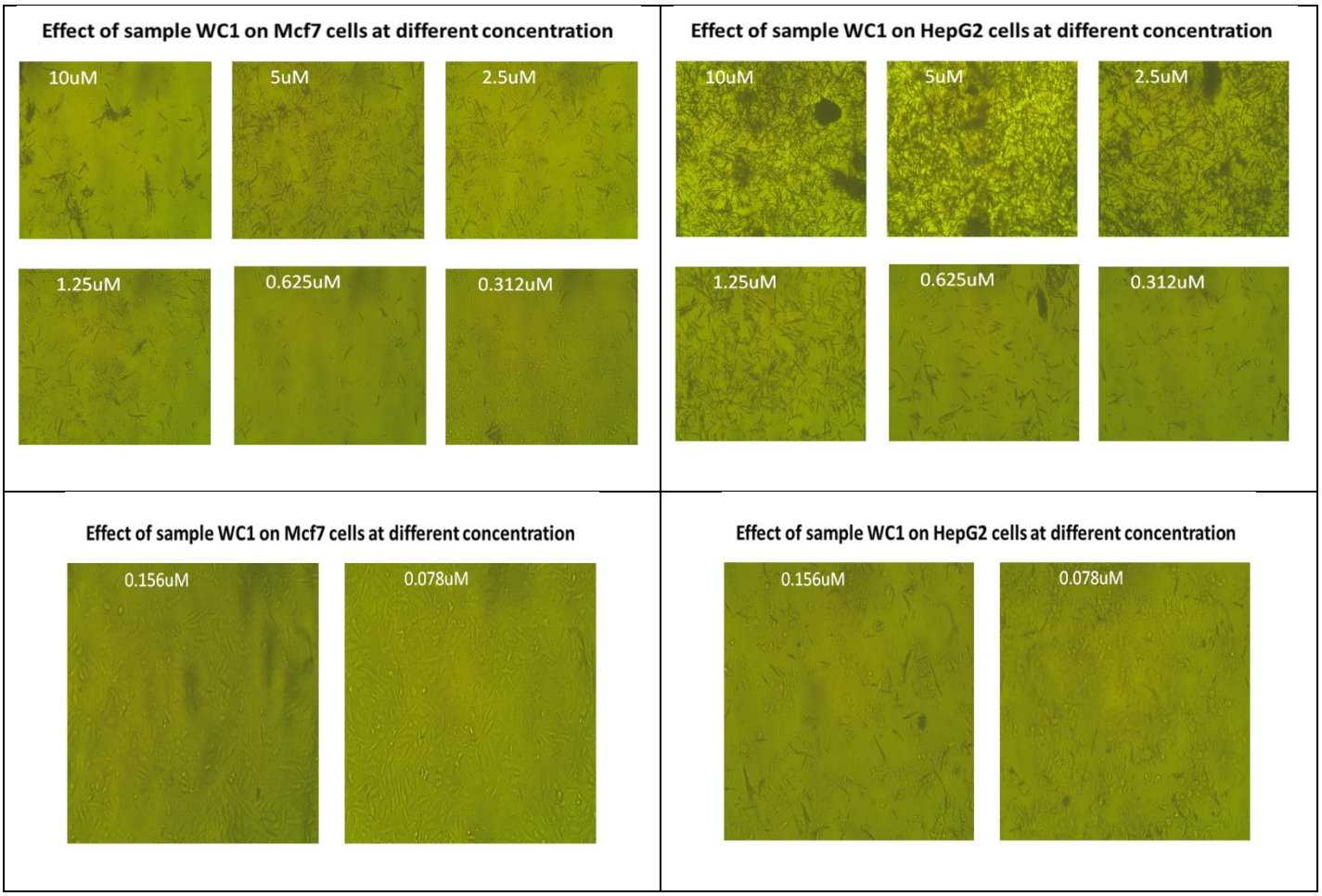


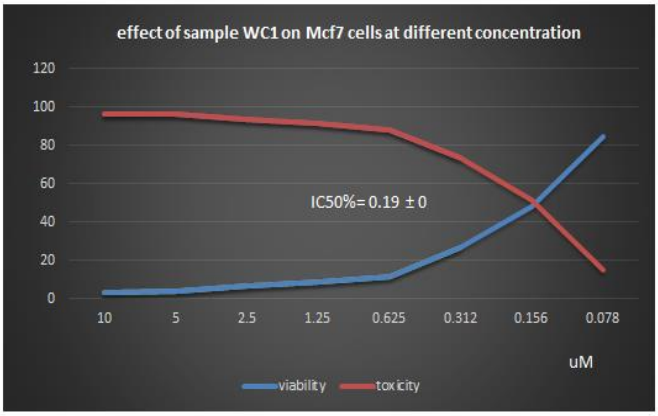


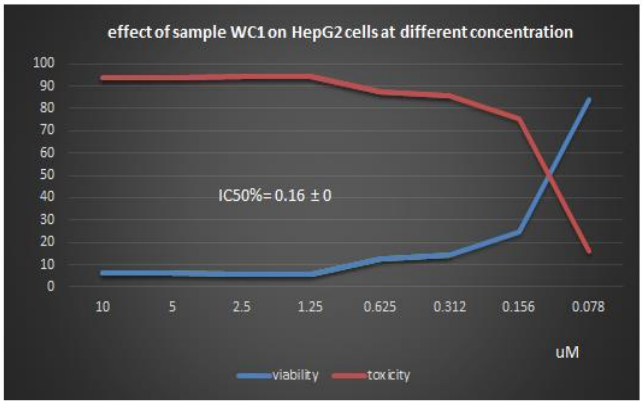
***** Cytotoxicity of WC-1 (compound 18c) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
MCF-7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
WC1	10	0.027	0.028	0.026	0.027	0.000577	3.492884864	96.50711514	0.19 ± 0
	5	0.029	0.027	0.03	0.028667	0.000882	3.708495041	96.29150496	
	2.5	0.052	0.048	0.05	0.05	0.001155	6.468305304	93.5316947	
	1.25	0.083	0.059	0.062	0.068	0.00755	8.796895213	91.20310479	
	0.625	0.094	0.088	0.091	0.091	0.001732	11.77231565	88.22768435	
	0.312	0.198	0.216	0.201	0.205	0.005568	26.52005175	73.47994825	
	0.156	0.365	0.381	0.374	0.373333	0.004631	48.2966796	51.7033204	
	0.078	0.655	0.639	0.672	0.655333	0.009528	84.77792152	15.22207848	
	0.312	0.768	0.777	0.771	0.772	0.002646	99.87063389	0.129366106	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
WC1	10	0.044	0.048	0.043	0.045	0.001528	6.130790191	93.86920981	0.16 ± 0
	5	0.045	0.046	0.044	0.045	0.000577	6.130790191	93.86920981	
	2.5	0.043	0.041	0.045	0.043	0.001155	5.858310627	94.14168937	
	1.25	0.046	0.04	0.04	0.042	0.002	5.722070845	94.27792916	
	0.625	0.083	0.092	0.099	0.091333	0.004631	12.44323342	87.55676658	
	0.312	0.108	0.094	0.109	0.103667	0.004842	14.12352407	85.87647593	
	0.156	0.196	0.165	0.188	0.183	0.009292	24.93188011	75.06811989	
	0.078	0.633	0.598	0.619	0.616667	0.010171	84.01453224	15.98546776	



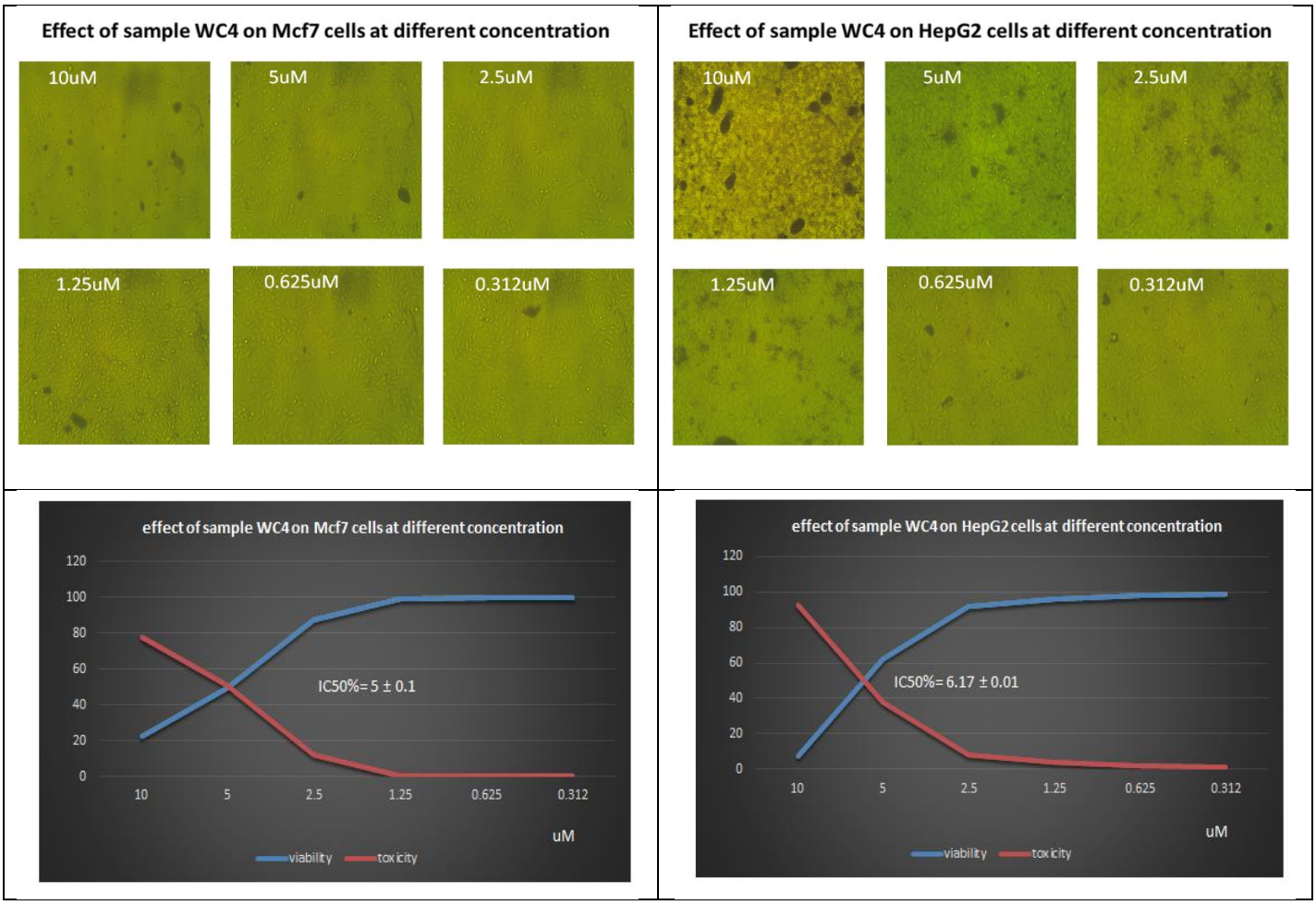




***** Cytotoxicity of WC-4 (compound 18d) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
MCF-7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
WC4	10	0.17	0.163	0.189	0.174	0.007767	22.50970246	77.49029754	5 ± 0.1
	5	0.388	0.364	0.391	0.381	0.008544	49.28848642	50.71151358	
	2.5	0.688	0.663	0.679	0.676667	0.007311	87.53773178	12.46226822	
	1.25	0.77	0.763	0.769	0.767333	0.002186	99.2669254	0.733074601	
	0.625	0.775	0.77	0.771	0.772	0.001528	99.87063389	0.129366106	
	0.312	0.768	0.777	0.771	0.772	0.002646	99.87063389	0.129366106	

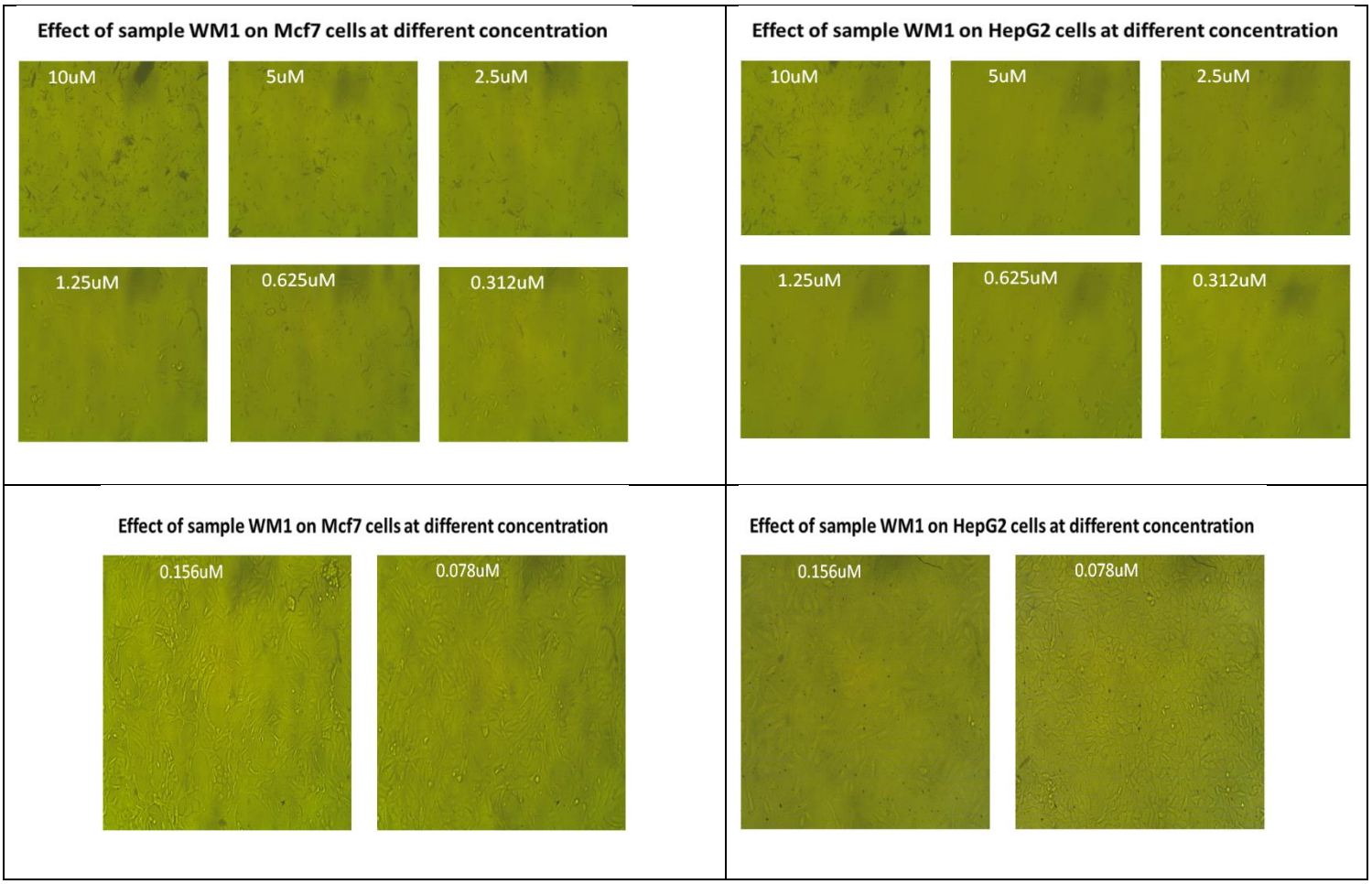
ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
WC4	10	0.063	0.048	0.052	0.054333	0.004485	7.40236149	92.59763851	6.17 ± 0.01
	5	0.463	0.442	0.459	0.454667	0.006438	61.94368756	38.05631244	
	2.5	0.673	0.682	0.67	0.675	0.003606	91.96185286	8.038147139	
	1.25	0.698	0.714	0.704	0.705333	0.004667	96.09445958	3.905540418	
	0.625	0.724	0.722	0.715	0.720333	0.002728	98.13805631	1.861943688	
	0.312	0.73	0.717	0.724	0.723667	0.003756	98.59218892	1.407811081	

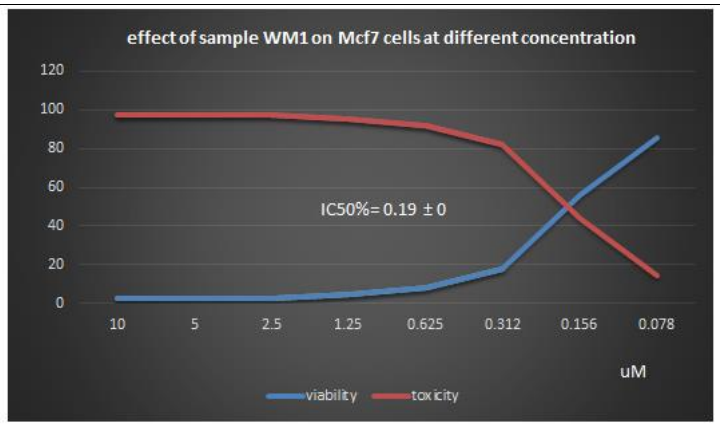


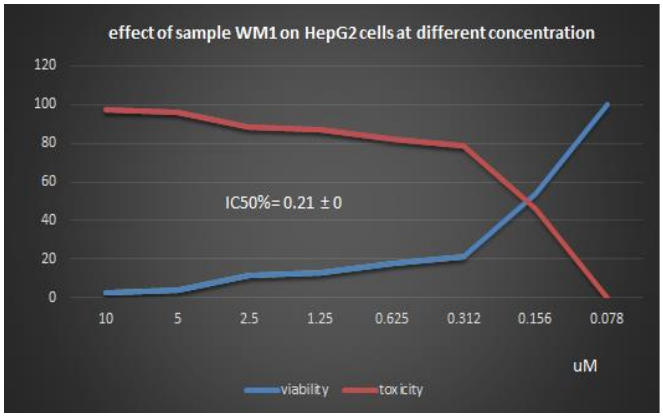
***** Cytotoxicity of WM-1 (compound 18e) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
MCF-7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
WM1	10	0.024	0.022	0.021	0.022333	0.000882	2.889176369	97.11082363	0.19 ± 0
	5	0.022	0.026	0.023	0.023667	0.001202	3.061664511	96.93833549	
	2.5	0.022	0.021	0.026	0.023	0.001528	2.97542044	97.02457956	
	1.25	0.043	0.036	0.032	0.037	0.003215	4.786545925	95.21345408	
	0.625	0.088	0.092	0.01	0.063333	0.026692	8.193186718	91.80681328	
	0.312	0.151	0.127	0.133	0.137	0.007211	17.72315653	82.27684347	
	0.156	0.437	0.413	0.452	0.434	0.011358	56.14489004	43.85510996	
	0.078	0.674	0.668	0.641	0.661	0.010149	85.51099612	14.48900388	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
WM1	10	0.019	0.018	0.022	0.019667	0.001202	2.67938238	97.32061762	0.21 ± 0
	5	0.026	0.037	0.03	0.031	0.003215	4.223433243	95.77656676	
	2.5	0.084	0.091	0.088	0.087667	0.002028	11.94368756	88.05631244	
	1.25	0.093	0.099	0.103	0.098333	0.002906	13.3969119	86.6030881	
	0.625	0.145	0.128	0.12	0.131	0.007371	17.84741144	82.15258856	
	0.312	0.166	0.148	0.152	0.155333	0.005457	21.16257947	78.83742053	
	0.156	0.397	0.382	0.406	0.395	0.007	53.8147139	46.1852861	
	0.078	0.736	0.73	0.732	0.732667	0.001764	99.81834696	0.181653043	





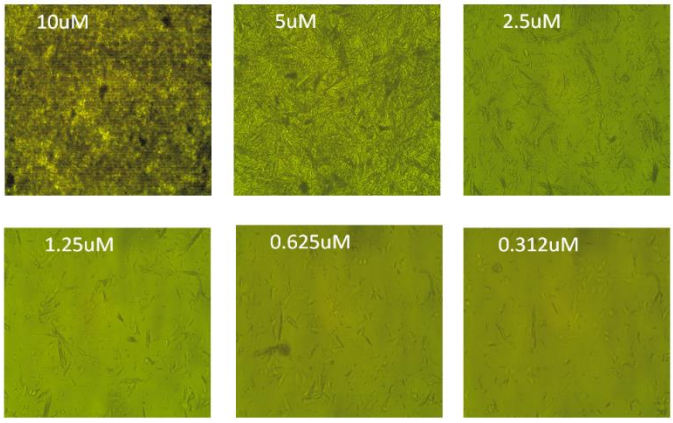


***** Cytotoxicity of WM-4 (compound 18f) against MCF-7 and HepG2**

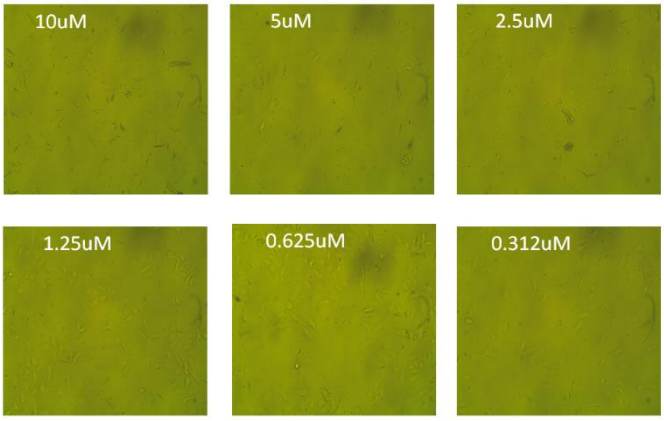
ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
MCF-7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
WM4	10	0.046	0.037	0.038	0.040333	0.002848	5.217766279	94.78223372	0.44 ± 0
	5	0.039	0.042	0.041	0.040667	0.000882	5.260888314	94.73911169	
	2.5	0.064	0.059	0.061	0.061333	0.001453	7.934454506	92.06554549	
	1.25	0.116	0.142	0.158	0.138667	0.012238	17.93876671	82.06123329	
	0.625	0.198	0.202	0.214	0.204667	0.004807	26.47692971	73.52307029	
	0.312	0.453	0.473	0.461	0.462333	0.005812	59.81026304	40.18973696	
	0.156	0.766	0.778	0.771	0.771667	0.00348	99.82751186	0.172488141	
	0.078	0.66	0.684	0.666	0.67	0.007211	86.67529107	13.32470893	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC_{50} ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
WM4	10	0.019	0.02	0.017	0.018667	0.000882	2.543142598	97.4568574	0.18 ± 0.01
	5	0.026	0.021	0.022	0.023	0.001528	3.133514986	96.86648501	
	2.5	0.02	0.018	0.024	0.020667	0.001764	2.815622162	97.18437784	
	1.25	0.142	0.118	0.137	0.132333	0.007311	18.02906449	81.97093551	
	0.625	0.183	0.155	0.138	0.158667	0.013119	21.61671208	78.38328792	
	0.312	0.214	0.183	0.199	0.198667	0.00895	27.06630336	72.93369664	
	0.156	0.331	0.308	0.356	0.331667	0.01386	45.18619437	54.81380563	
	0.078	0.593	0.563	0.585	0.580333	0.008969	79.06448683	20.93551317	

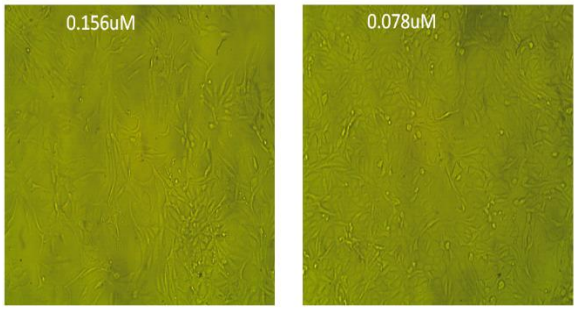
Effect of sample WM4 on Mcf7 cells at different concentration



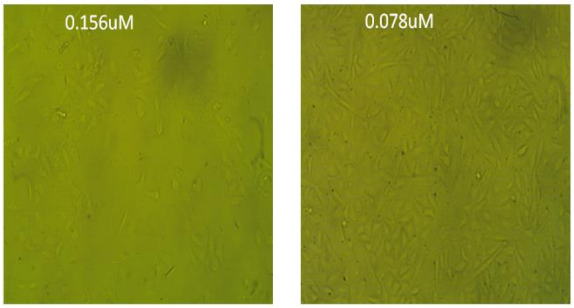
Effect of sample WM4 on HepG2 cells at different concentration

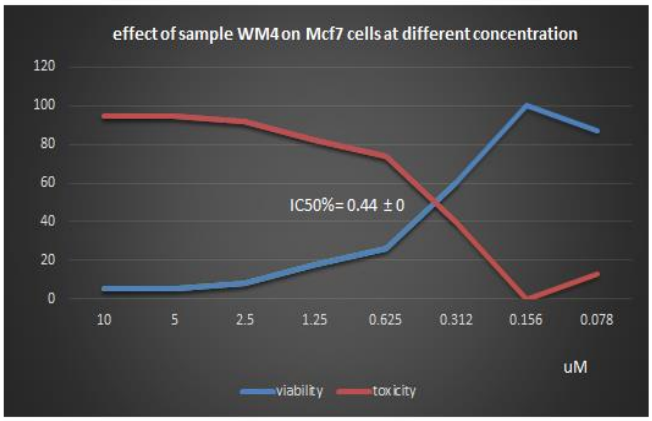


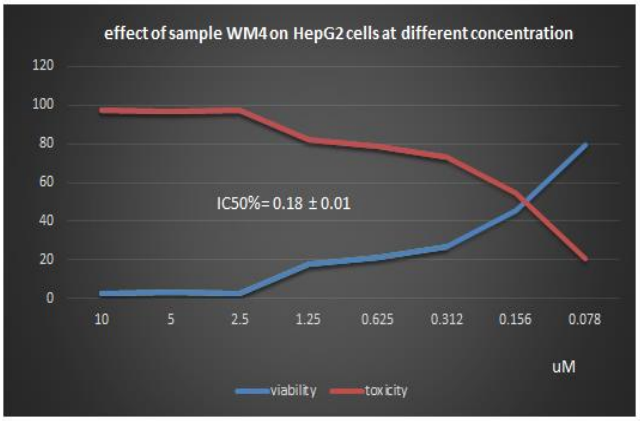
Effect of sample WM4 on Mcf7 cells at different concentration



Effect of sample WM4 on HepG2 cells at different concentration





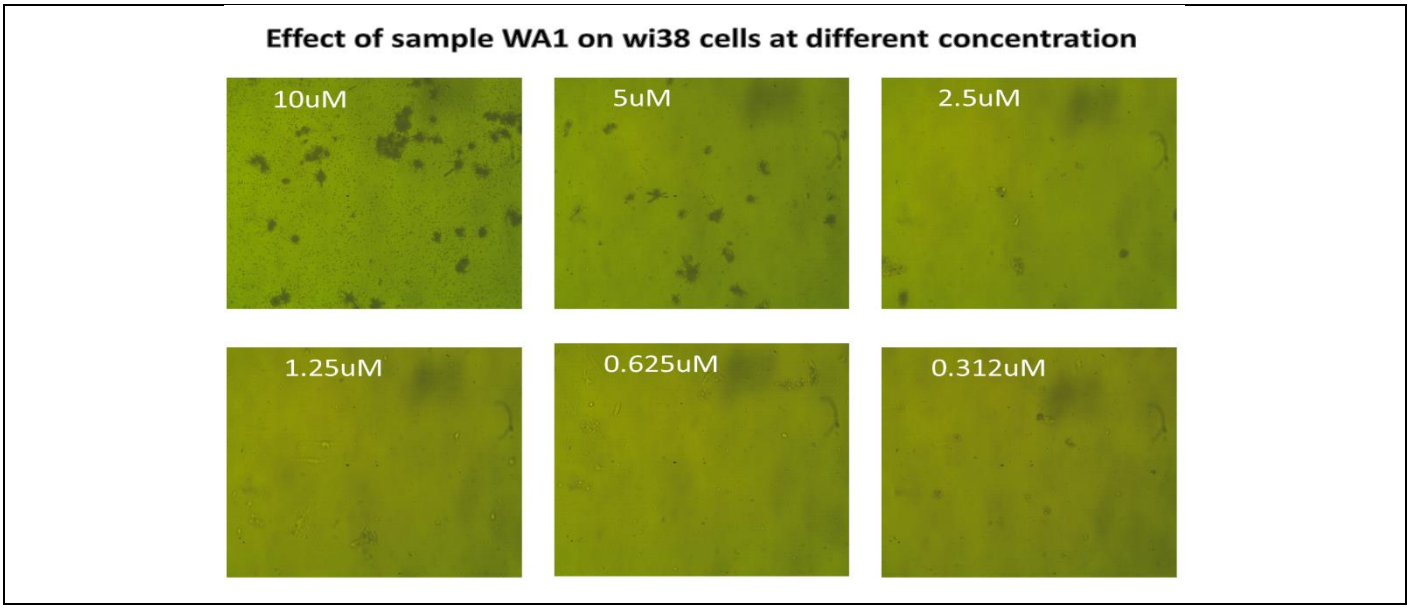


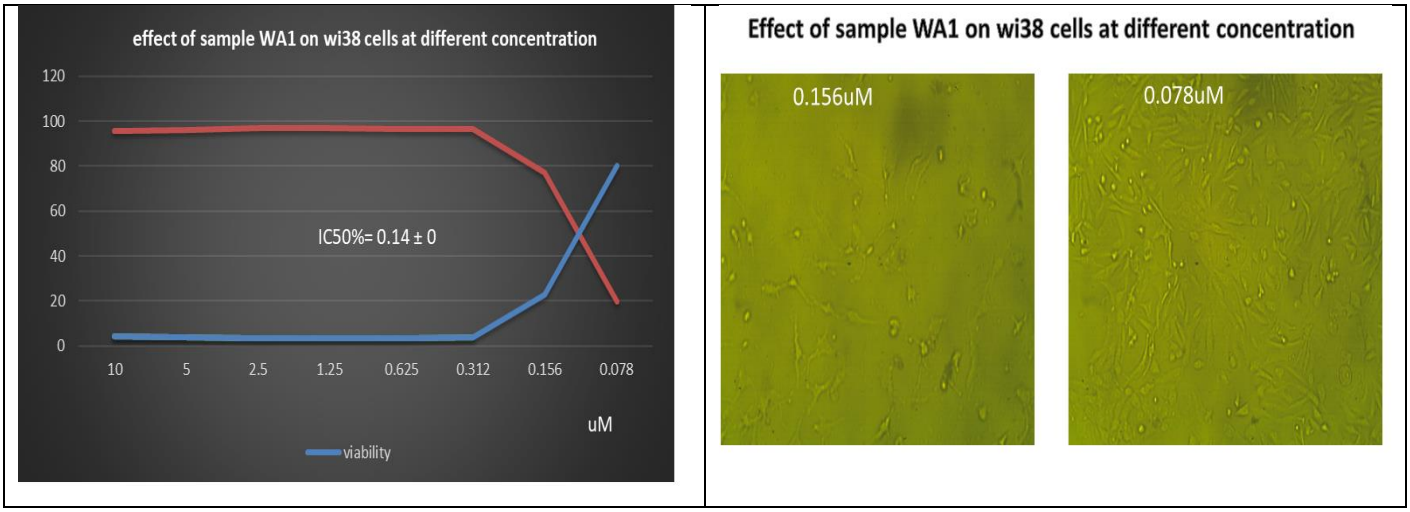
S.2.2. Safety assay

The safety profile of compound **14** was checked on against normal human lung cells (WI-38) to determine the treatments concentrations that do not depict toxic effects against the tested cells. A portion of 100.0 µl of 6×10^4 cell/ml cells was seeded into each well of a 96-well plate and then the plates were incubated at 37°C in a humidified 5.0% CO_2 incubator for 24 h. At the end of incubation period, the exhausted medium was replaced with 100.0 µl of different concentrations of the designated treatment (prepared in RPMI medium starting from 1.0 mM). The inoculated plates were incubated at the same growth conditions for another 24 h. At the end of incubation, cellular viability was assessed using MTS assay kit (Promega) according to the manual instruction.

***** Cytotoxicity of WA-1 (compound 14) against normal human lung cells (WI-38)**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
Wi38	--------	0.672	0.67	0.668	0.67	0.001155	100	0	uM
WA1	10	0.022	0.035	0.03	0.029	0.003786	4.328358209	95.67164179	0.14 ± 0
	5	0.026	0.025	0.029	0.026667	0.001202	3.980099502	96.0199005	
	2.5	0.021	0.02	0.023	0.021333	0.000882	3.184079602	96.8159204	
	1.25	0.02	0.02	0.025	0.021667	0.001667	3.233830846	96.76616915	
	0.625	0.021	0.026	0.023	0.023333	0.001453	3.482587065	96.51741294	
	0.312	0.026	0.023	0.025	0.024667	0.000882	3.68159204	96.31840796	
	0.156	0.152	0.177	0.132	0.153667	0.013017	22.93532338	77.06467662	
	0.078	0.534	0.551	0.528	0.537667	0.006888	80.24875622	19.75124378	

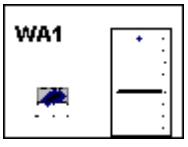


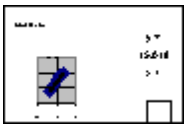


S.2.3. *In vitro* VEGFR-2 inhibition

Inhibitory activity of compound **14** against VEGFR-2 was evaluated using Human VEGFR-2 ELISA kit (VEGFR2(KDR) Kinase Assay Kit Catalog # 40325)(BPS Bioscience, San Diego, USA). A specific antibody for VEGFR-2 was seeded on a 96-well plate and 100 µL of the standard solution or the tested compound was added, all were incubated at room temperature for 2.5 hours. Then washed, 100 µL of the prepared biotin antibody was added, then incubated at room temperature for additional 1 hour. Washed, 100 µL of streptavidin solution was added then incubated for 45 min. at room temperature. Washed again, 100 µL of TMB Substrate reagent was added and incubated for 30 min. at room temperature. 50 µL of the stop solution was added, then read at 450 nm immediately. The standard curve was drawn, concentrations on the X-axis and the absorbance on the Y-axis.

VEGFR2

code	IC50	conc	log	%inh	T2	T1	ΔT	RFU2	RFU1	ΔRFU	slope	K.Activity
WA1		100	2	95.3	30	0	30	4.67	0	4.67	3.333	5.604
		10	1	88.2	30	0	30	11.84	0	11.84	3.333	14.208
		1	0	71.8	30	0	30	28.22	0	28.22	3.333	33.864
		0.1	-1	46.5	30	0	30	53.48	0	53.48	3.333	64.176
		0.01	-2	31.6	30	0	30	68.41	0	68.41	3.333	82.092
EC				0	30	0	30	100	0	100	3.333	120
code	IC50	conc	log	%inh	T2	T1	ΔT	RFU2	RFU1	ΔRFU	slope	K.Activity

Sorafenib	100	2	95.7	30	0	30	4.29	0	4.29	3.333	5.148
	10	1	89	30	0	30	11.03	0	11.03	3.333	13.236
	1	0	75.8	30	0	30	24.16	0	24.16	3.333	28.992
	0.1	-1	57.4	30	0	30	42.55	0	42.55	3.333	51.06
	0.01	-2	36.4	30	0	30	63.62	0	63.62	3.333	76.344
EC			0	30	0	30	100	0	100	3.333	120

S.2.4. Cell cycle analysis

The MCF-7 cells were distributed at a density of 2.5×10^5 cells per flask on to the tissue culture flasks (T-75 flask; Nunc A/S) containing RPMI-1640 medium supplemented with 2 % FBS. After incubation for 72h with compound **14** at a concentration of 0.041 µM, the cells were washed thrice with ice-cold PBS buffer and fixed in 70 % cold ethyl alcohol at −20 °C overnight. The contents from each of the flask were then centrifuged at 12,000×*g* for 20 min followed by washing and treatment with 100 µl PBS containing 1 mM RNase A (Qiagen, Hilden, Germany). After incubation for 45 min at 37 °C, the cells were stained with 400 µl propidium iodide (PI; 50 µg/ml; Sigma-Aldrich) for 30 min. The cells were filtered through a 40-µm nylon mesh (BD Biosciences, San Jose, CA, USA), and the DNA content was analyzed using Epics XL-MCL™ Flow Cytometer.

S.2.5. Apoptosis analysis

The annexin V externalization assay for apoptosis was performed using flow cytometry after treatment MCF-7 cells with compound **14** (at concentrations of 0.041 µM) for 72 h as described in the vendor's protocol (BD Pharmingen, BD Biosciences, San Jose, USA). Briefly, untreated and cells treated with compound **15**, from 60 to 70% confluent plates were trypsinized, washed in PBS and resuspended (1×10^6 cells/ml) in binding buffer (10 mM HEPES, pH 7.4, 140 mM NaCl, 2.5 mM $CaCl_2$). A fraction (100 µl/1×10^5 cells) of the cell suspension was incubated with 5 µl annexin V conjugated to FITC and 5 µl propidium iodide (PI) for 15 min at 25 °C in the dark. 400 µl of binding buffer was added to the suspension and apoptosis was measured immediately using a Becton Dickinson FACScan analyzer as described by Raza *et al*., 2008. The apoptotic cells were estimated as the percentage of cells that stained positive for Annexin V-FITC while remaining impermeable to PI (AV+/PI−). This method also distinguished viable cells (AV−/PI−) and cells undergoing necrosis (AV+/PI+).

S.2.6. *In vitro* assay for BAX and Bcl-2 using RT-PCR Tanique

The molecular anticancer mode of action of compound **14** was investigated by screening their ability to affect the level of BAX and Bcl-2 using specific forward and reverse primers and RTq-PCR technique in MCF-7 cells (chosen as the most sensitive cancer cell line). After cellular treatment, MCF-7 cell line was cultured into 12 well plates (6×10^3 cell/ml) for 24 hour at a concentration of 0.04 µM. After treatment, total RNA extraction was performed using RNA extraction kit (Qiagen, Germany). Then, 1 ug of the obtained RNA was used to synthesize cDNA using cDNA synthesis kit (Promega Corporation, Madison, USA) as recommended by the manufacturer. Simultaneously, GAPDH was used as internal control. The RTq-PCR was done using SYBR Green dye (QuantiTect SYBR Green PCR Kits) and Light Cycler fluorimeter (Bio-RAD S1000 Tm thermal cycler). The PCR cycling program was as follows: 95ºC for 2 min, followed by 40 cycles of 95ºC for 30s, 55ºC for 30 s, and 60ºC for 45s, and finally 60ºC for 5 min.

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

Primers
BAX : F 5'- TCAGGATGCGTCCACCAAGAAG -3',
BAX : R 5'- TGTGTCCACGGCGGCAATCATC-3'.
Bcl-2 : F 5'- ATCGCCCTGTGGATGACTGAGT -3',
Bcl-2 : R 5'- GCCAGGAGAAATCAAACAGAGGC -3'.
GAPDH : F 5'- GTCTCCTCTGACTTCAACAGCG-3'
GAPDH : R 5'- ACCACCCTGTTGCTGTAGCCAA-3'

Sample	Gene Expression		
	Control cells	Test cells	FLD

MCF7

Ser	code	Conc	GAPDH	Bax	ΔCTC	GAPDH	Bax	ΔCTE	ΔΔ CT	2^ ΔΔCT
			HC	TC	TC-HC	HE	TE	TE-HE	ΔCTE-ΔCTC	E=1.872
1	WA1		21.66	33.92	12.26	21.86	31.07	9.21	-3.05	6.7691
5	Control		21.66	33.92	12.26	21.66	33.92	12.26	0	1

Ser	code	Conc	GAPDH	Bcl2	ΔCTC	GAPDH	Bcl2	ΔCTE	ΔΔ CT	2^ ΔΔCT
			HC	TC	TC-HC	HE	TE	TE-HE	ΔCTE-ΔCTC	E=1.872
1	WA1		21.66	26.84	5.18	21.86	29.97	8.11	2.93	0.1593
5	Control		21.66	26.84	5.18	21.66	26.84	5.18	0	1

S.2.7. Wound healing assay (Migration assay)

Rat of migration (RM) = (wi – wf) / t

Wi =average of intial wound width um

Wf= average of final wound width um

T = time span of the assay in hours

Percentage of wound clouser % =$\{ (A_{t=0} - A_{t=\Delta t}) / A_{t=0} \} 100$

$A_{t=0}$ = intial wound area

$A_{t=\Delta t}$ = wound area after n hours

Area difference % = (Ai – Af)

Ai = intial area

Af = final area

MCF-7 cells were grown to 95.0% confluency in a complete DMEM medium and then the wounds were formed using a plastic tip. After washing with pre-warmed PBS, the cells were incubated in the specific medium or compound **14** treatment. After incubation at 37°C and 5.0% CO_2 for 24h, the cells were washed with PBS and the wounds distance was determined as the scratch width of the treated and untreated groups using ImageJ software.

S3. *In silico* studies

S.3.1. Docking studies

Protein Preparation: The crystal structure of VEGFR-2 [PDB ID: 2OH4, resolution: 2.05 Å] was obtained from Protein Data Bank (https://www.rcsb.org). At first, the crystal structure of the VEGFR-2 complexed with the co-crystallized ligand was prepared by removing crystallographic water molecules. Only one chain was retained besides the co-crystallized ligand. The selected protein chain was protonated using the following setting. The used electrostatic functional form was GB/VI with a distance cut-off of 15 Å. The used value of the dielectric constant was 2 with an 80 dielectric constant of the used solvent. The used Van der Waals functional form was 800R3 with a distance cut-off of 10 Å. Then, the energy of the protein chain was minimized using Hamiltonian AM1 implanted in Molecular Operating Environment (MOE 2019 and MMFF94x (Merck molecular force field) for structural optimization. Next, the active site of the target protein was defined for ligand docking and redocking (in case of validation of docking protocol). The active site of the protein was identified as the residues that fall within the 5 Å distance from the perimeter of the co-crystallized ligand.

Ligand Preparation: 2D structures of the synthesized compounds and the standard compound, sorafenib were drawn using ChemBioDraw Ultra 14.0 and saved in MDL-SD file format. The 3D structures of the ligands were protonated, and the structures were optimized by energy minimization using MM2 force-field and 10000 iteration steps of 2 fs. The conformationally optimized ligands were used for docking studies.

Docking Setup and Validation of Docking Protocol: The protein-ligand docking studies were carried out using MOE version 2019. Validation of the docking protocol was carried out by redocking the co-crystallized reference ligand against the isolated pocket of VEGFR-2. The docking protocol was validated by comparing the heavy atoms RMSD value of the re-docked ligand pose with the corresponding co-crystallized reference ligand structure.

The docking setup for the tested compounds was established according to the protocol followed in the validation step. For each docking run, 30 docked solutions were generated using ASE for scoring function and rigid receptor for refinement. The pose with ideal binding mode was selected for further investigations. The docking results were visualized using Discovery Studio (DS) 4.0.

Analysis of the docking results was carried out by comparing the interactions and docking score obtained for the docked ligands with that of the re-docked reference molecule.

S.3.2. MD simulations

To assess the robustness of the VEGFR-2_**14** complex and to compare the structure of the apo and holo protein, a 100 ns classical unbiased MD simulation was run in GROMACS 2021. We utilized the solution builder module of the CHARMM-GUI server to generate the input files. Using the transferable intermolecular potential 3 points (TIP3P) water model, we solvated and centered the apo protein and the docked complex in a cube of 8.3 nm in length with 1 nm of padding, and then neutralized the system with NaCl ions at a concentration of 0.154 M. The VEGFR-2 protein's amino acid parameters, the TIP3P water model, and the neutralizing ions were all obtained with the help of the CHARMM36m force field. Parameters for compound **14** were calculated with the help of the CHARMM general force field (CGenFF).

We utilized GROMACS 2021 to perform the dynamics and used periodic boundary conditions (PBCs) in all three spatial dimensions. The potential energy of the system was minimized so that atomic collisions may be avoided. During the equilibration process, the temperature was brought to 310 K and the pressure was brought to 1 atm. In detail, the minimizing step was set to be converged at 100,000 minimization steps or when the maximum force on any atom was 100 $KJ.mol^{-1}.nm^{-1}$. By using the Velocity Rescale technique, we were able to achieve thermal equilibrium in a canonical (NVT) ensemble. For pressure equilibration, we employed a Berendsen barostat with the isothermal-isobaric (NPT) ensemble. The production run in an NPT ensemble was started for 100 ns with the Nose-Hoover thermostat at 310 degrees Kelvin and the Parrinello-Rahman barostat set at 1 atm. The LINear Constraint Solver (LINCS) was used to impose length constraints on the hydrogen-bonded atoms . We utilized Particle Mesh Ewald (PME) to calculate the electrostatics with a threshold of 1.2 nm. By using a time step of 1 femtosecond during equilibration and 2 femtoseconds during production, the Newtonian equations of motion were integrated using the leap-frog algorithm. One thousand frames were captured at 0.1 ns intervals throughout the simulation. We centered the protein in the middle of the simulation box and made it whole again using the trjconv command and then analyzed it using VMD TK scripts. VEGFR-2 and **14** root mean square deviation (RMSD) values were determined. Root mean square fluctuation (RMSF), the radius of gyration (RoG), the ligand-protein center of mass separation, and the

number of hydrogen bonds were all computed to see their dynamic behavior with time.

MM-GBSA

When calculating the ligand's binding energy, we utilized the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) method implemented in the gmx_MMPBSA program. In addition, decomposition analysis was carried out to calculate the binding energy contributed by each amino acid within 1 nm of the ligand. An ionic strength of 0.154 M and a solvation method (igb) value of 5 were selected. The dielectric constant was set to 1.0 inside and 78.5 outside the structure. The MM-GBSA approach is shown in Equation 1.

$$\Delta G = < G_{complex} - (G_{receptor} + G_{ligand}) > \quad \text{Equation 1}$$

Where < > represents the average of the enclosed free energies of complex, receptor, and ligand over the frames used in the calculation. In our approach, we used the whole trajectory (a total of 1000 frames). Different energy terms can be calculated according to Equations 2 to 6 as follows:

$$\Delta G_{binding} = \Delta H - T\Delta S \quad \text{Equation 2}$$

$$\Delta H = \Delta E_{gas} + \Delta E_{sol} \quad \text{Equation 3}$$

$$\Delta E_{gas} = \Delta E_{ele} + \Delta E_{vdW} \quad \text{Equation 4}$$

$$\Delta E_{solv} = E_{GB} + E_{SA} \quad \text{Equation 5}$$

$$E_{SA} = \gamma.SASA \quad \text{Equation 6}$$

Where:

ΔH is the enthalpy which can be calculated from gas-phase energy (E_{gas}) and solvation-free energy (E_{sol}). $-T\Delta S$ is the entropy contribution to the free binding energy. E_{gas} is composed of electrostatic and van der Waals terms; E_{ele}, E_{vdW}, respectively. E_{sol} can be calculated from the polar solvation energy (E_{GB}) and nonpolar solvation energy (E_{SA}) which is estimated from the solvent-accessible surface area.

ProLIF Analysis

we monitored which amino acids interacted with the ligand and how often using the python package Protein-Ligand Interaction Fingerprints (ProLIF). We next used TTclust to cluster the

trajectories, and the obtained representative frames were used with the Protein-Ligand interaction profiler (PLIP) to extract the interactions as 3D conformations.

Principal Component Analysis:

When the mass-weighted covariance matrix (C) of the positional fluctuations of a subset of atoms in a protein is analyzed using a Principal Component Analysis (PCA), the presence of coordinated motion at large amplitudes in MD trajectories becomes apparent. The alpha carbon atoms of amino acids Glu826:Leu1161 were used in the PCA analysis as a subset of atoms to detect the coordinated motion (15). After equilibration, the protein's configuration in each trajectory was used as the reference structure while performing the analyses that used a single trajectory. In contrast, the last frame from the equilibration of the apo system was chosen as the reference for the concatenated trajectory analyses. The PCA technique, in particular, uses the information revealed by diagonalizing the C matrix to determine the eigenvectors and eigenvalues that define the atomic motions' direction and amplitude. For any given system, the first PC displays the largest motion, whereas subsequent PCs show smaller motions. We were able to analyze the C matrix in GROMACS by utilizing the gmx covar command to diagonalize the matrix and the gmx anaeig command to complete the analysis.

Essential subspace size was determined based on the cumulative eigenvalues with respect to the number of eigenvectors used, where the variance maintained by the selected eigenvectors was shown. Additionally, the scree plot was made by plotting the eigenvalue of each eigenvector against its index number. Moreover, it is well known that the distribution of the initial eigenvectors does not follow a Gaussian distribution. Therefore, this was used as an additional criterion in determining the number of eigenvectors of the essential subspace.

For each principal component (pi) of the covariance matrix, the cosine content (ci) may be calculated from the C matrix; this absolute value goes from 0 (no cosine) to 1 (perfect cosine). The equation of cosine content is as follows:

$$c_i = \frac{2}{T} \left(\int \cos(i\pi t\, p_i(t)\, dt\right)^2 \left(\int p_i^2\,(t)\, dt\right)^{-1}$$

Where T is the time of the simulation. Insufficient sampling has been associated with abnormally high ci values, which indicate random motion. The large-scale behavior of proteins is similar to

diffusion when the cosine content of the first few PCs is close to 1. All trajectories' cosine contents were calculated by analyzing the first 10 PCs.

Free Energy Landscape (FEL):

The exploration of a protein's free-energy landscape can be achieved by using a conformational sampling approach, which allows for the investigation of conformations that are similar to the native state structure. In this study, we employed the molecular dynamics (MD) simulation method to fulfill this purpose. To visually represent the free-energy landscape (FEL) in two dimensions, we define the probability of observing the system in a specific state, characterized by certain variables of interest (known as reaction coordinates), as proportionate to the exponential of the negative ratio between the free energy G_α associated with that state and the product of the Boltzmann constant k and the absolute temperature T ($e^{-G\alpha/kT}$). The complete FEL can be derived from these calculations.

$$G\alpha = -kT \ln\left(\frac{P(q_\alpha)}{P_{max}(q)}\right)$$

Where *k* is the Boltzmann constant, *T* is the temperature of simulation, $P(q_\alpha)$ is an estimate of the probability density function obtained from a histogram of the MD data and $P_{max}(q)$ is the probability of the most probable state. We derived two-dimensional representations of the free-energy landscapes by considering two separate reaction coordinates: qi and qj. These representations were obtained by analyzing the joint probability distributions, referred to as P(qi,qj), which describe the system [https://www.sciencedirect.com/science/article/pii/S1093326309000175]. The GROMACS gmx sham command was utilized for this purpose.

S.3.4. Density Function Theory (DFT) calculations

After optimization at the DFT/B3LYP/6-31+G (d, p) theory level using the Gaussian 09(D.01) program, the quantum chemistry calculations were performed using the Multiwfn, AIMALL, and Gauss View 5.0 interface tools. AIMAll and Multiwfn programs were employed for the examination of the quantum theory of atoms in molecules (QTAIM). Numerous analyses, including those of electrostatic potential (ESP), total density of states (TDOS), frontiers molecular

orbitals (FMO), and total density of states (TDOS), have been done at the same theoretical level. The following mathematical formulas were used to determine the global reactivity receptors:

$IP = -E_{HOMO}$

$EA = -E_{LUMO}$

$\mu = (IP+EA)/2$

$\eta = (IP-EA)$

$\chi = -\mu$

$\omega = \mu^2/(2\eta)$

$\sigma = 1/\eta$

$\Delta N = -(\mu/\eta)$

$\Delta E = -\omega$

$E_{gap} = E_{LUMO} - E_{HOMO}$

where (μ) chemical potential, (σ) is softness, (η) is global hardness, (EA) is electron affinity (EA), (χ) is electronegativity, (IP) is ionization potential, (ΔE) is energy change, (ω) is electrophilicity index, and E_{gap} is the band gap between HOMO and LUMO orbitals.

Table S1. The QTAIM parameters (a.u.) at bond critical points (BCPs) of compound **14.**

BCP #	Atoms	(ρ)	(∇2ρ)	K(r)	G(r)	V(r)	H(r)
1	C1 - C2	0.256466	-0.62583	0.215394	0.058938	-0.27433	-0.21539
2	C1 - O3	0.403276	-0.06569	0.69715	0.680728	-1.37788	-0.69715
3	C1 - N4	0.307285	-0.96937	0.459028	0.216686	-0.67571	-0.45903
4	C7 - C8	0.305403	-0.81588	0.29739	0.09342	-0.39081	-0.29739
5	C5 - C6	0.307829	-0.83398	0.304248	0.095753	-0.4	-0.30425
6	N4 - C5	0.287599	-0.87836	0.401897	0.182308	-0.58421	-0.4019
7	C7 - H30	0.289265	-1.07565	0.307155	0.038244	-0.3454	-0.30716
8	C6 - C7	0.315141	-0.86272	0.318005	0.102325	-0.42033	-0.318
9	C5 - C10	0.310864	-0.84684	0.309388	0.097678	-0.40707	-0.30939
10	C8 - C11	0.270685	-0.67357	0.233694	0.065302	-0.299	-0.23369
11	C9 - C10	0.312945	-0.85315	0.313479	0.100193	-0.41367	-0.31348
12	C8 - C9	0.306615	-0.81848	0.299902	0.095281	-0.39518	-0.2999
13	C10 - H32	0.283261	-1.01006	0.29477	0.042255	-0.33702	-0.29477

14	C11 - N13	0.370332	-0.80026	0.638515	0.43845	-1.07696	-0.63851
15	C11 - C12	0.254702	-0.60264	0.211313	0.060652	-0.27197	-0.21131
16	N13 - N14	0.343245	-0.63967	0.331974	0.172056	-0.50403	-0.33197
17	C12 - H33	0.28361	-1.0183	0.296322	0.041747	-0.33807	-0.29632
18	S15 - C16	0.195178	-0.35743	0.148929	0.059573	-0.2085	-0.14893
19	N14 - H33	0.016896	0.074351	-0.00335	0.015235	-0.01188	0.003353
20	N14 - C16	0.386386	-1.19527	0.66379	0.364971	-1.02876	-0.66379
21	S15 - C19	0.199388	-0.3705	0.154346	0.06172	-0.21607	-0.15435
22	N17 - C20	0.27094	-0.72164	0.383482	0.203071	-0.58655	-0.38348
23	N18 - C19	0.366414	-0.65692	0.628632	0.464402	-1.09303	-0.62863
24	C16 - N17	0.302056	-0.94084	0.431031	0.195822	-0.62685	-0.43103
25	N17 - N18	0.36515	-0.70103	0.366969	0.191711	-0.55868	-0.36697
26	C19 - C26	0.273059	-0.69495	0.239353	0.065616	-0.30497	-0.23935
27	C20 - C21	0.309461	-0.83984	0.307283	0.097322	-0.40461	-0.30728
28	N14 - H36	0.011404	0.042045	-0.00188	0.008636	-0.00676	0.001876
29	C21 - H36	0.289575	-1.07853	0.308196	0.038563	-0.34676	-0.3082
30	C22 - C24	0.312254	-0.86796	0.312791	0.0958	-0.40859	-0.31279
31	C20 - C23	0.313525	-0.85547	0.315032	0.101165	-0.4162	-0.31503
32	C23 - C25	0.308978	-0.83039	0.306392	0.098794	-0.40519	-0.30639
33	C21 - C24	0.314404	-0.85689	0.317056	0.102834	-0.41989	-0.31706
34	C22 - C25	0.310971	-0.84673	0.311019	0.099337	-0.41036	-0.31102
35	C26 - C43	0.257541	-0.62592	0.216893	0.060413	-0.27731	-0.21689
36	C43 - H44	0.280541	-0.99062	0.289722	0.042067	-0.33179	-0.28972
37	C26 - O27	0.401084	0.12538	0.683283	0.714628	-1.39791	-0.68328
38	N4 - H28	0.341929	-1.82397	0.507798	0.051807	-0.5596	-0.5078
39	C6 - H29	0.286376	-1.03455	0.300719	0.042082	-0.3428	-0.30072
40	C9 - H31	0.286588	-1.03623	0.300838	0.041781	-0.34262	-0.30084
41	C12 - H34	0.273922	-0.92699	0.277701	0.045955	-0.32366	-0.2777
42	C12 - H35	0.273361	-0.92275	0.276863	0.046175	-0.32304	-0.27686
43	C24 - H37	0.284988	-1.03287	0.298673	0.040454	-0.33913	-0.29867
44	C25 - H38	0.285253	-1.02363	0.299185	0.043279	-0.34246	-0.29919
45	C23 - H39	0.287031	-1.05078	0.302545	0.03985	-0.3424	-0.30255
46	O40 - C41	0.249211	-0.39723	0.354413	0.255107	-0.60952	-0.35441
47	C22 - O40	0.286527	-0.35775	0.438485	0.349048	-0.78753	-0.43849

48	C41 - H50	0.288072	-1.06136	0.302664	0.037323	-0.33999	-0.30266
49	C41 - H42	0.282696	-1.00809	0.291013	0.038991	-0.33	-0.29101
50	C43 - H45	0.275798	-0.95177	0.281808	0.043865	-0.32567	-0.28181
51	C43 - H46	0.274106	-0.93684	0.278831	0.04462	-0.32345	-0.27883
52	C2 - H47	0.280352	-0.98973	0.28971	0.042278	-0.33199	-0.28971
53	C6 - H49	0.008741	0.035179	-0.0021	0.006692	-0.00459	0.002103
54	C2 - H48	0.273833	-0.93054	0.278134	0.045498	-0.32363	-0.27813
55	C2 - H49	0.278397	-0.96613	0.285613	0.044081	-0.32969	-0.28561
56	C41 - H51	0.282711	-1.00819	0.291041	0.038995	-0.33004	-0.29104

S.3.5. ADMET studies

ADMET descriptors (absorption, distribution, metabolism, excretion and toxicity) of the compounds were determined using Discovery studio 4.0. Sorafenib was used as a reference molecule. At first, the CHARMM force field was applied then the tested compounds were prepared and minimized according to the preparation of small molecule protocol.

- **Preparation of the tested compounds:**

In this protocol, the general-purpose panel was utilized with the activation of the Prepare ligand option. The change ionization was switched on the true option using the Rule based as an ionization method. In Rule based task, we used the carboxylate as an acid ionization. Additionally, the primary, secondary, and tertiary amines were selected as Base ionization. The ionization enumeration option was switched on the one protomer. Under the filter smart option, we selected all options. The false option was selected for tasks Generate tautomers, generate isomers, Fix bad valencies, and parallel processing. The generate coordinates task was switched on the 3D option. Finally, the duplicate structure task was activated on the remove option.

- **Running of ADMET protocol**

In this protocol, the small molecules panel was utilized with the activation of the ADMET descriptors option. Then, we selected the prepared compounds as the input ligands. Further, all the ADMET parameters (aqueous solubility, Blood brain barrier, intestinal absorption, CYP2D6, and

plasma protein binding) were selected. Then, the output of the running protocol was visualized to give the ADMET chart.

S.3.6 Toxicity studies

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

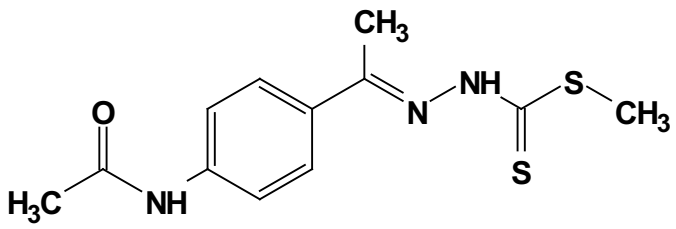
- **Preparation of the tested compounds:**

In this protocol, the general-purpose panel was utilized with the activation of the Prepare ligand option. The change ionization was switched on the true option using the Rule based as an ionization method. In Rule based task, we used the carboxylate as an acid ionization. Additionally, the primary, secondary, and tertiary amines were selected as Base ionization. The ionization enumeration option was switched on the one protomer. Under the filter smart option, we selected all options. The false option was selected for tasks Generate tautomers, generate isomers, Fix bad valencies, and parallel processing. The generate coordinates task was switched on the 3D option. Finally, the duplicate structure task was activated on the remove option.

- **Running of Toxicity protocol**

In this protocol, the small molecules panel was utilized with the activation of the toxicity prediction (extensible) option. Then, we selected the prepared compounds as the input ligands. Further, the different toxicity models were selected from the model panel. The similarity search task was activated to be true. The detailed report task was switched on as a PDF file. Then, the output of the running protocol was visualized to give the toxicity PDF report.

1H NMR of ANS



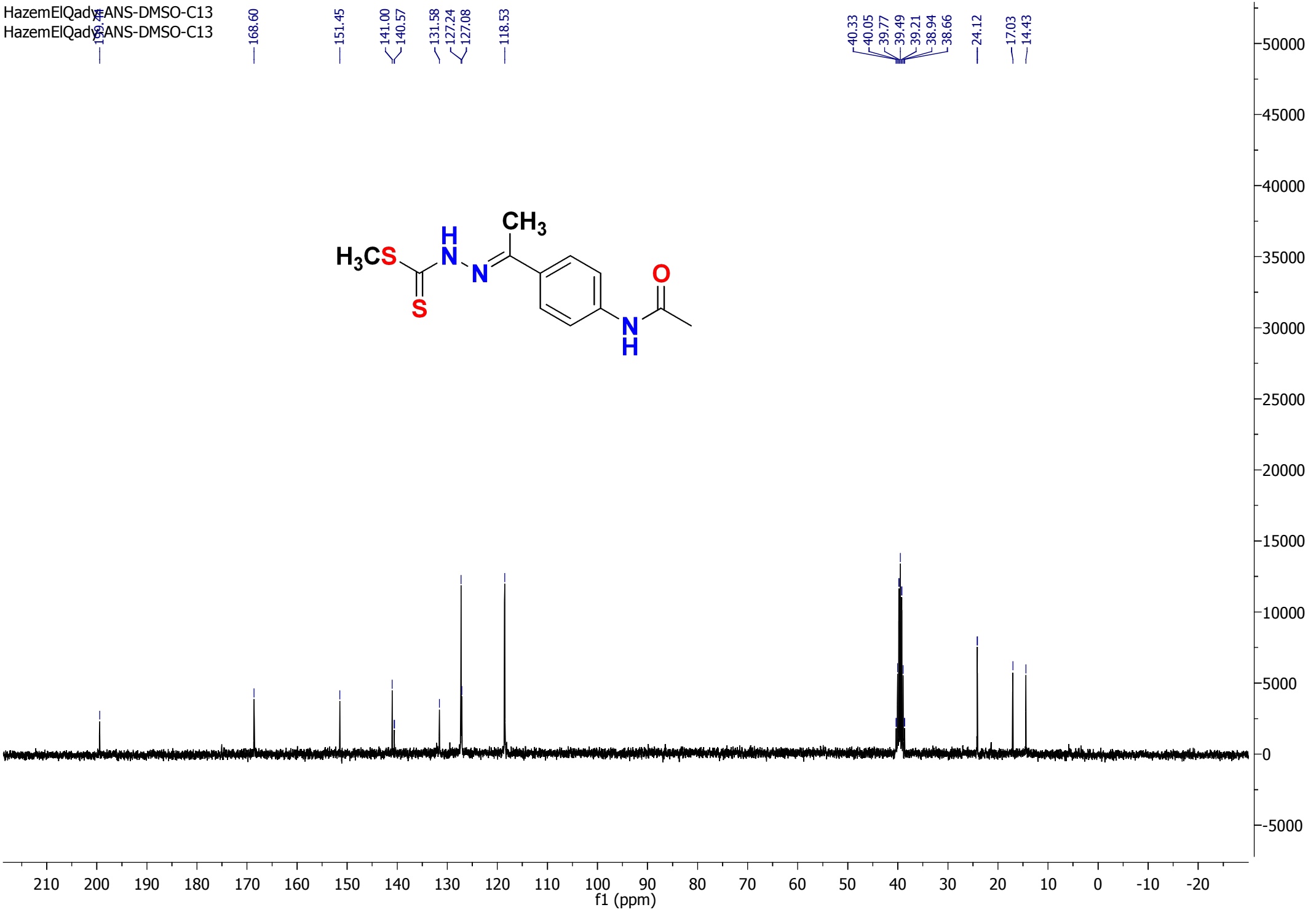
12.35 10.10 7.79 7.78 7.77 7.77 7.76 7.63 7.62 7.62 7.61 7.60 2.47 2.32 2.03

1.00 1.14 2.39 2.23 3.33 3.38 3.30

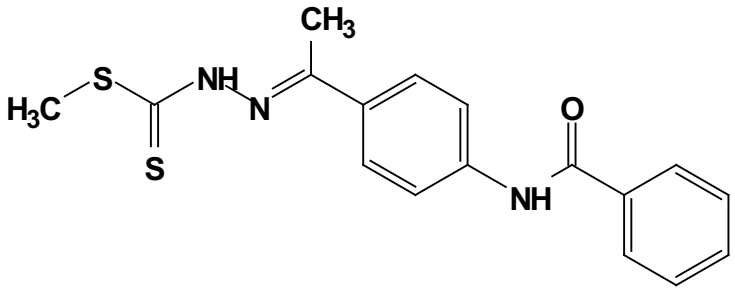
f1

HazemElQadyANS-DMSO-C13

HazemElQadyANS-DMSO-C13



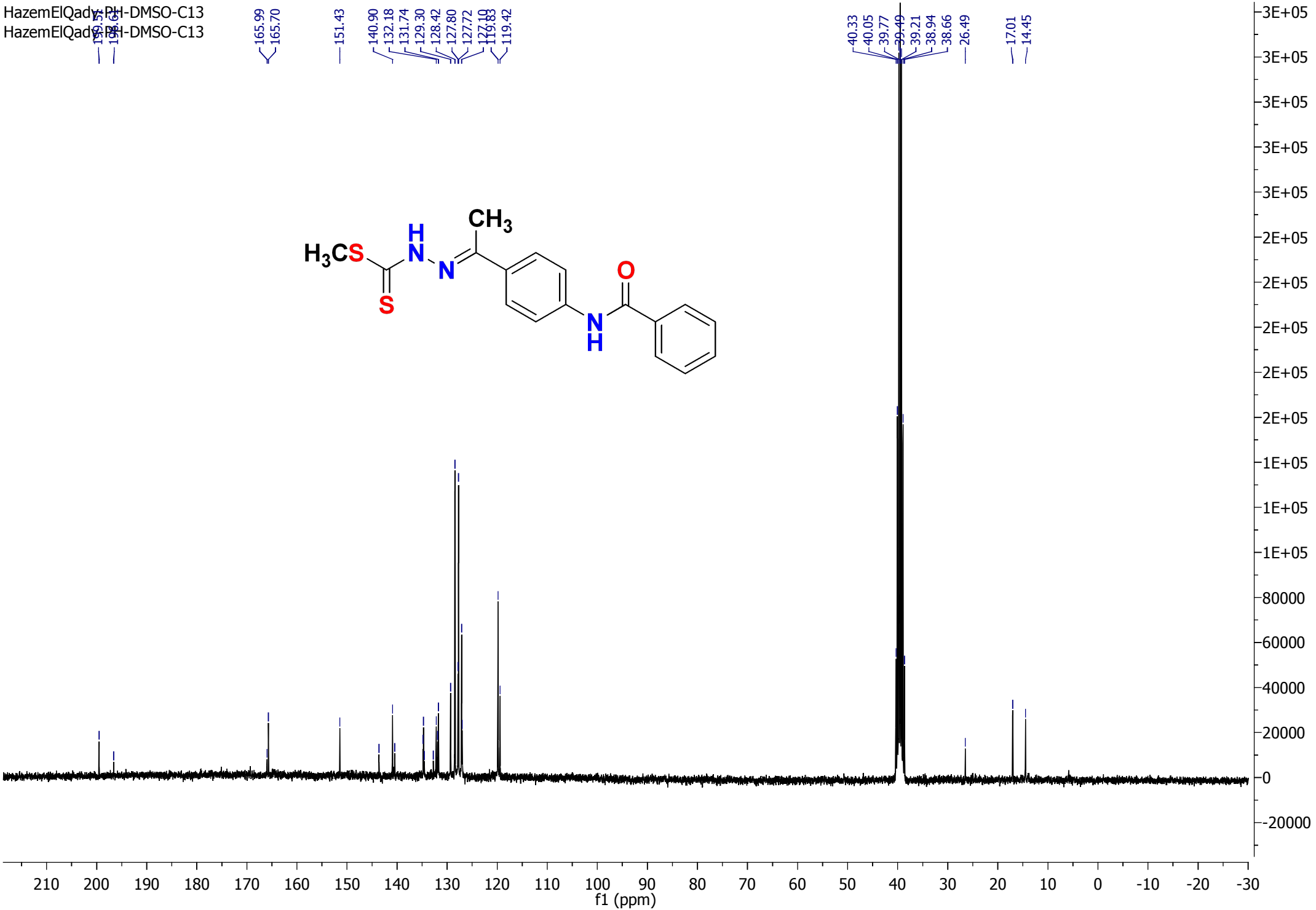
1H NMR of PH



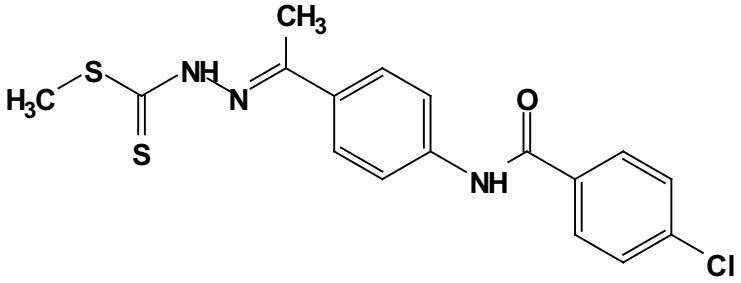
12.39
10.41
7.95
7.94
7.94
7.93
7.93
7.93
7.84
7.58
7.53
7.53
7.52
2.49
2.35

1.00
1.02
3.21
3.13
1.29
2.08
3.13
3.16

f1

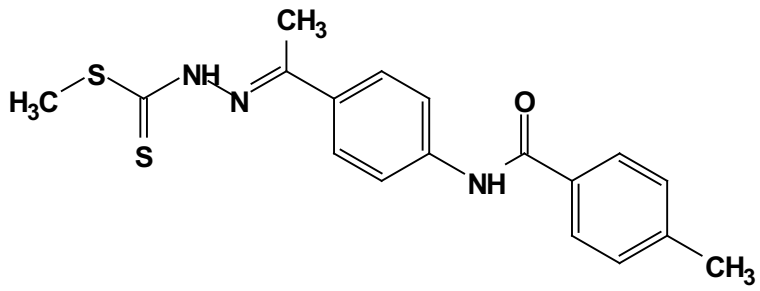


1H NMR of CNS


CH3
S
NH
H3C
N
S
O
NH
Cl

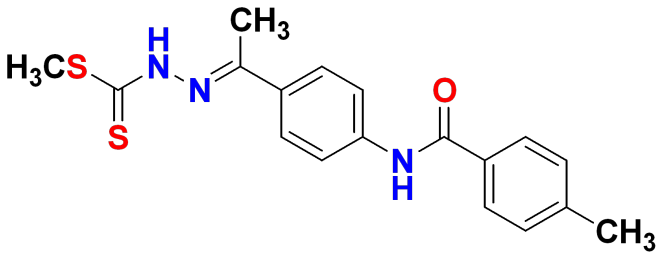
12.39
10.46
7.99
7.98
7.97
7.96
7.94
7.91
7.89
7.86
7.84
7.83
7.82
7.61
7.60
7.59
7.58
2.52
2.35
1.00
1.02
3.01
3.32
2.04
3.16
3.09
f1

1H NMR of MNS

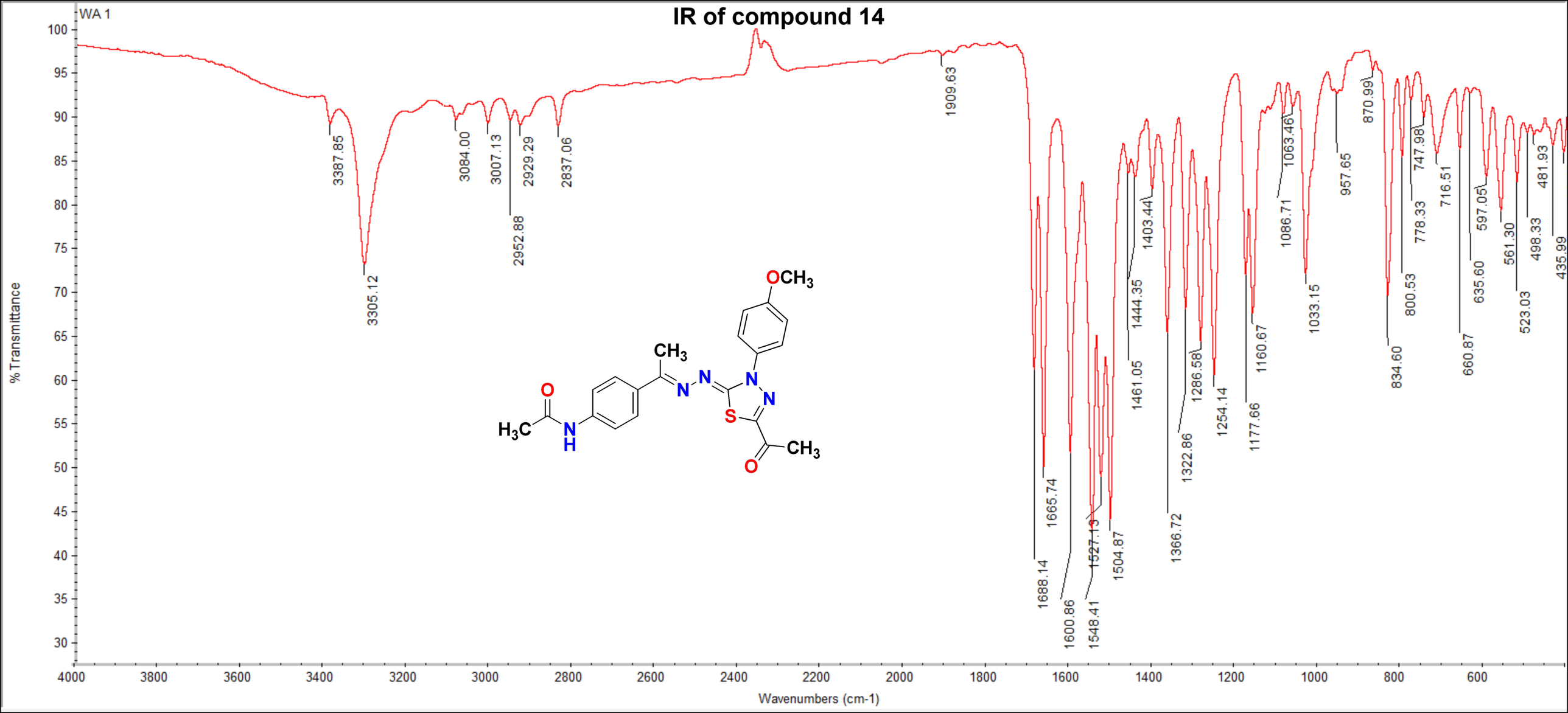

CH3
S
NH
H3C
N
S
O
NH
CH3

12.39
10.31
7.93
7.92
7.87
7.85
7.84
7.34
7.33
7.31
2.52
2.36
2.35
0.98
1.10
1.15
5.06
2.39
3.04
3.06
3.11
f1

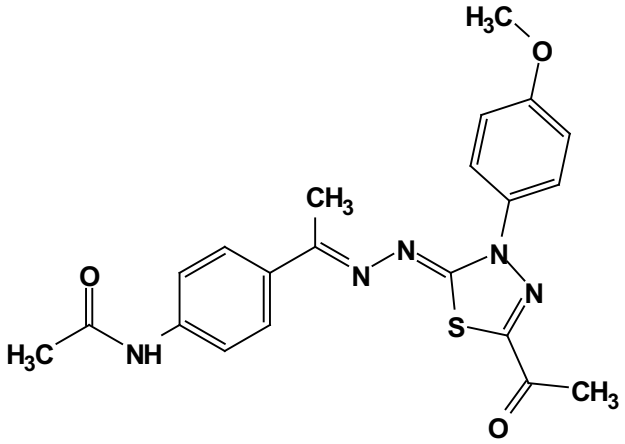
HazemElQadyMNS-DMSO-C13
HazemElQadyMNS-DMSO-C13



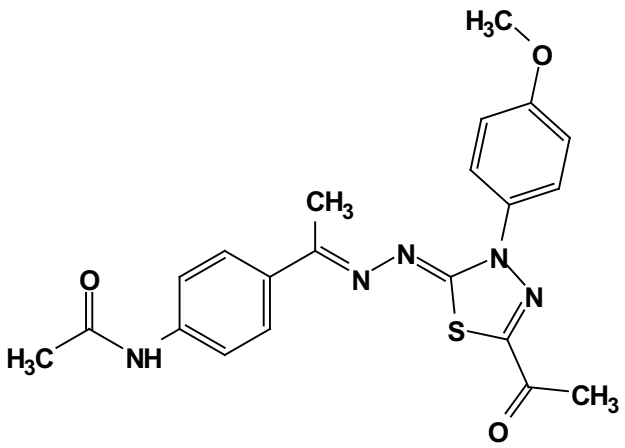
IR of compound 14



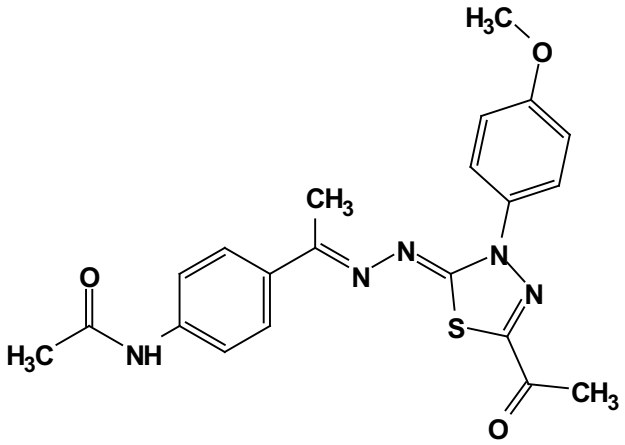
1H NMR of compound 14



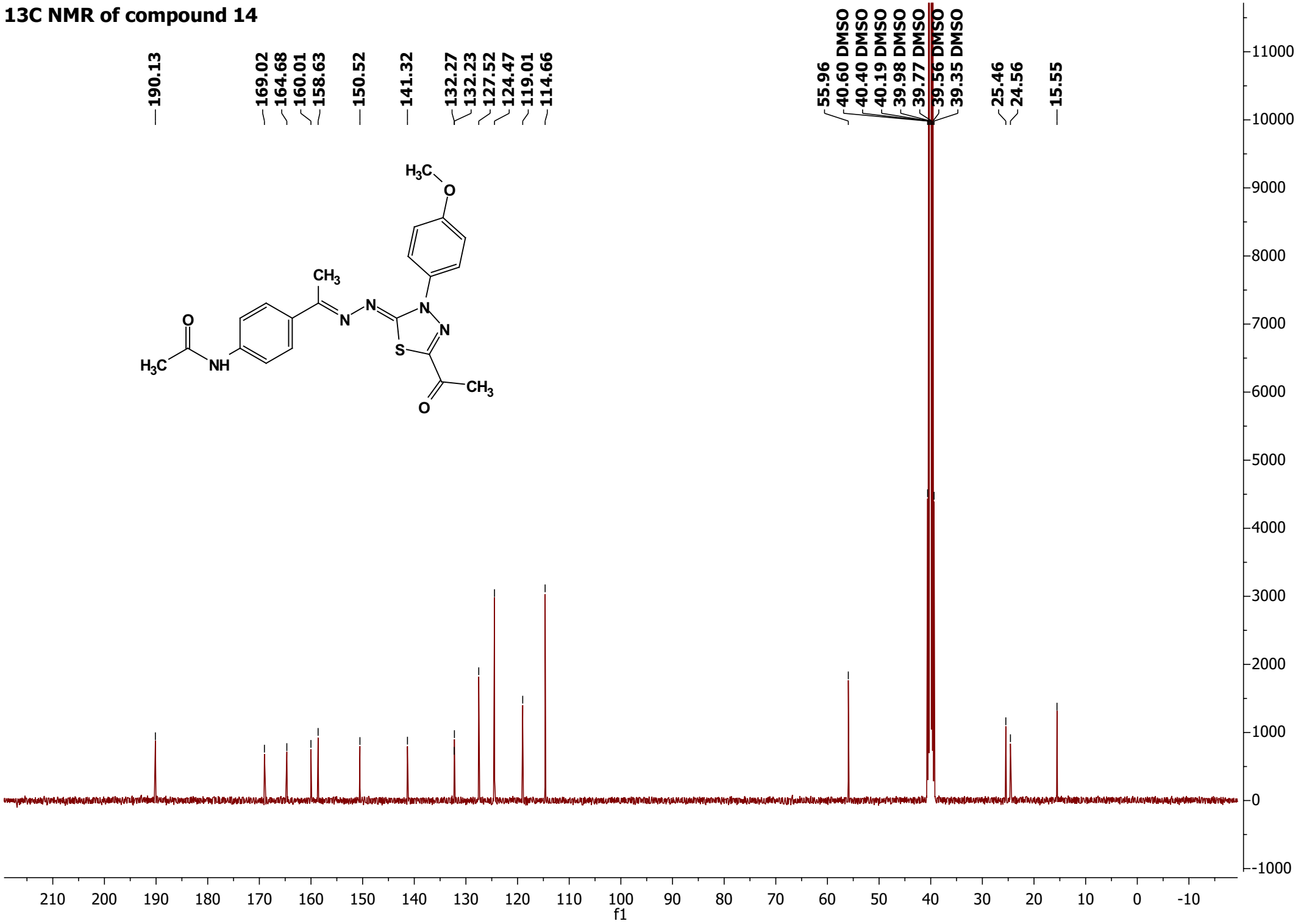
1H NMR of compound 14



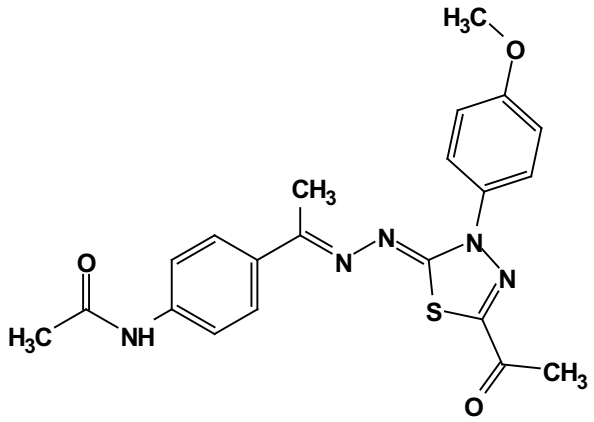
1H NMR of compound 14



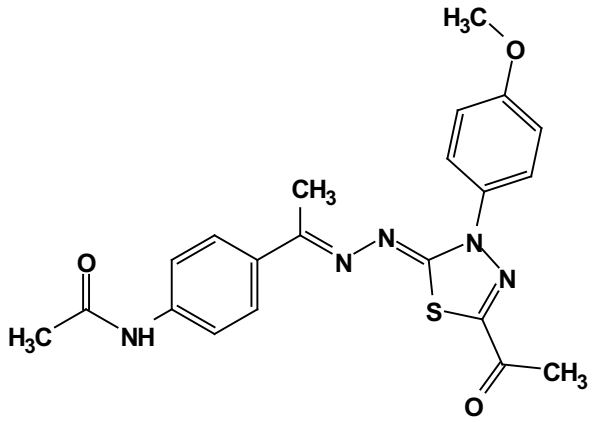
13C NMR of compound 14



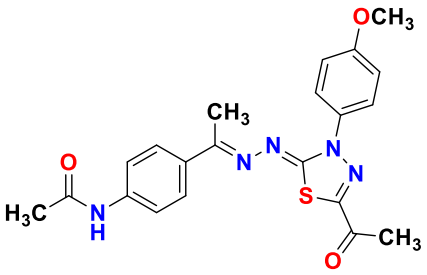
13C NMR of compound 14



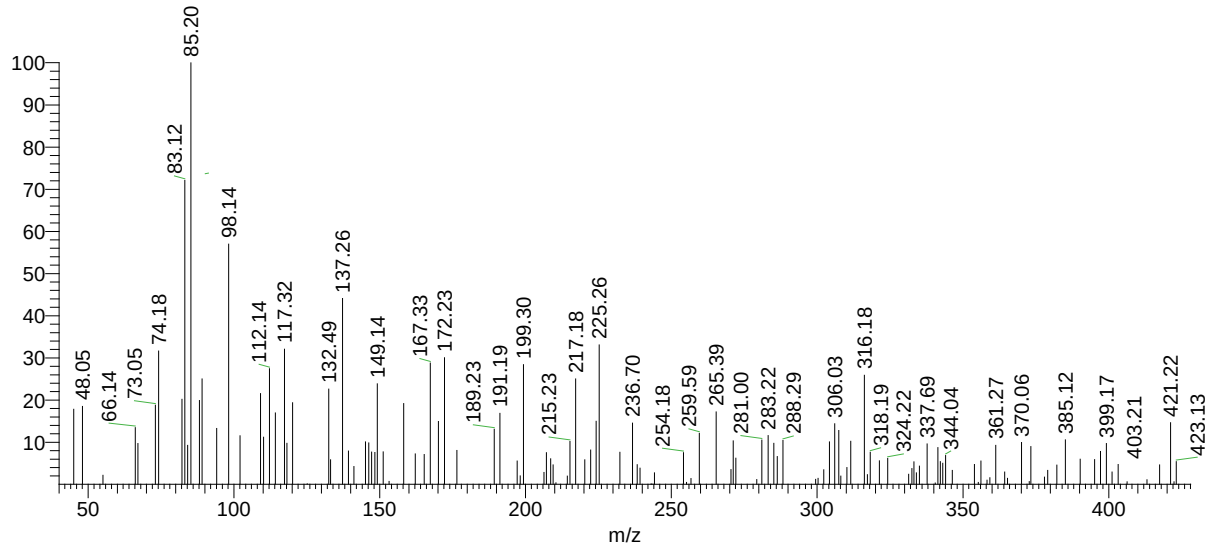
13C NMR of compound 14



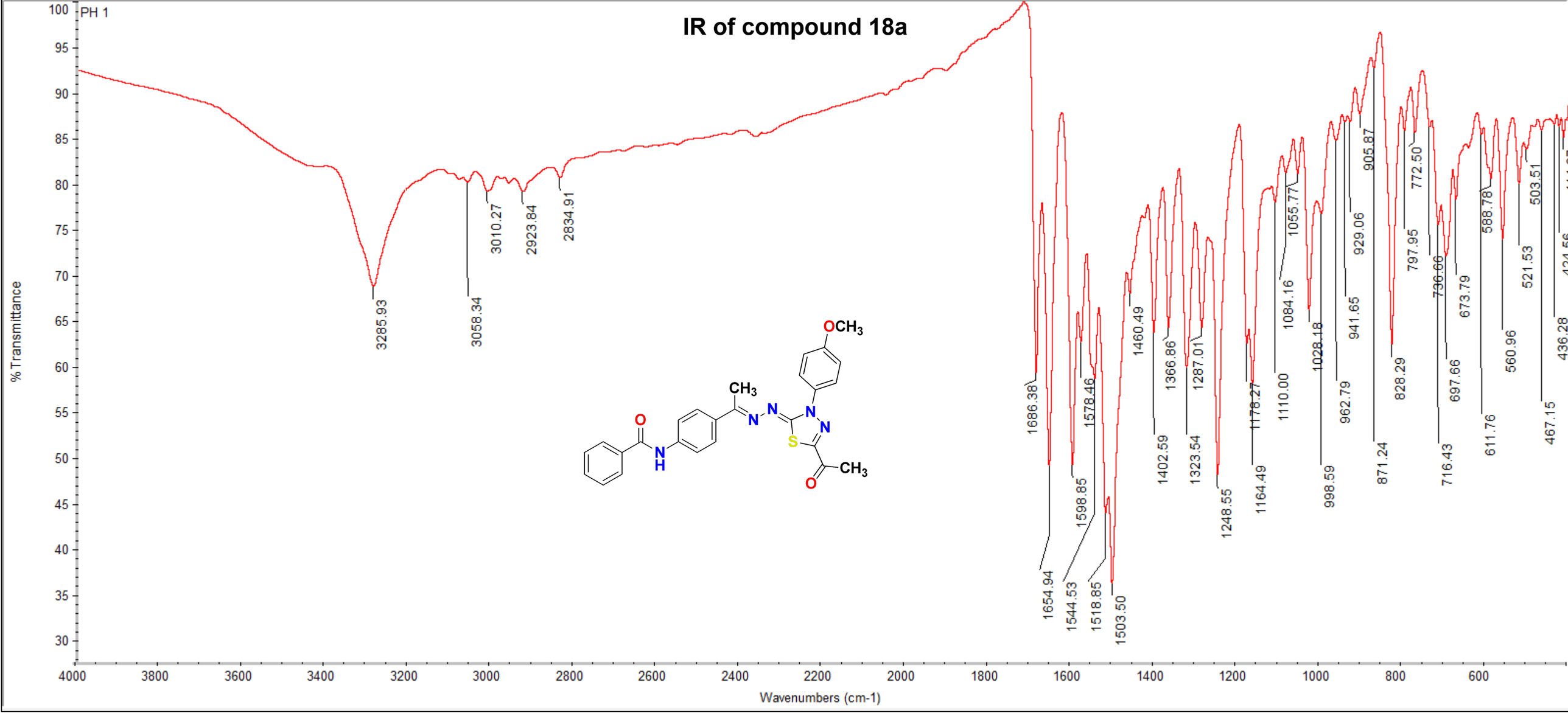
Mass spec. of compound 14



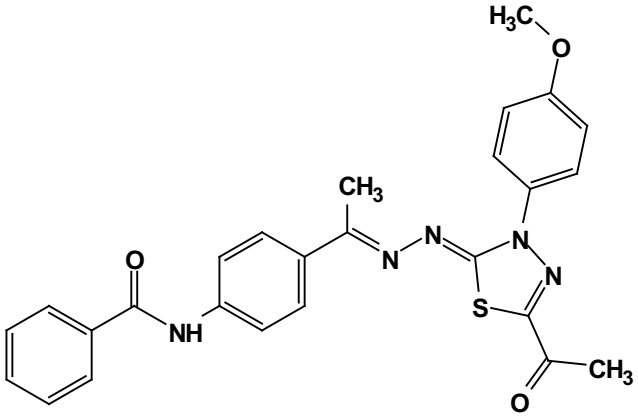
ibrahim-hassan-wa1 #195 RT: 3.28 AV: 1 SB: 2 2.59 , 2.56 NL: 7.55E3
T: {0,0} + c EI Full ms [40.00-1000.00]



IR of compound 18a



1H NMR of compound 18a

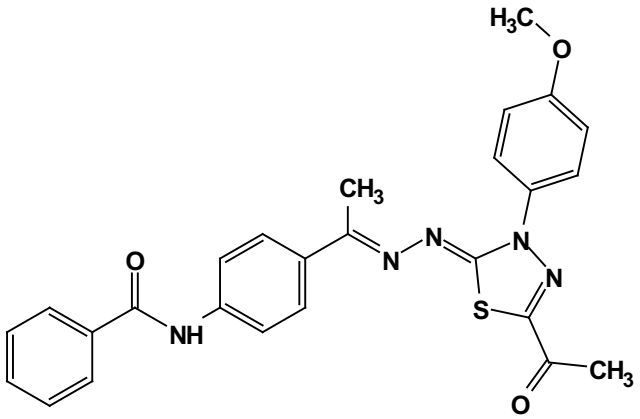


10.46 8.01 8.01 8.00 8.00 7.99 7.99 7.98 7.97 7.95 7.94 7.93 7.93 7.92 7.90 7.65 7.65 7.64 7.63 7.63 7.61 7.61 7.61 7.58 7.57 7.56 7.55 7.54 7.54 7.17 7.16 7.15 7.14 7.14 3.85 3.37 HDO 2.59 2.41

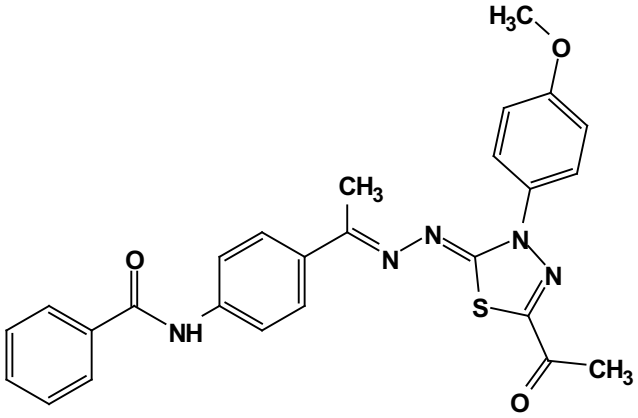
1.06 3.31 5.25 1.32 1.97 2.06 3.00 3.17 3.08

f1

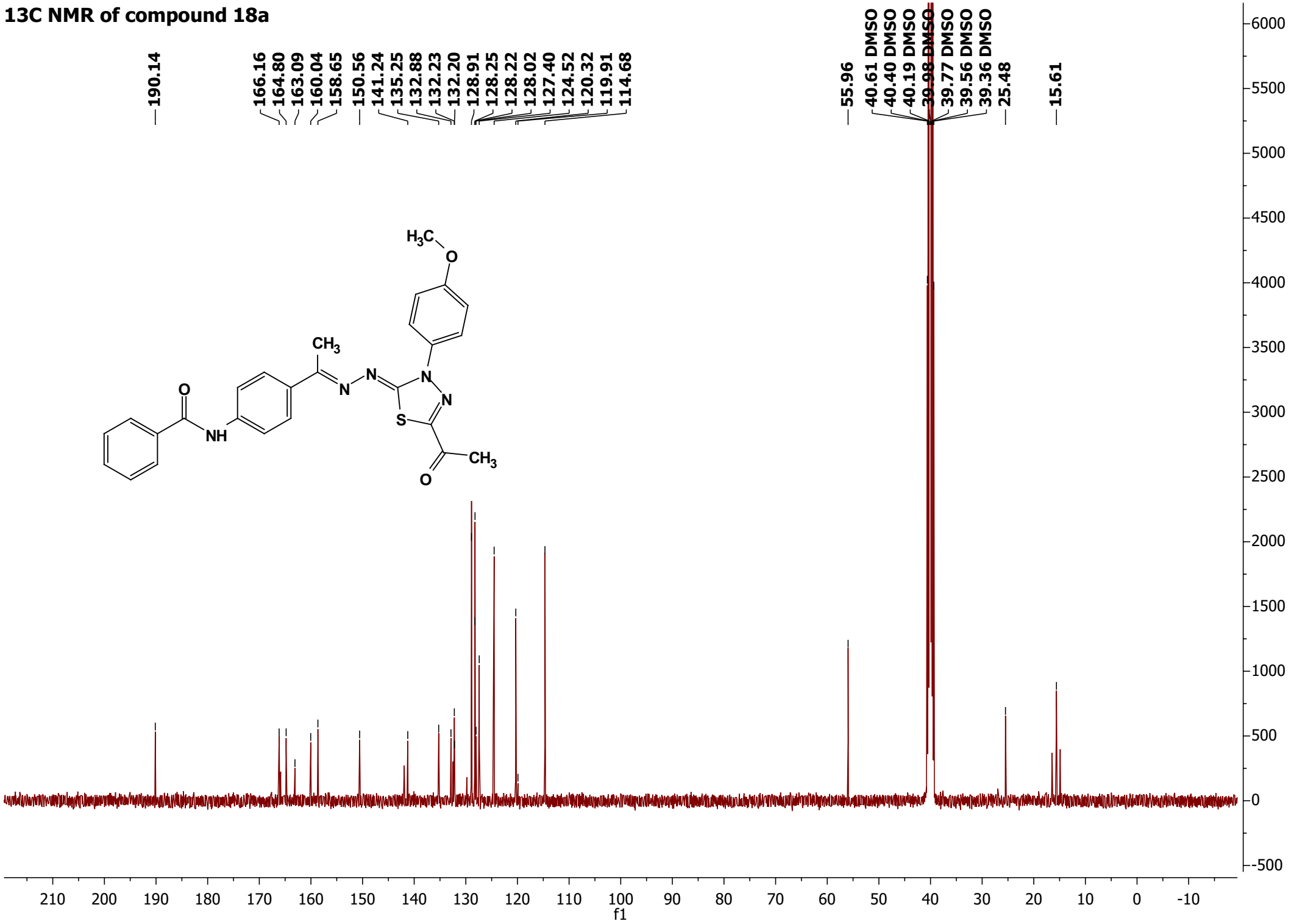
1H NMR of compound 18a



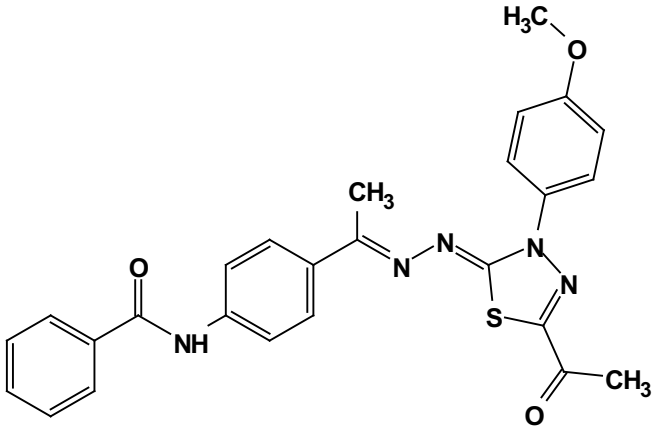
1H NMR of compound 18a



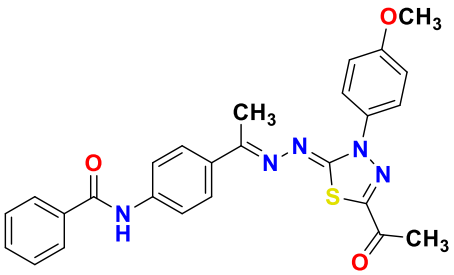
13C NMR of compound 18a



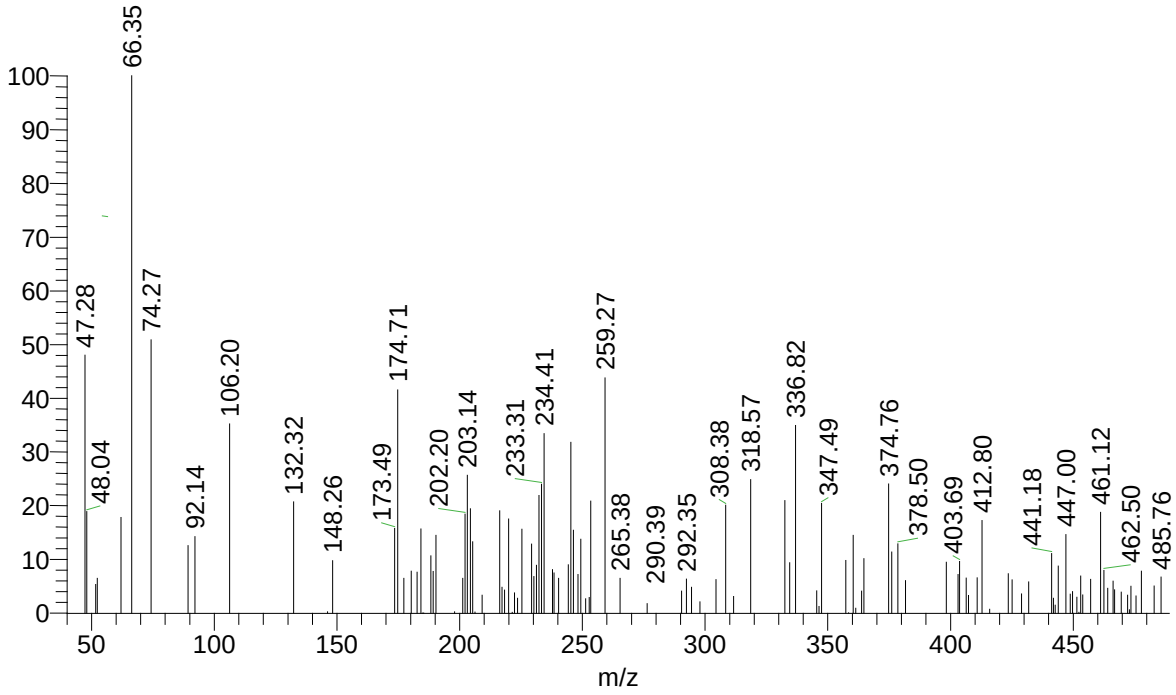
13C NMR of compound 18a



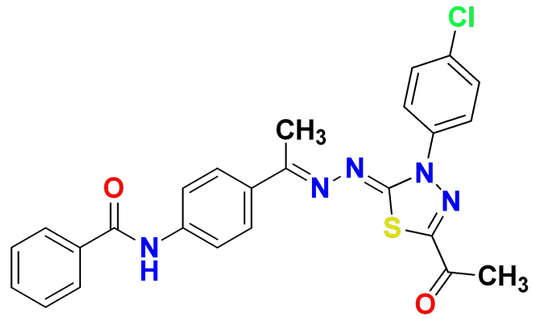
Mass spec. of compound 18a



ibrahim-hassan-PH-1 #4-20 RT: 0.08-0.35 AV: 17 SB: 26 1.21-1.34 , 0.87-1.14 NL: 4.76E2
T: {0,0} + c EI Full ms [40.00-1000.00]



IR of compound 18b



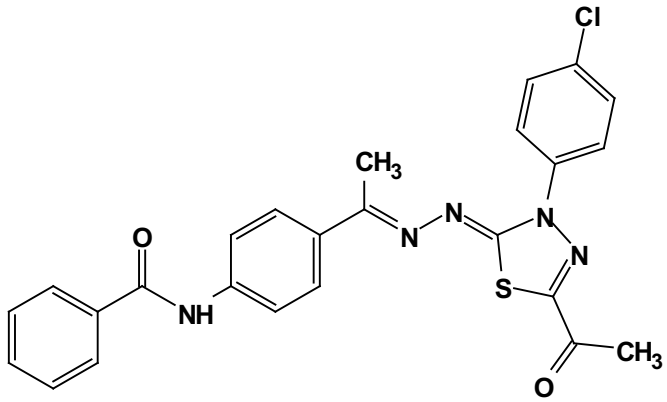
PH 4

% Transmittance

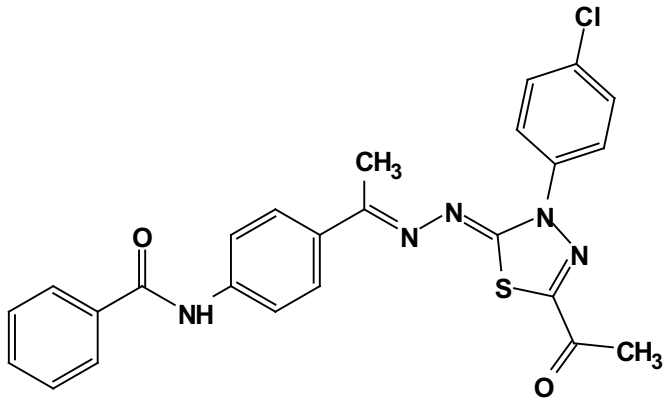
Wavenumbers (cm-1)

3331.32
3103.17
3067.96
3014.35
1682.22
1654.89
1597.98
1553.34
1521.09
1486.12
1446.93
1405.01
1362.90
1323.67
1282.98
1180.57
1156.69
1108.17
1091.12
1055.34
1028.36
1011.64
946.56
925.87
900.97
865.25
829.96
793.18
742.12
715.10
690.68
672.88
613.64
580.14
562.34
519.80
504.03
477.42
434.24

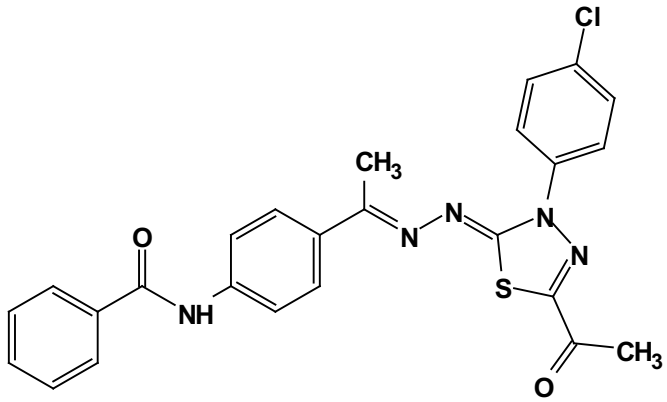
1H NMR of compound 18b



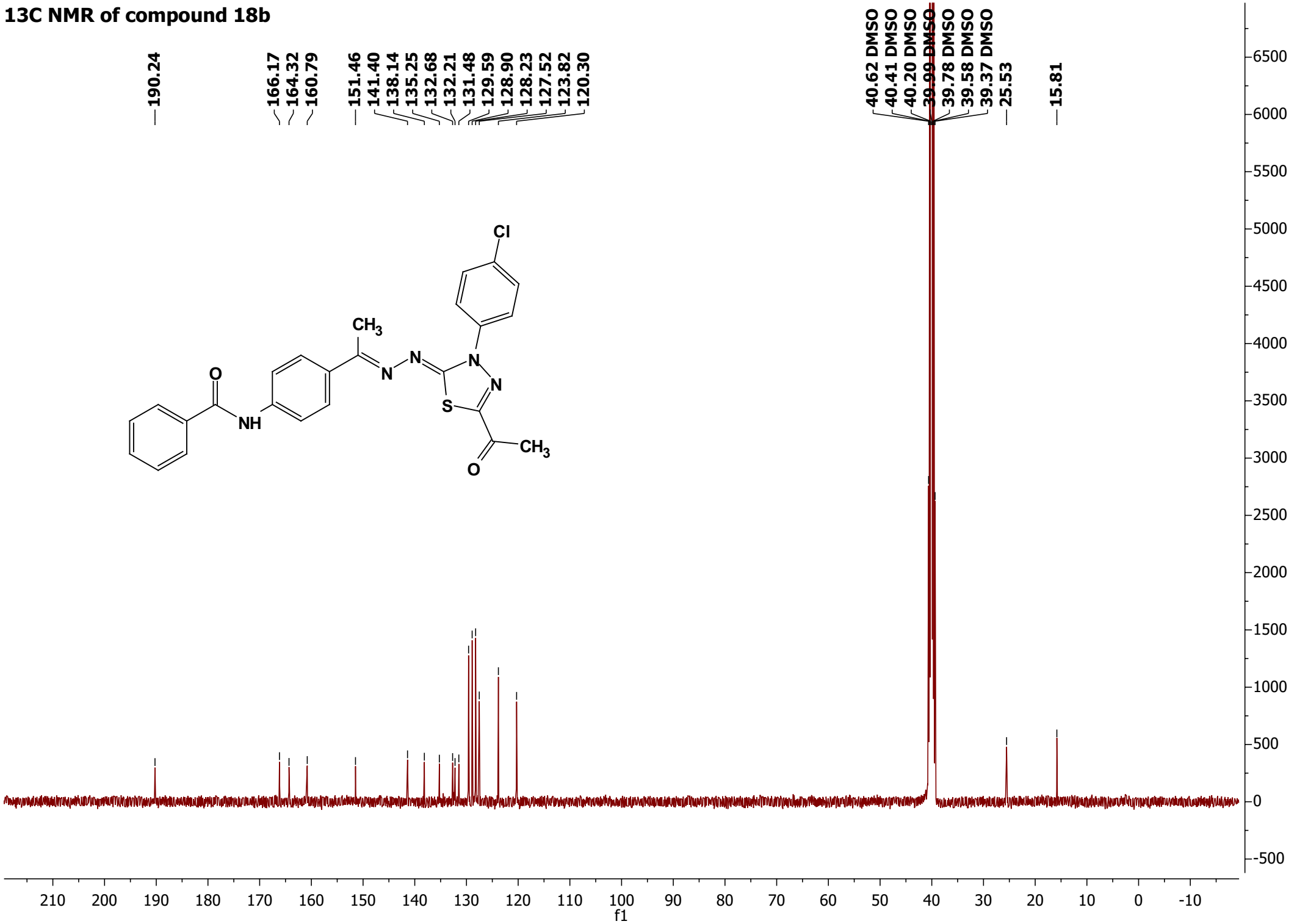
1H NMR of compound 18b



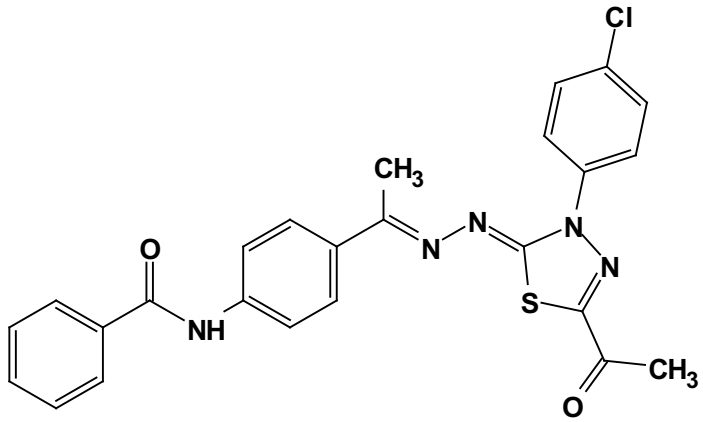
1H NMR of compound 18b



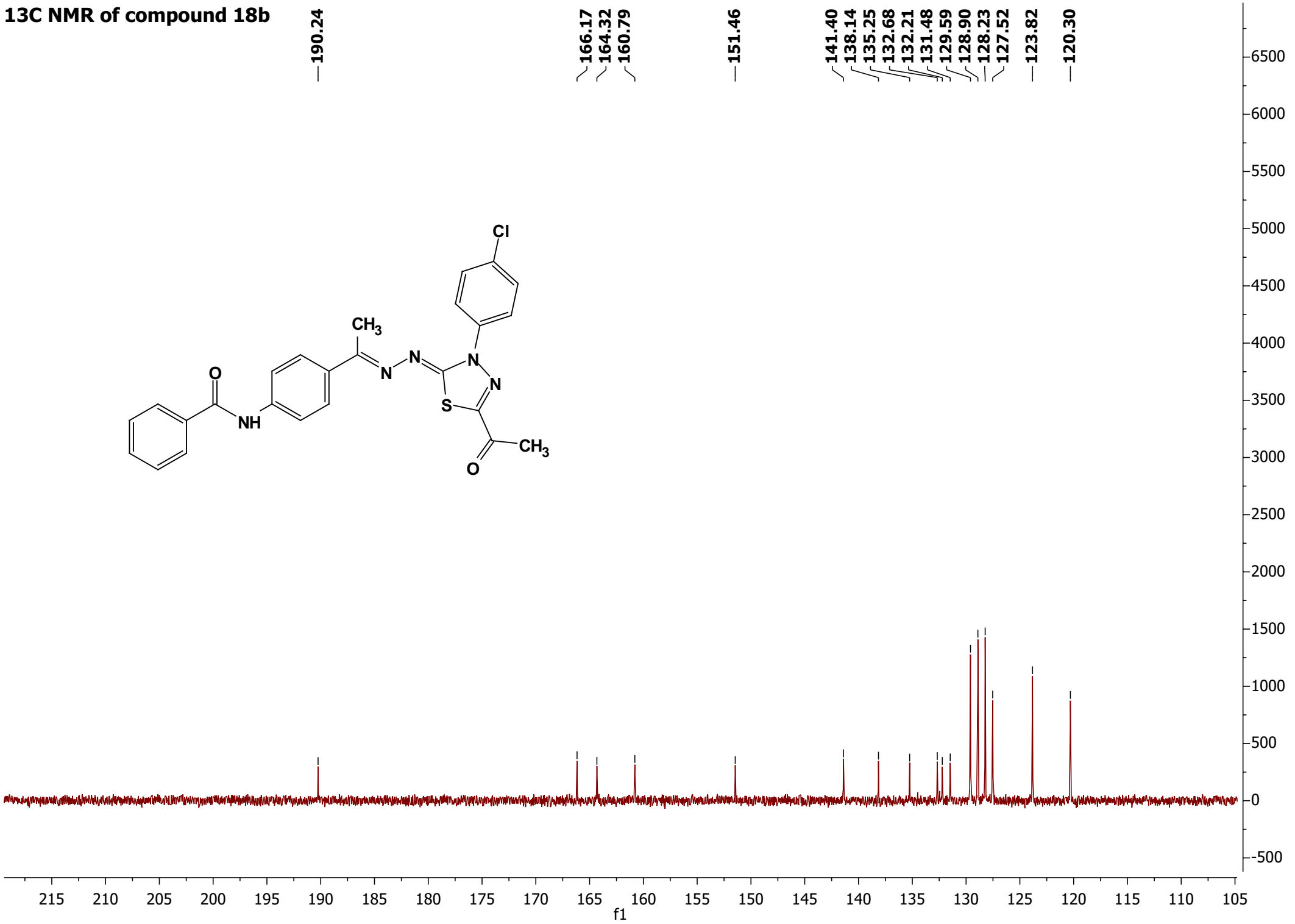
13C NMR of compound 18b



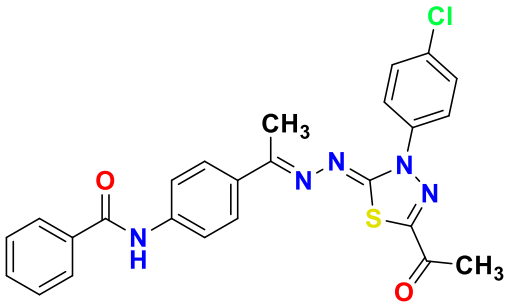
13C NMR of compound 18b



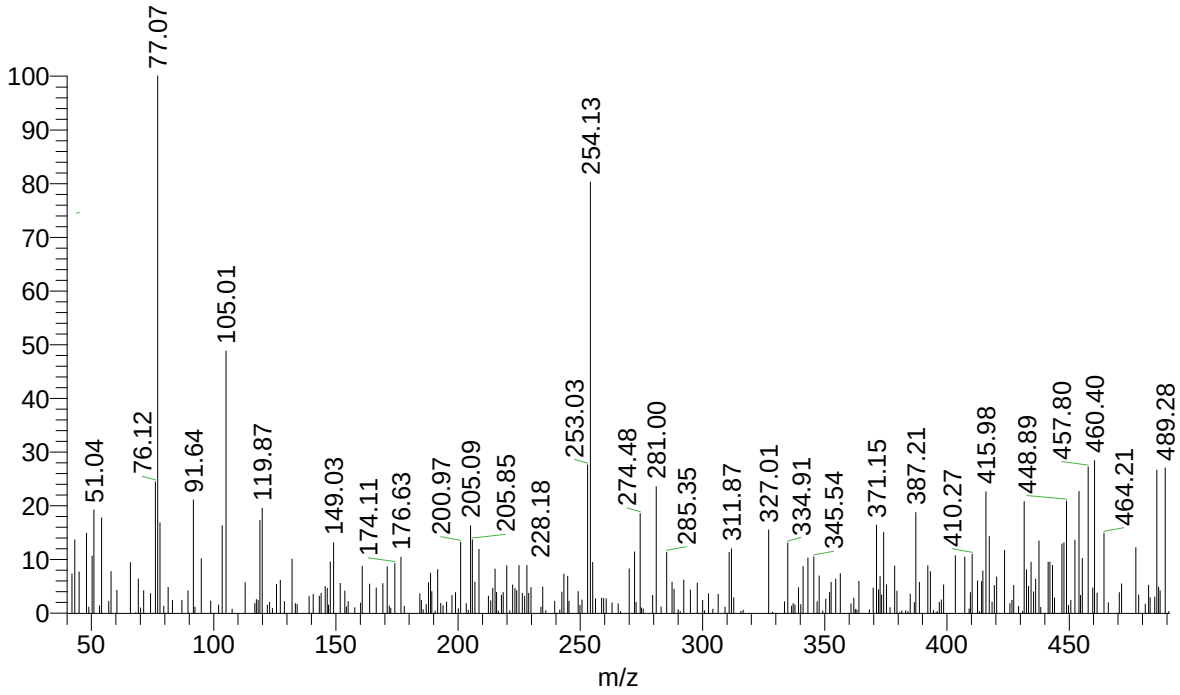
13C NMR of compound 18b



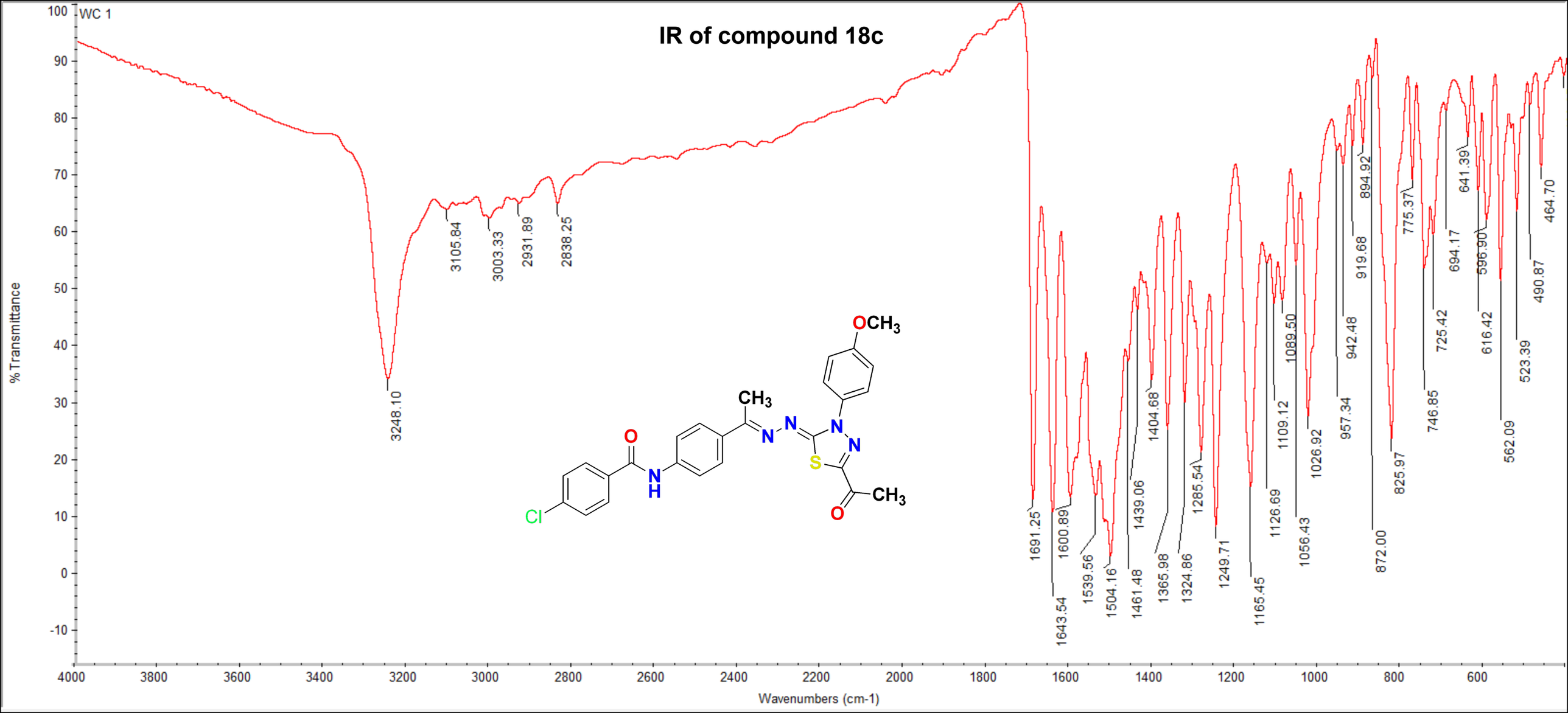
Mass spec. of compound 18b



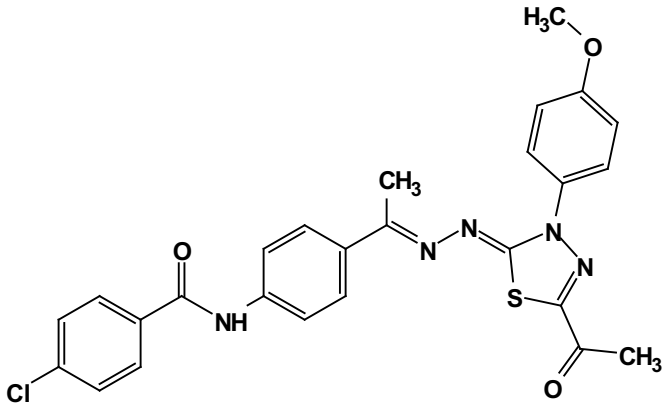
ibrahim-hassan-ph4 #87 RT: 1.47 AV: 1 SB: 26 1.21-1.34 , 0.87-1.14 NL: 3.38E3
T: {0,0} + c EI Full ms [40.00-1000.00]



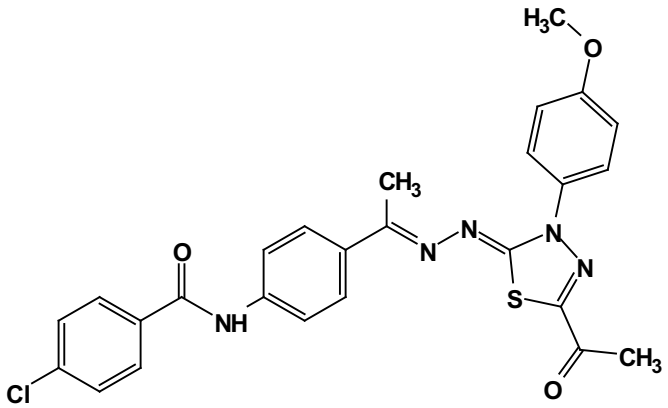
IR of compound 18c



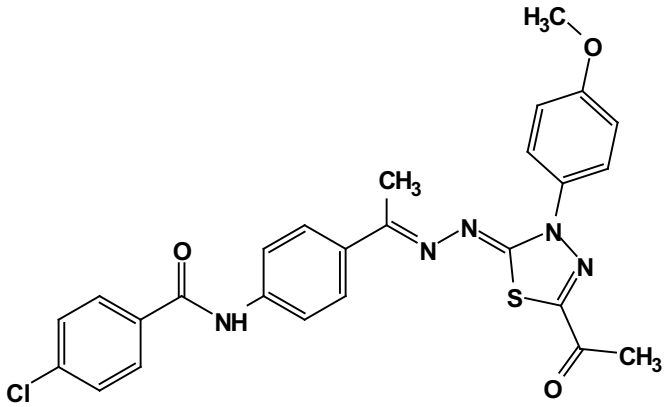
1H NMR of compound 18c



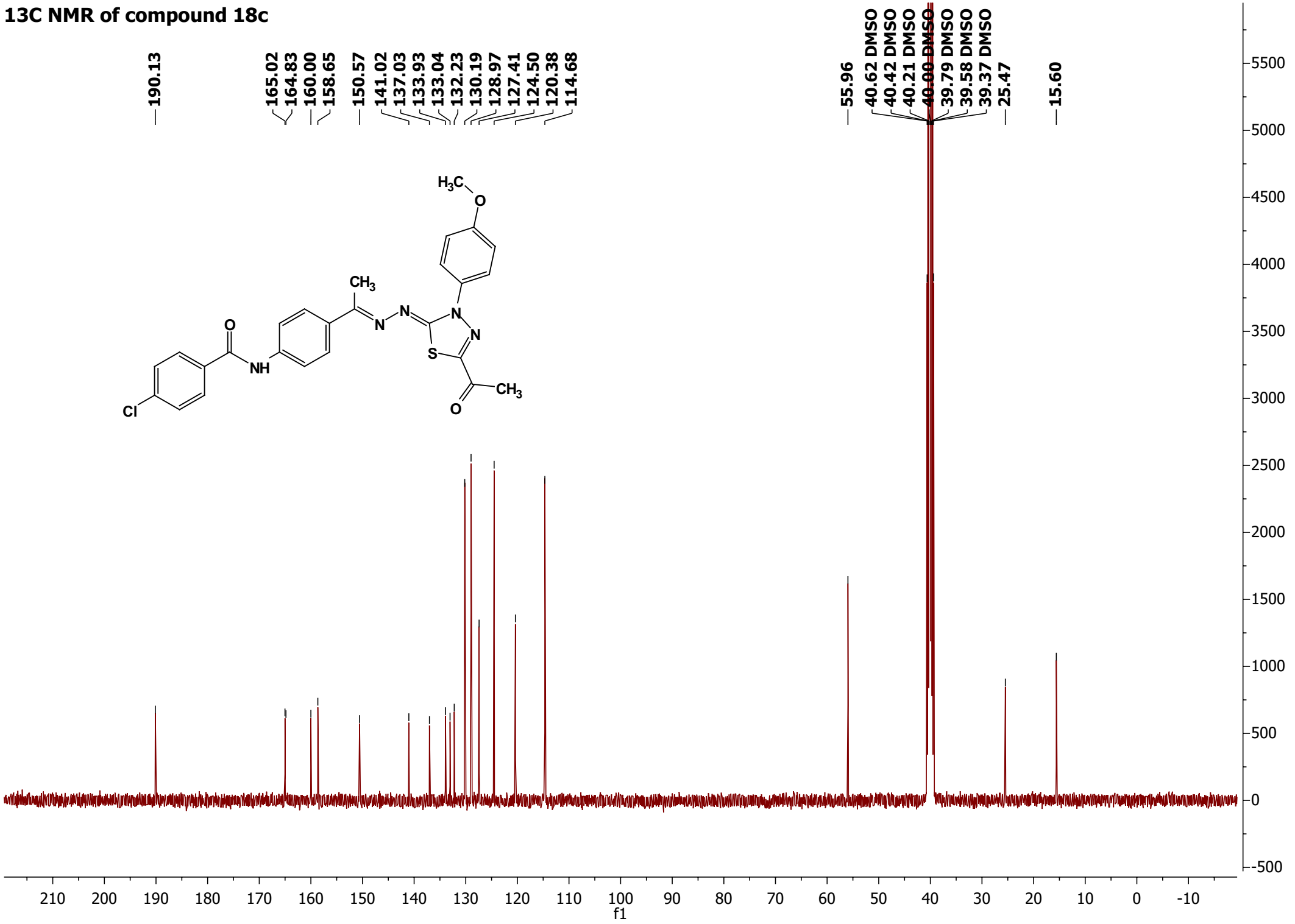
1H NMR of compound 18c



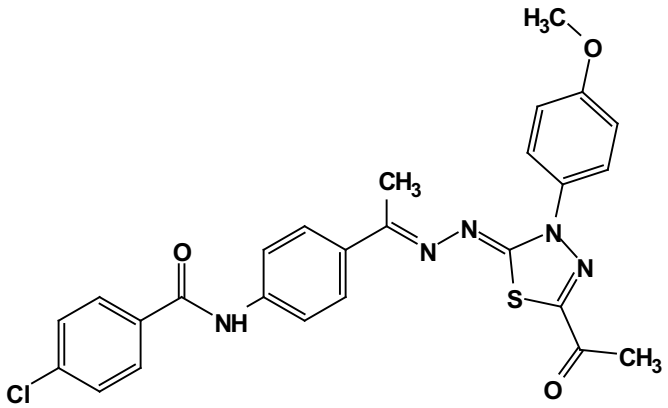
1H NMR of compound 18c



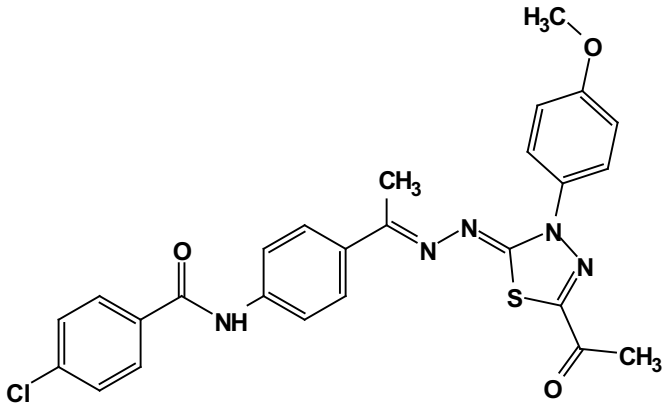
13C NMR of compound 18c



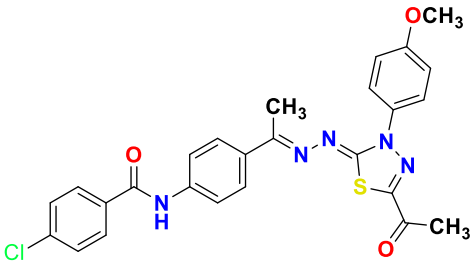
13C NMR of compound 18c



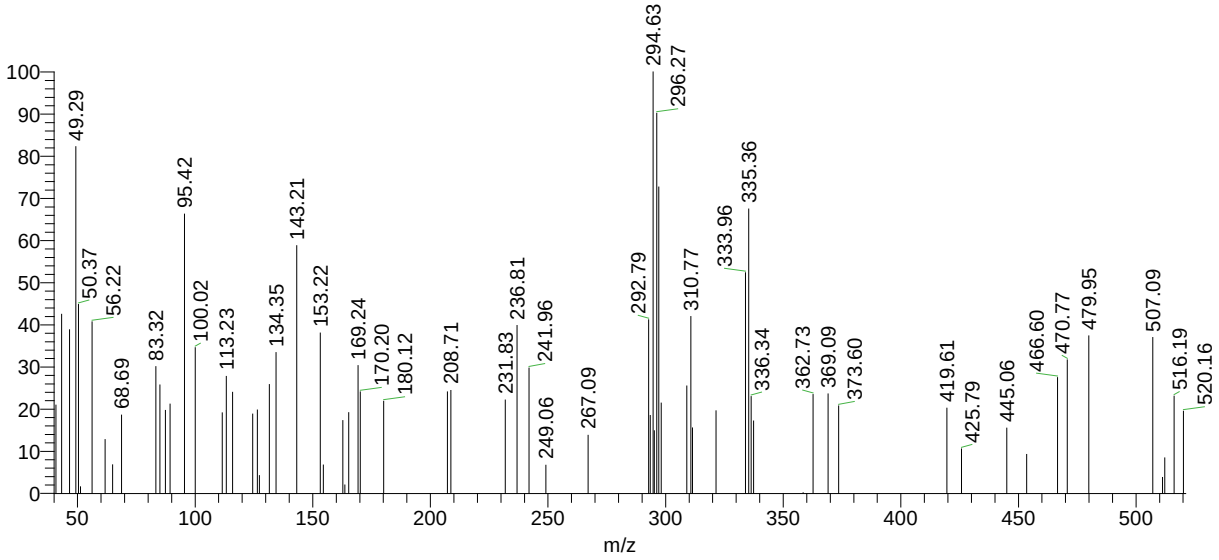
13C NMR of compound 18c



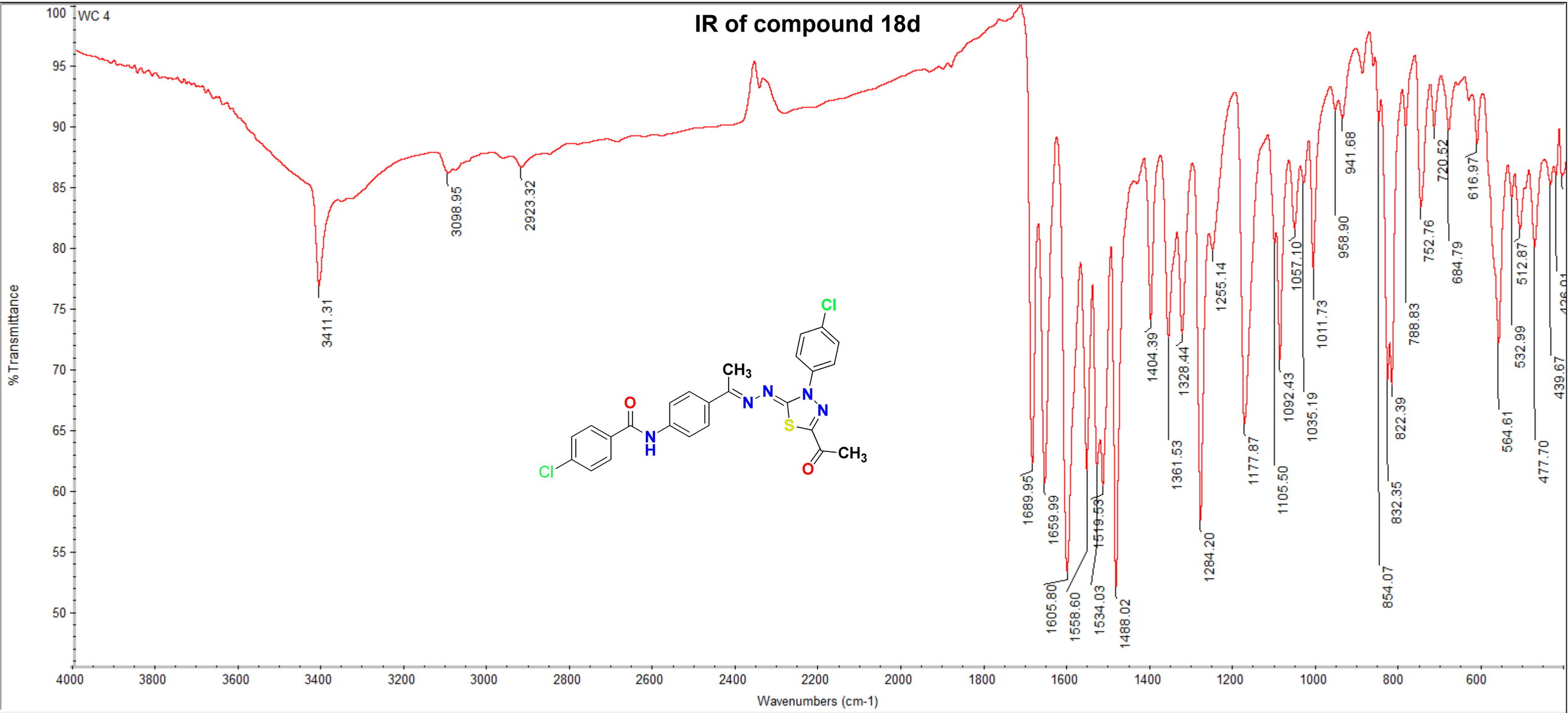
Mass spec. of compound 18c



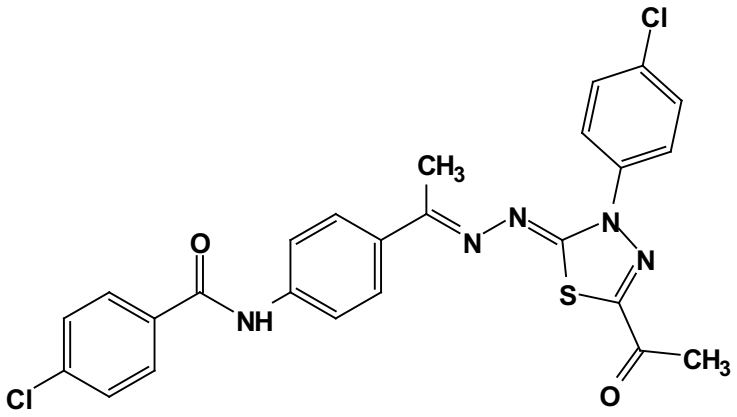
ibrahim-hassan-wc1 #180-182 RT: 3.03-3.06 AV: 3 SB: 15 3.53-3.63 , 3.46-3.58 NL: 1.78E2
T: {0,0} + c EI Full ms [40.00-1000.00]



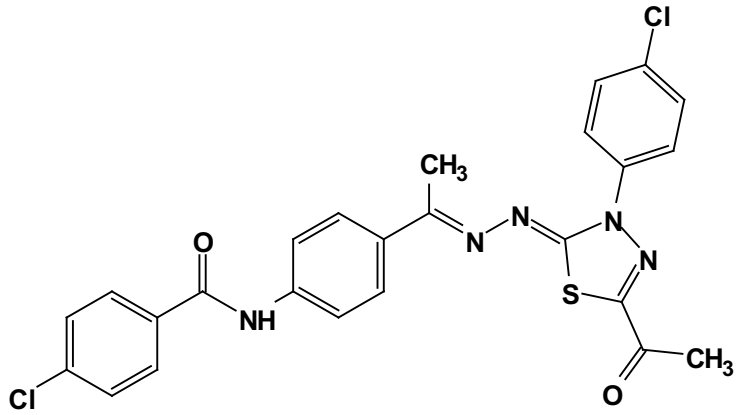
IR of compound 18d



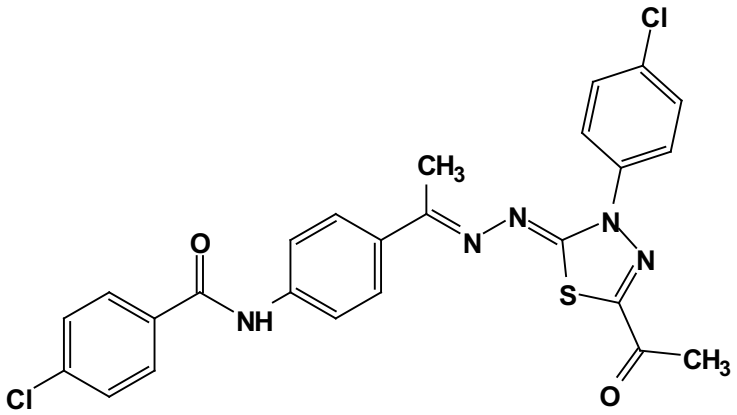
1H NMR of compound 18d



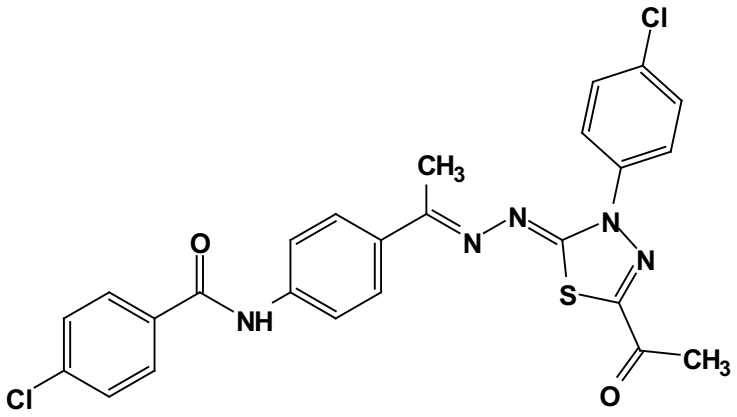
1H NMR of compound 18d



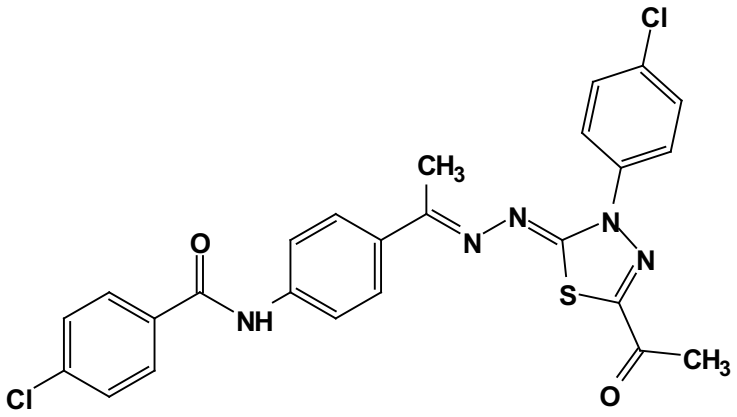
1H NMR of compound 18d



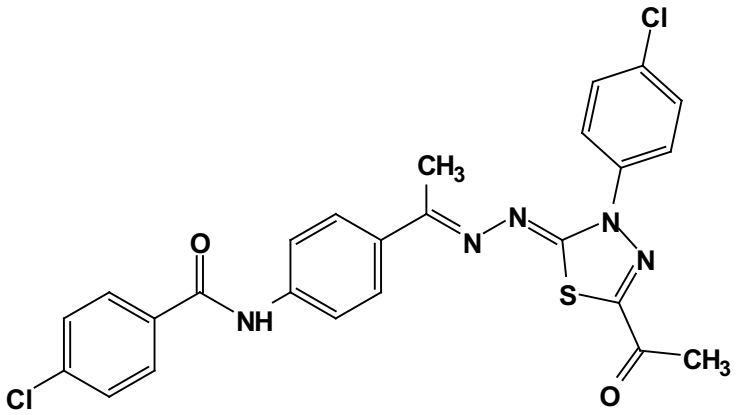
13C NMR of compound 18d



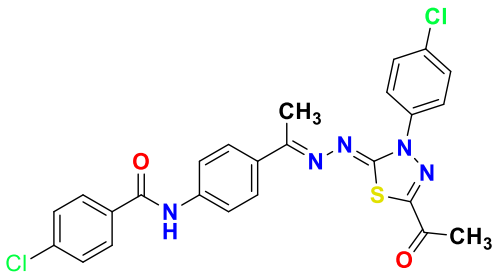
13C NMR of compound 18d



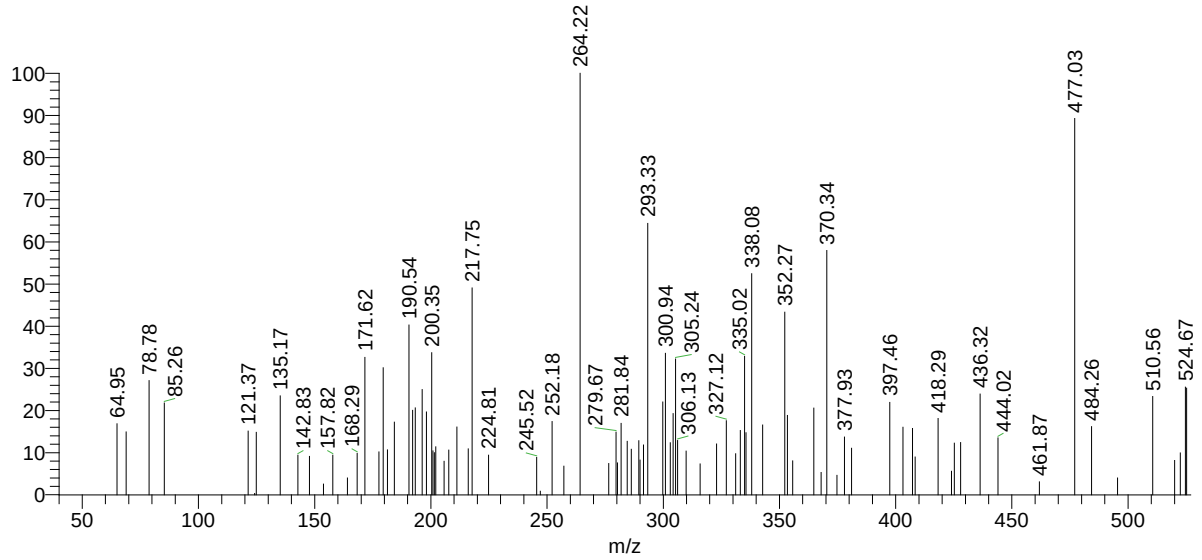
13C NMR of compound 18d

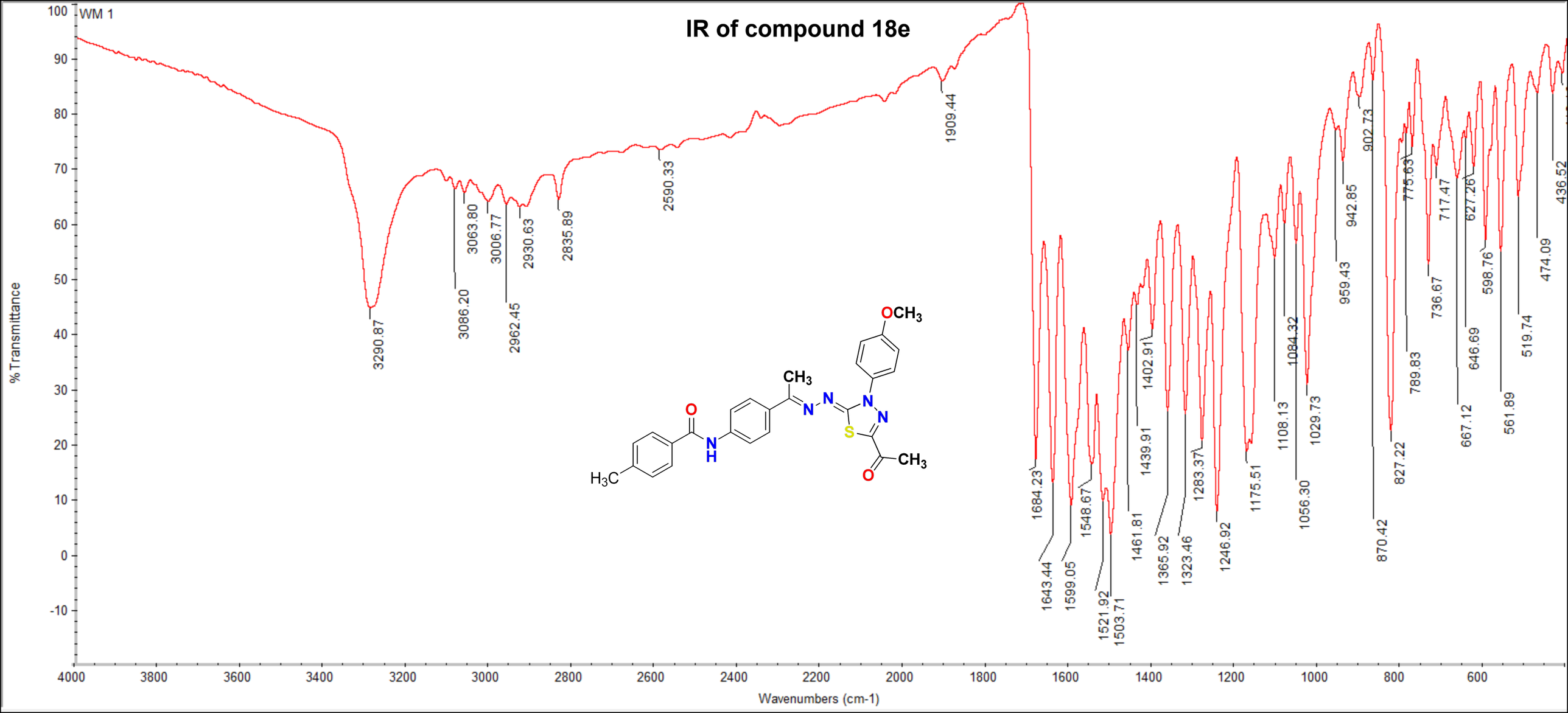


Mass spec. of compound 18d

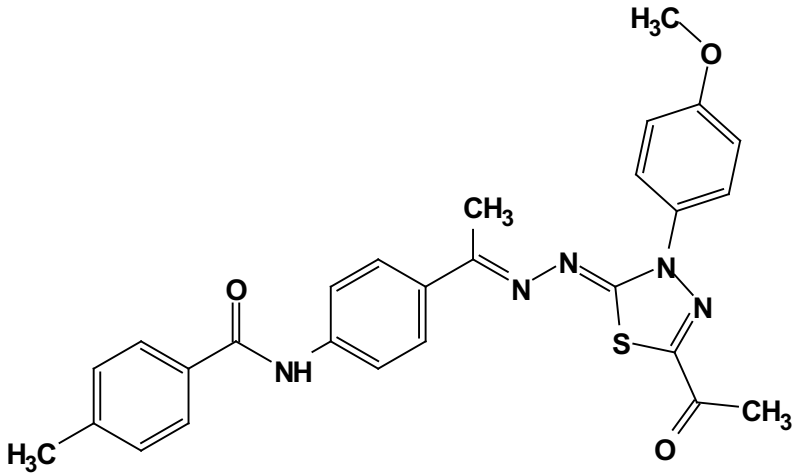


ibrahim-hassan-wc4 #119-120 RT: 2.01-2.03 AV: 2 SB: 26 1.21-1.34 , 0.87-1.14 NL: 3.05E2
T: {0,0} + c EI Full ms [40.00-1000.00]

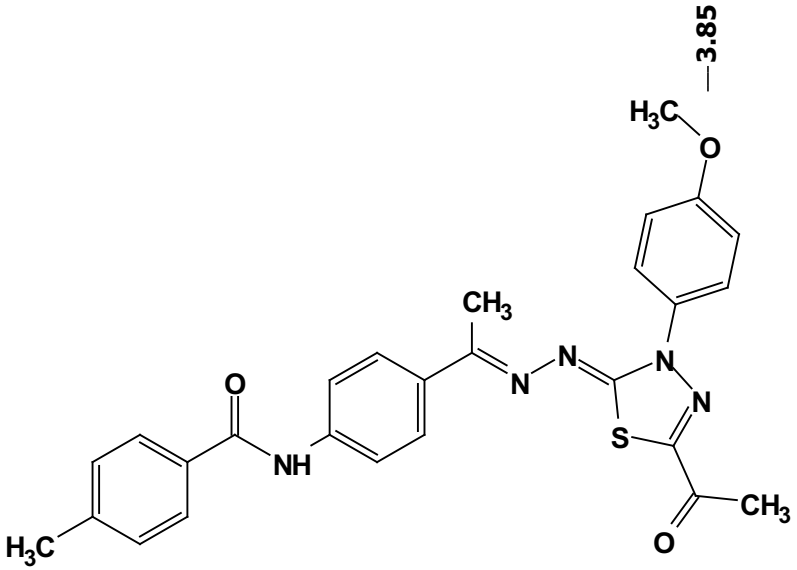



IR of compound 18e
WM 1
% Transmittance
Wavenumbers (cm-1)
3290.87
3086.20
3063.80
3006.77
2962.45
2930.63
2835.89
2590.33
1909.44
1684.23
1643.44
1599.05
1548.67
1521.92
1503.71
1461.81
1439.91
1402.91
1365.92
1323.46
1283.37
1246.92
1175.51
1108.13
1084.32
1056.30
1029.73
959.43
942.85
902.73
870.42
827.22
789.83
775.63
736.67
717.47
667.12
646.69
627.26
598.76
561.89
519.74
474.09
436.52
OCH3
CH3
O
N
H
N
N
N
S
O
CH3
H3C

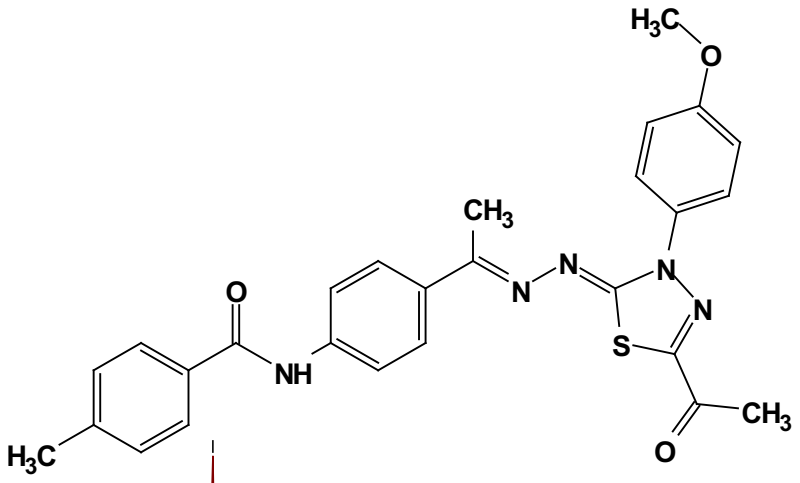
1H NMR of compound 18e



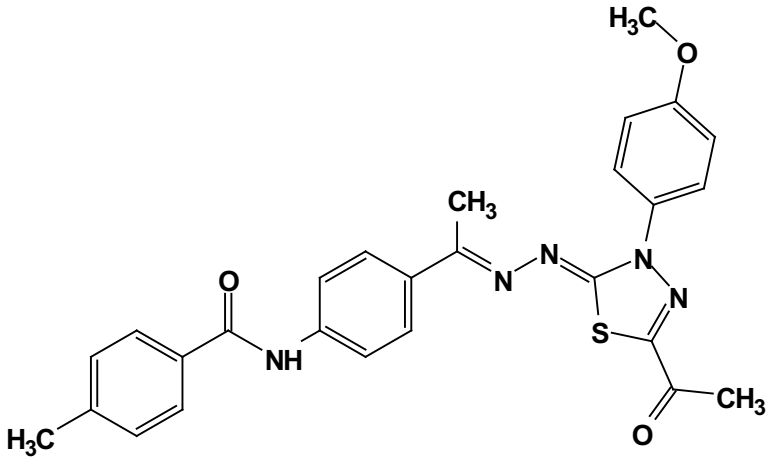
1H NMR of compound 18e



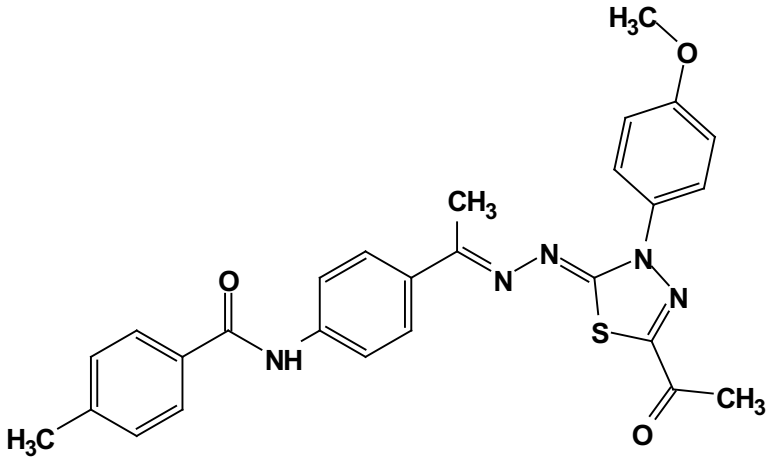
1H NMR of compound 18e



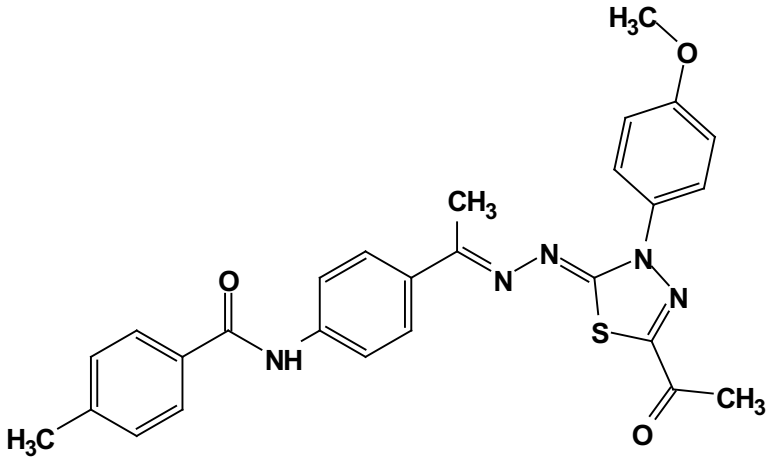
13C NMR of compound 18e



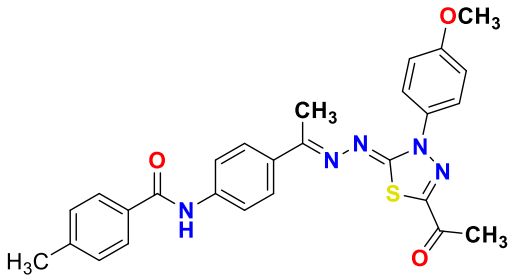
13C NMR of compound 18e



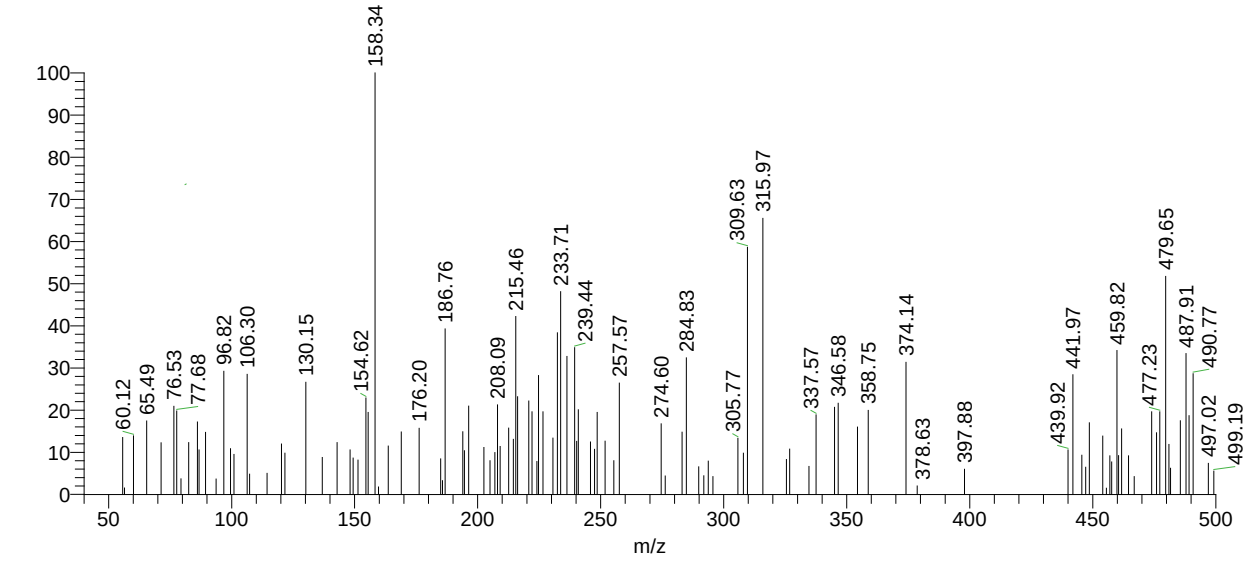
13C NMR of compound 18e

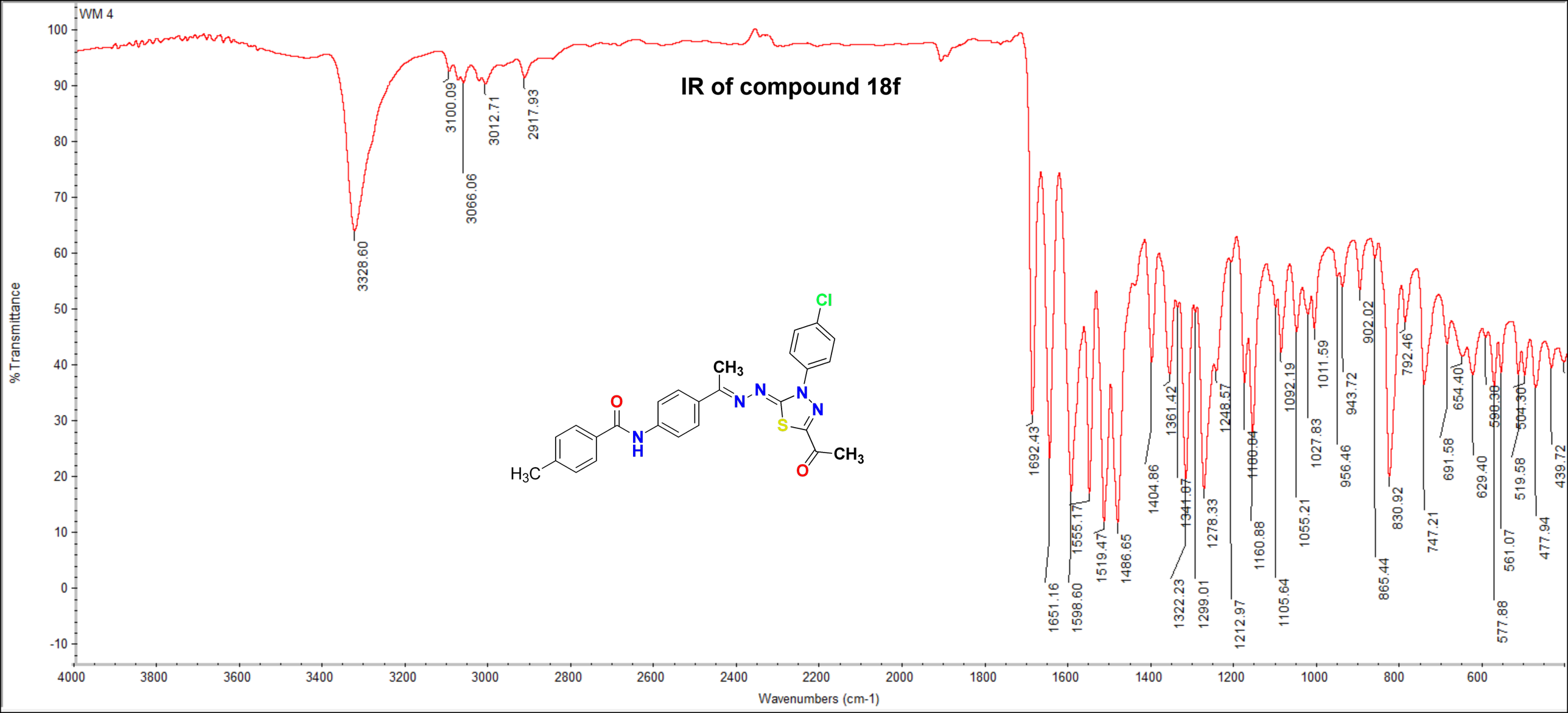


Mass spec. of compound 18e

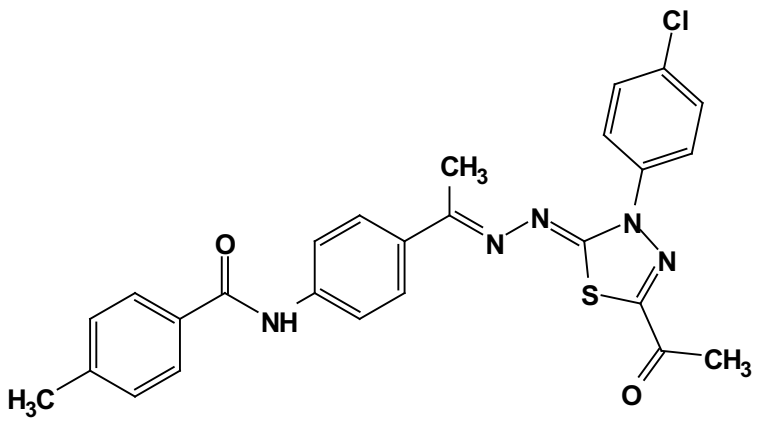


ibrahim-hassan-wm1 #28-29 RT: 0.49-0.50 AV: 2 SB: 26 1.21-1.34 , 0.87-1.14 NL: 2.91E2
T: {0,0} + c EI Full ms [40.00-1000.00]

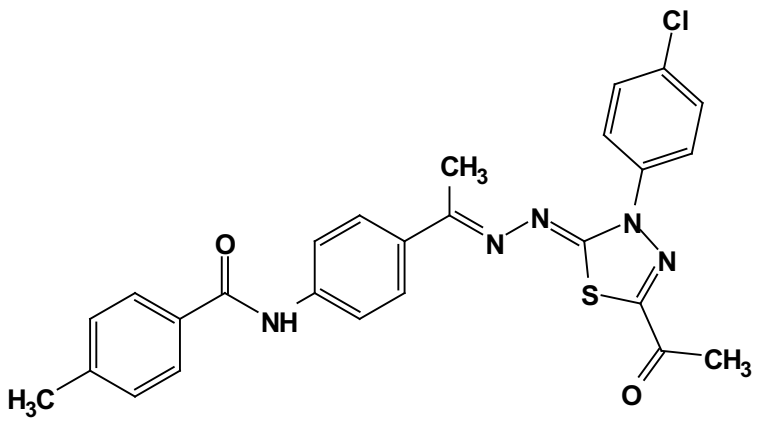



WM 4
IR of compound 18f
% Transmittance
Wavenumbers (cm-1)
3328.60
3100.09
3066.06
3012.71
2917.93
1692.43
1651.16
1598.60
1555.17
1519.47
1486.65
1404.86
1361.42
1341.07
1322.23
1299.01
1278.33
1248.57
1212.97
1180.84
1160.88
1105.64
1092.19
1055.21
1027.83
1011.59
956.46
943.72
902.02
865.44
830.92
792.46
747.21
691.58
654.40
629.40
590.30
577.88
561.07
519.58
504.30
477.94
439.72
Cl
CH3
O
N
N
N
N
S
O
CH3
H3C

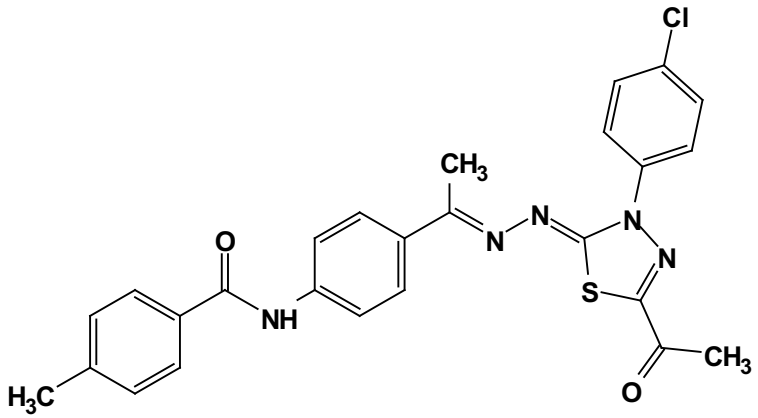
1H NMR of compound 18f



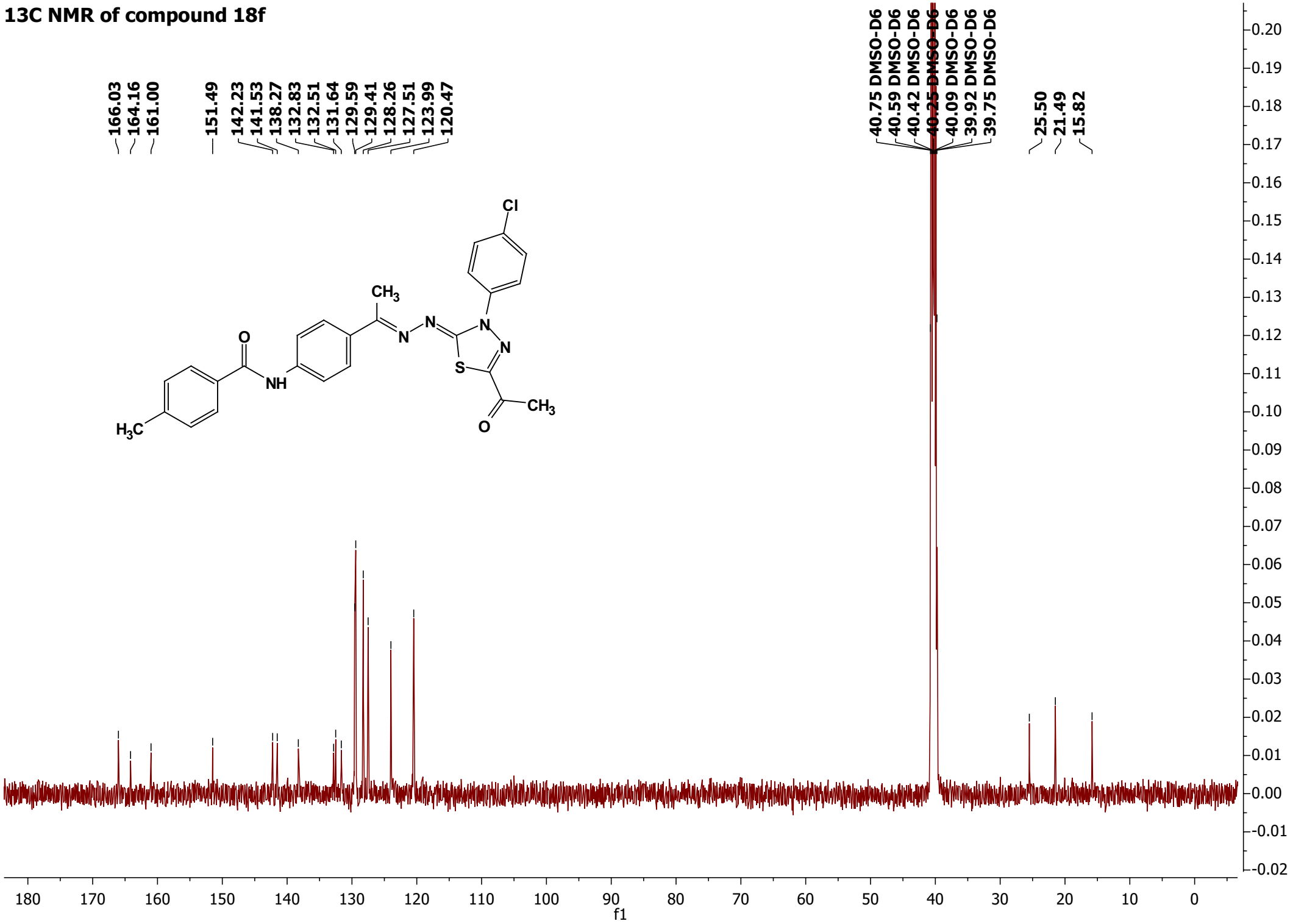
1H NMR of compound 18f



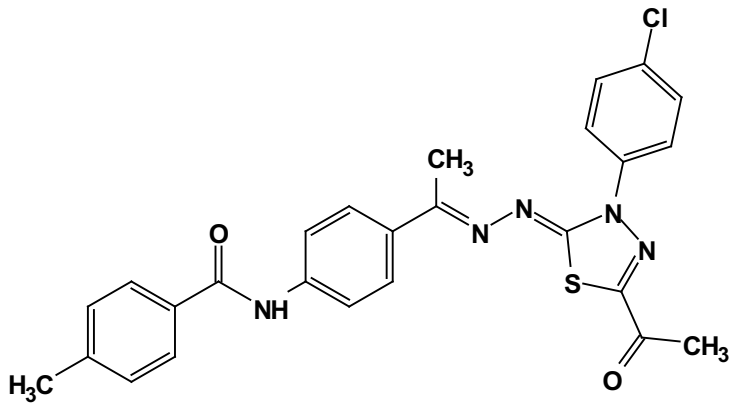
1H NMR of compound 18f



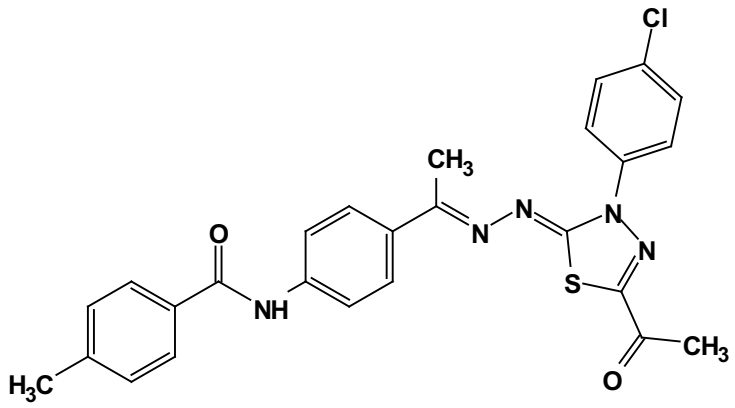
13C NMR of compound 18f



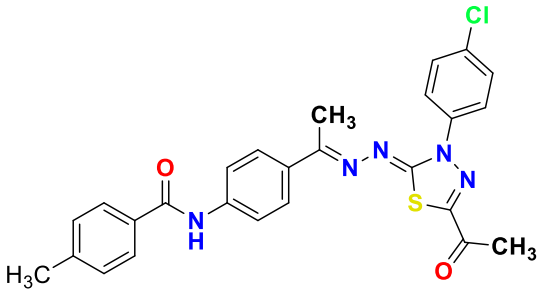
13C NMR of compound 18f



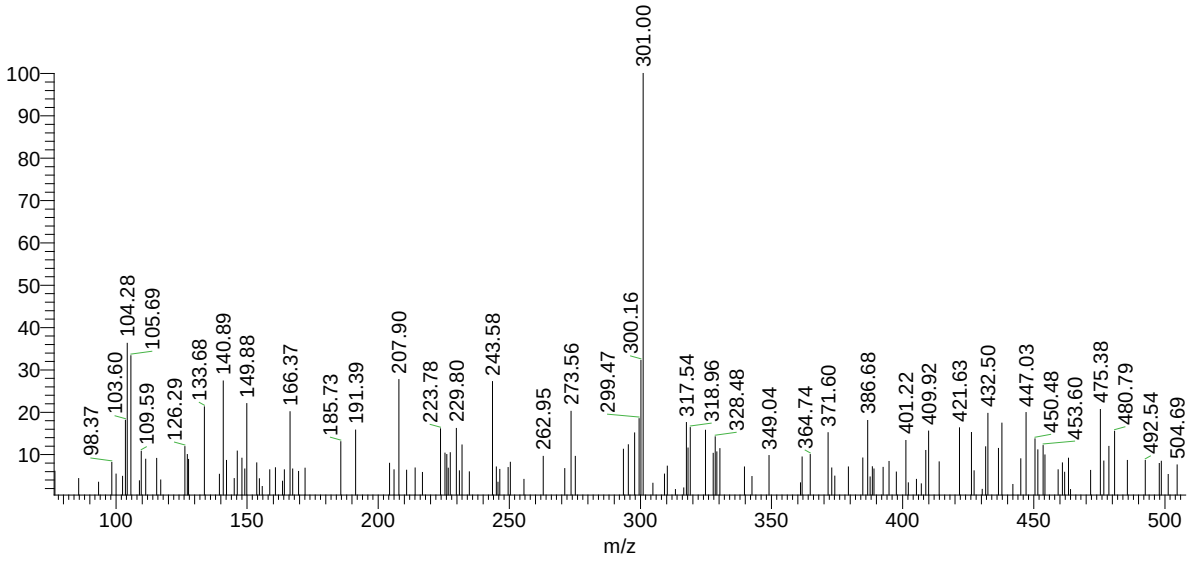
13C NMR of compound 18f

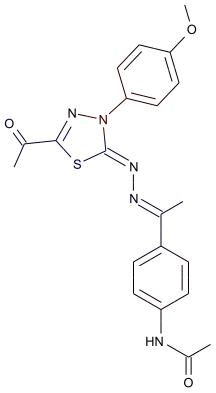


Mass spec. of compound 18f



ibrahim-hassan-wm4 #156 RT: 2.63 AV: 1 SB: 26 1.21-1.34 , 0.87-1.14 NL: 1.18E3
T: {0,0} + c EI Full ms [40.00-1000.00]





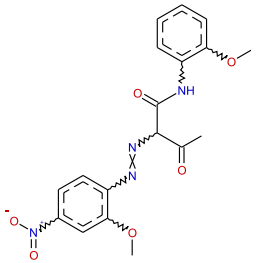
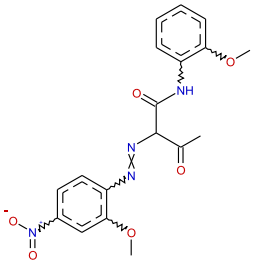
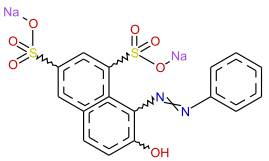
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.679
Enrichment: 1.22
Bayesian Score: -2.56
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 7.78e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	C.I. Pigment Yellow 74	6358-31-2	1936-15-8
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.559	0.560	0.594
Reference	EMIC	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411

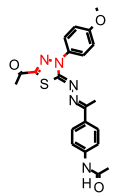
Model Applicability

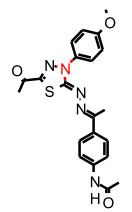
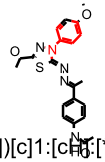
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

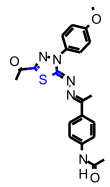
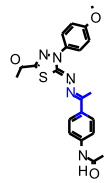
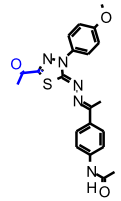
Feature Contribution

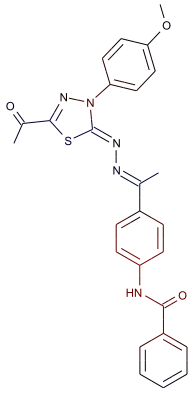
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1325991669	 [*]N1[*][*]C(=N1)[*]	0.362	7 out of 8

SCFP_12	10	 [*]N([*])[*]	0.306	1774 out of 2287
SCFP_12	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.304	957 out of 1235

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	571795252	 [*]C(=[*])C(=O)C	-0.663	31 out of 107



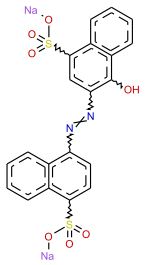
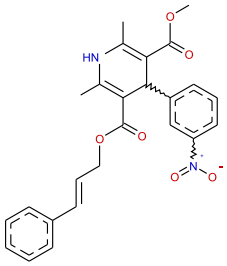
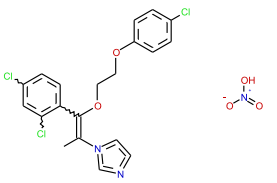
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.725
Enrichment: 1.3
Bayesian Score: -0.696
Mahalanobis Distance: 12.7
Mahalanobis Distance p-value: 1.91e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3567-69-9	99522-79-9	83621-06-1
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.588	0.612	0.613
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

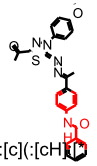
Model Applicability

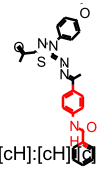
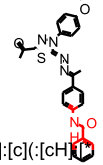
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

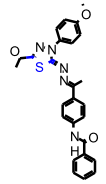
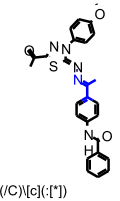
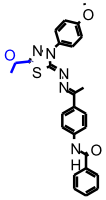
Feature Contribution

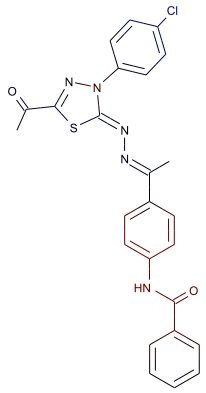
Top features for positive contribution

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

SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	124026986	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	0.429	33 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	571795252	 [*]C(=[*])C(=O)C	-0.663	31 out of 107



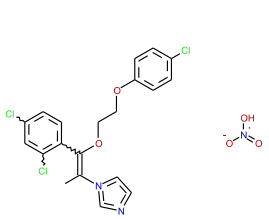
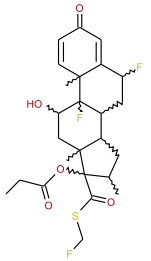
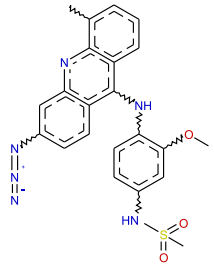
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.668
Enrichment: 1.2
Bayesian Score: -2.95
Mahalanobis Distance: 12.6
Mahalanobis Distance p-value: 4.07e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	83621-06-1	FLUTICASONE	110004-69-8
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Distance	0.614	0.628	0.629
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	PDR 1994	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

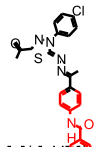
Model Applicability

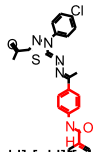
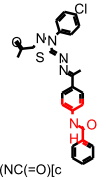
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

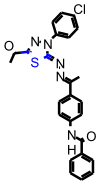
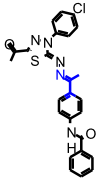
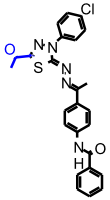
Feature Contribution

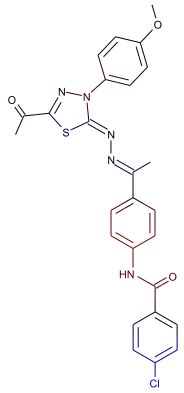
Top features for positive contribution

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

SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	2096901122	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	0.429	33 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	571795252	 [*]C(=[*])C(=O)C	-0.663	31 out of 107



$C_{26}H_{22}ClN_5O_3S$

Molecular Weight: 520.00258

ALogP: 5.392

Rotatable Bonds: 7

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.643

Enrichment: 1.15

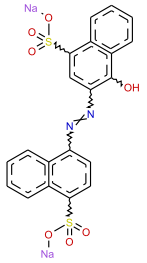
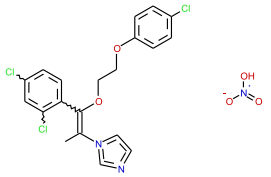
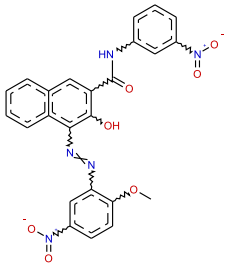
Bayesian Score: -3.78

Mahalanobis Distance: 13

Mahalanobis Distance p-value: 4.69e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3567-69-9	83621-06-1	6471-49-4
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Distance	0.618	0.633	0.641
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

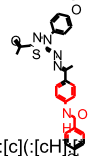
Model Applicability

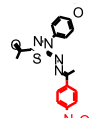
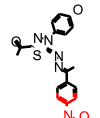
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

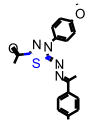
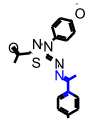
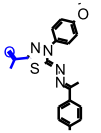
Feature Contribution

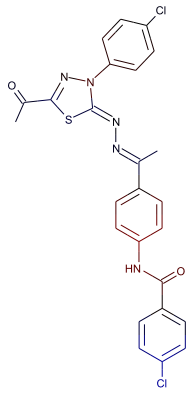
Top features for positive contribution

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

SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	2096901122	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	0.429	33 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	571795252	 [*]C(=[*])C(=O)C	-0.663	31 out of 107



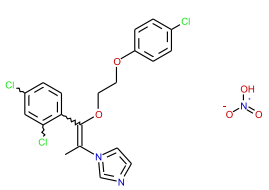
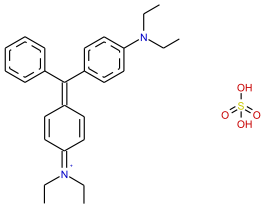
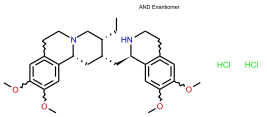
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.639
Enrichment: 1.14
Bayesian Score: -3.91
Mahalanobis Distance: 12.6
Mahalanobis Distance p-value: 3.32e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	83621-06-1	633-03-4	316-42-7
Structure			
Actual Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Distance	0.670	0.671	0.675
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

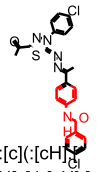
Model Applicability

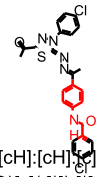
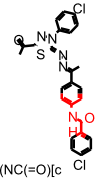
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

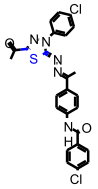
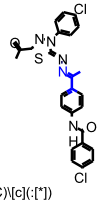
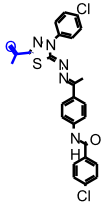
Feature Contribution

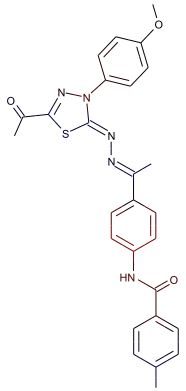
Top features for positive contribution

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

SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	2096901122	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	0.429	33 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	571795252	 [*]C(=[*])C(=O)C	-0.663	31 out of 107



$C_{27}H_{25}N_5O_3S$

Molecular Weight: 499.58409

ALogP: 5.214

Rotatable Bonds: 7

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.711

Enrichment: 1.27

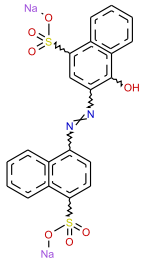
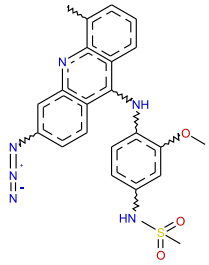
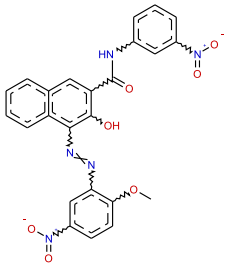
Bayesian Score: -1.3

Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 1.81e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	3567-69-9	110004-69-8	6471-49-4
Structure			
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen
Distance	0.608	0.634	0.637
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

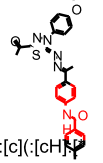
Model Applicability

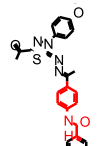
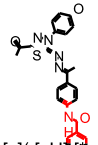
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

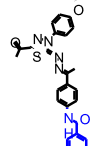
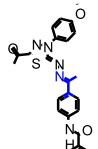
Feature Contribution

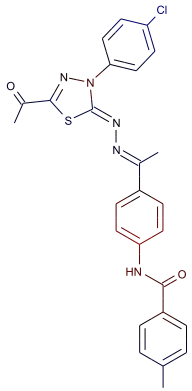
Top features for positive contribution

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

SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	124026986	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	0.429	33 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-1555568408	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.762	0 out of 2
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2



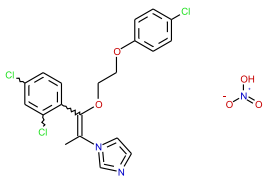
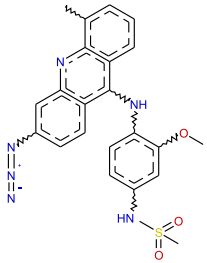
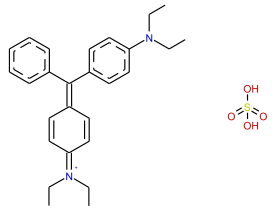
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.634
Enrichment: 1.14
Bayesian Score: -4.05
Mahalanobis Distance: 12.6
Mahalanobis Distance p-value: 3.63e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	83621-06-1	110004-69-8	633-03-4
Structure			
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen
Distance	0.654	0.654	0.663
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

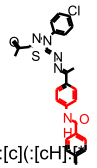
Model Applicability

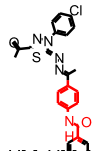
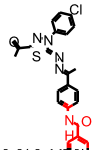
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

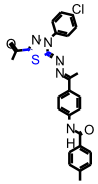
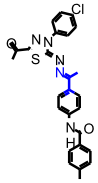
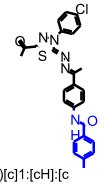
Feature Contribution

Top features for positive contribution

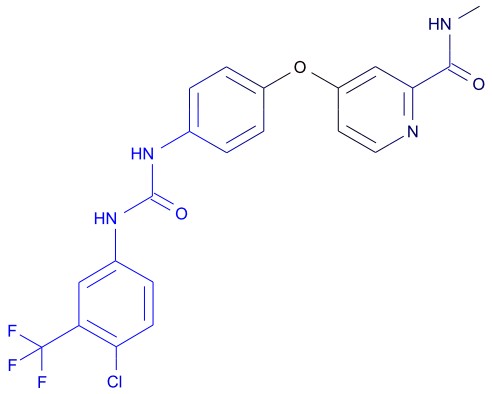
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	555539852	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.447	22 out of 24

SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	124026986	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	0.429	33 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	-1555568408	 [*]NC(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.762	0 out of 2

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

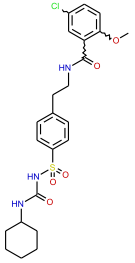
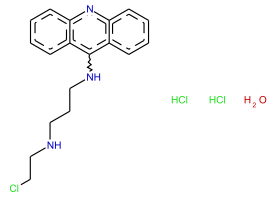
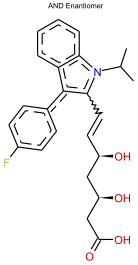
Rotatable Bonds: 6

Acceptors: 4

Donors: 3

TOPKAT_Ames_Mutagenicity

Structural Similar Compounds

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

Model Prediction

Prediction: Non-Mutagen

Probability: 0.0531

Enrichment: 0.0951

Bayesian Score: -19.7

Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 2.73e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

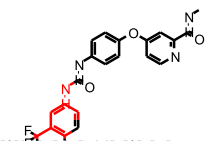
Model Applicability

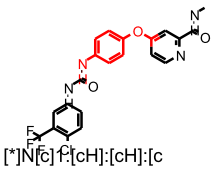
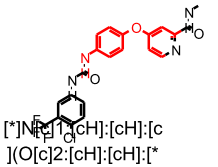
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

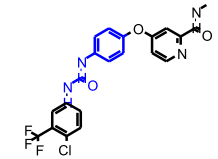
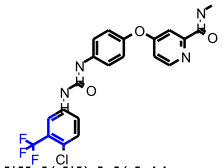
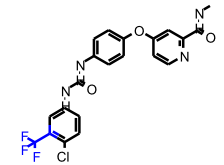
Feature Contribution

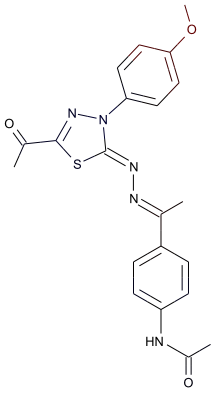
Top features for positive contribution

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

SCFP_12	1208843554		0.337	6 out of 7
SCFP_12	-1943080297		0.304	5 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	816802409		-1.82	0 out of 9
SCFP_12	-1903175541		-1.51	3 out of 30
SCFP_12	-300280774		-1.51	3 out of 30



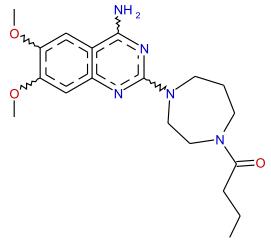
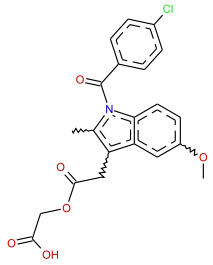
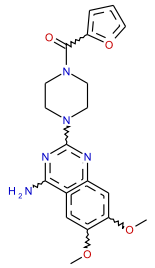
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.553
Enrichment: 1.05
Bayesian Score: 0.136
Mahalanobis Distance: 10.7
Mahalanobis Distance p-value: 0.00314

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Bunazosin .HCl (Free base form)	Acemetacin	Prazosin .HCl (Free base form)
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.590	0.612	0.613
Reference	Kiso to Rinsho 17:914-924; 1983	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 17:57-62; 1979

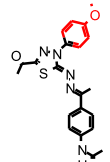
Model Applicability

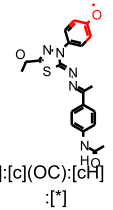
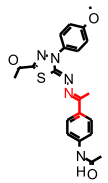
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

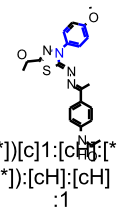
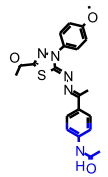
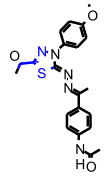
Feature Contribution

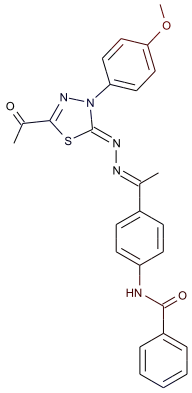
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1237755852	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.453	8 out of 9

SCFP_6	591469355	 [*]:[cH]:[c](OC):[cH] :[*]	0.411	10 out of 12
SCFP_6	-331724199	 [*]\N=C(/C)\[c](:[*]) :[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH] :1	-0.449	6 out of 19
SCFP_6	2097618059	 [*]:[cH]:[c](NC(=O)C) :[cH]:[*]	-0.422	0 out of 1
SCFP_6	-704135030	 [*]C(=[*])C1=N[*][*]S 1	-0.422	0 out of 1



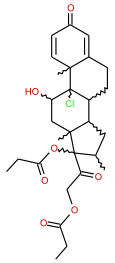
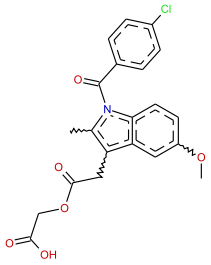
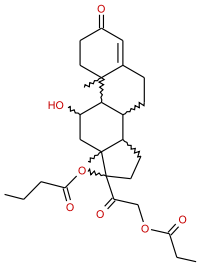
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.602
Enrichment: 1.15
Bayesian Score: 1.42
Mahalanobis Distance: 11.8
Mahalanobis Distance p-value: 7.25e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Beclomethasone Dipropionate	Acemetacin	Hydrocortisone-17-butyrate-21-propionate
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.614	0.616	0.644
Reference	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 21:441-466; 1981

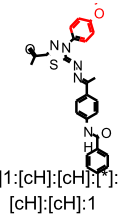
Model Applicability

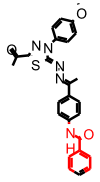
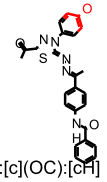
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC17 out of range. Value: 3.4886. Training min, max, SD, explained variance: -2.7025, 2.8536, 1.067, 0.0167.

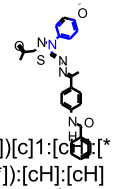
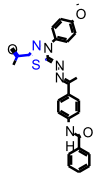
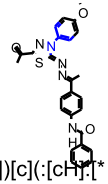
Feature Contribution

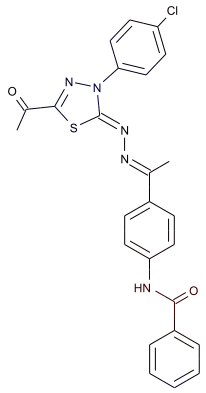
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1237755852	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.453	8 out of 9

SCFP_6	282594097		0.441	3 out of 3
SCFP_6	591469355		0.411	10 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1380909229		-0.449	6 out of 19
SCFP_6	-704135030		-0.422	0 out of 1
SCFP_6	1334669481		-0.355	10 out of 28



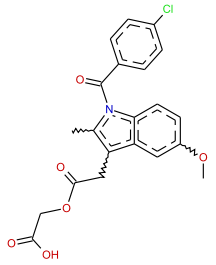
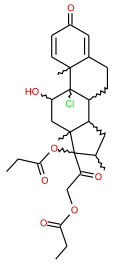
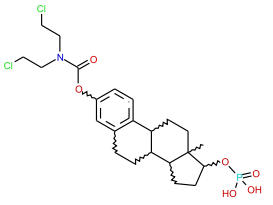
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.533
Enrichment: 1.01
Bayesian Score: -0.387
Mahalanobis Distance: 11.4
Mahalanobis Distance p-value: 0.000251

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Acemetacin	Beclomethasone Dipropionate	Estramustine Phosphate Disodium (Free acid form)
Structure			
Actual Endpoint	Non-Toxic	Toxic	Non-Toxic
Predicted Endpoint	Non-Toxic	Toxic	Non-Toxic
Distance	0.630	0.662	0.662
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 20(6):1219-1236; 1980

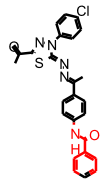
Model Applicability

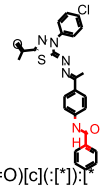
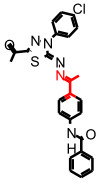
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

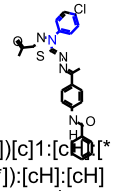
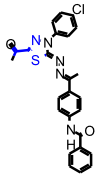
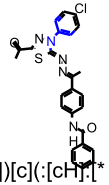
Feature Contribution

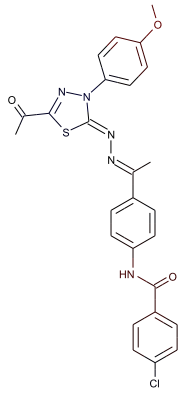
Top features for positive contribution

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

SCFP_6	1257084377		0.362	14 out of 18
SCFP_6	-331724199		0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1380909229		-0.449	6 out of 19
SCFP_6	-704135030		-0.422	0 out of 1
SCFP_6	1334669481		-0.355	10 out of 28



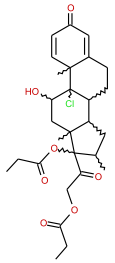
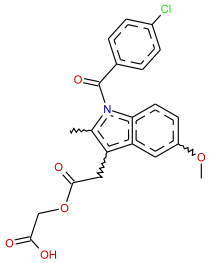
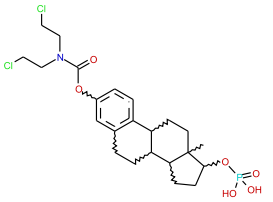
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.61
Enrichment: 1.16
Bayesian Score: 1.62
Mahalanobis Distance: 11
Mahalanobis Distance p-value: 0.00117

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Beclomethasone Dipropionate	Acemetacin	Estramustine Phosphate Disodium (Free acid form)
Structure			
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic
Distance	0.637	0.639	0.657
Reference	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 20(6):1219-1236; 1980

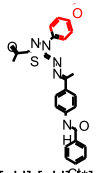
Model Applicability

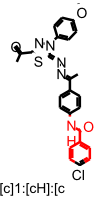
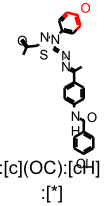
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

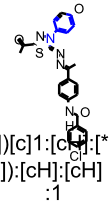
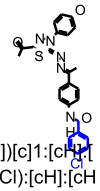
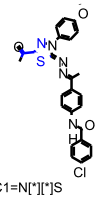
Feature Contribution

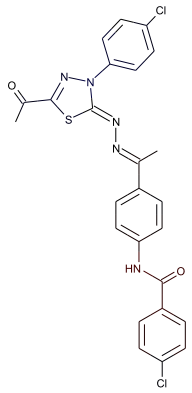
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1237755852	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.453	8 out of 9

SCFP_6	282594097		0.441	3 out of 3
SCFP_6	591469355		0.411	10 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1380909229		-0.449	6 out of 19
SCFP_6	1915307678		-0.422	0 out of 1
SCFP_6	-704135030		-0.422	0 out of 1



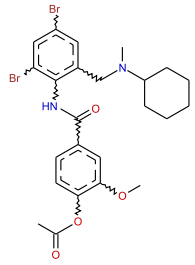
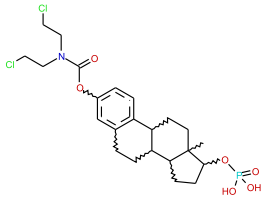
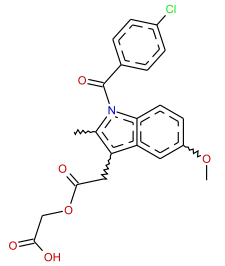
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.543
Enrichment: 1.03
Bayesian Score: -0.127
Mahalanobis Distance: 10.7
Mahalanobis Distance p-value: 0.0025

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Brovanexine .HCl (Free base form)	Estramustine Phosphate Disodium (Free acid form)	Acemetacin
Structure			
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic
Distance	0.651	0.669	0.693
Reference	Kiso to Rinsho 16(13):7179-7195; 1982	Oyo Yakuri 20(6):1219-1236; 1980	Oyo Yakuri 22(6):777-786; 1981

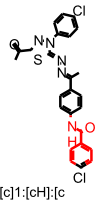
Model Applicability

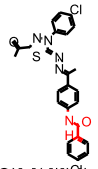
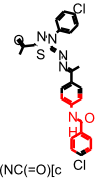
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

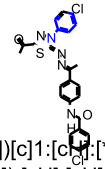
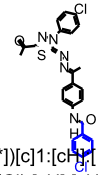
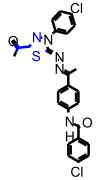
Feature Contribution

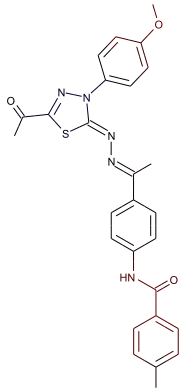
Top features for positive contribution

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

SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	0.362	14 out of 18
SCFP_6	-232641495	 [*]:[cH]:[c](NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1):[cH]:[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH] :1	-0.449	6 out of 19
SCFP_6	1915307678	 [*]C(=[*])[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	-0.422	0 out of 1
SCFP_6	-704135030	 [*]C(=[*])C1=N[*][*]S 1	-0.422	0 out of 1



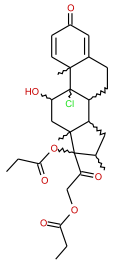
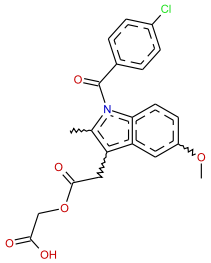
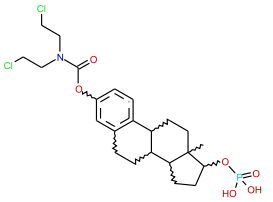
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.64
Enrichment: 1.22
Bayesian Score: 2.38
Mahalanobis Distance: 11.5
Mahalanobis Distance p-value: 0.000188

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Beclomethasone Dipropionate	Acemetacin	Estramustine Phosphate Disodium (Free acid form)
Structure			
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic
Distance	0.631	0.639	0.654
Reference	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 20(6):1219-1236; 1980

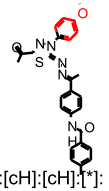
Model Applicability

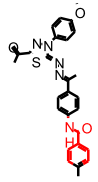
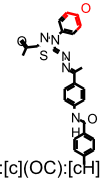
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

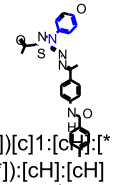
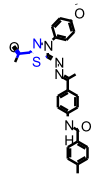
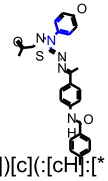
Feature Contribution

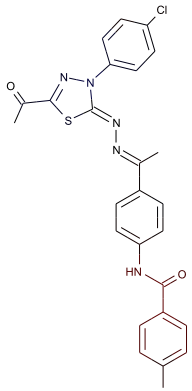
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1237755852	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.453	8 out of 9

SCFP_6	282594097	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.441	3 out of 3
SCFP_6	591469355	 [*]:[cH]:[c](OC):[cH]:[*]	0.411	10 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.449	6 out of 19
SCFP_6	-704135030	 [*]C(=[*])C1=N[*][*]S1	-0.422	0 out of 1
SCFP_6	1334669481	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.355	10 out of 28



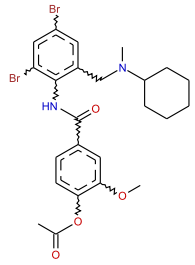
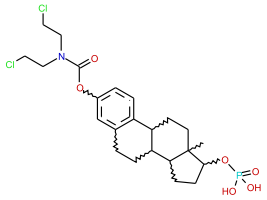
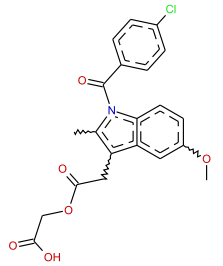
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Toxic
Probability: 0.578
Enrichment: 1.1
Bayesian Score: 0.793
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.000554

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Brovanexine .HCl (Free base form)	Estramustine Phosphate Disodium (Free acid form)	Acemetacin
Structure			
Actual Endpoint	Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Toxic	Non-Toxic	Non-Toxic
Distance	0.656	0.659	0.665
Reference	Kiso to Rinsho 16(13):7179-7195; 1982	Oyo Yakuri 20(6):1219-1236; 1980	Oyo Yakuri 22(6):777-786; 1981

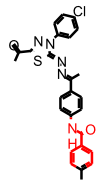
Model Applicability

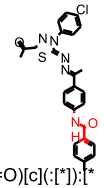
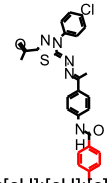
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC2 out of range. Value: 6.7281. Training min, max, SD, explained variance: -4.772, 6.5229, 2.509, 0.0922.
2. OPS PC14 out of range. Value: 4.4177. Training min, max, SD, explained variance: -3.5766, 3.955, 1.214, 0.0216.

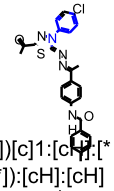
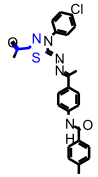
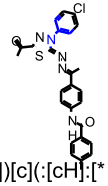
Feature Contribution

Top features for positive contribution

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

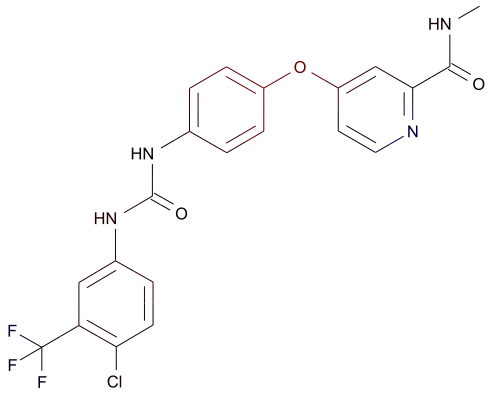
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	0.362	14 out of 18
SCFP_6	795925860	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH] :1	-0.449	6 out of 19
SCFP_6	-704135030	 [*]C(=[*])C1=N[*][*]S 1	-0.422	0 out of 1
SCFP_6	1334669481	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.355	10 out of 28

Sorafenib

TOPKAT_Developmental_Toxicity_Potential



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Toxic

Probability: 0.592

Enrichment: 1.13

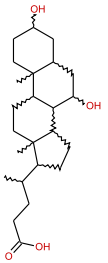
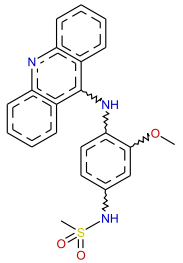
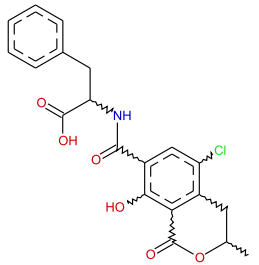
Bayesian Score: 1.15

Mahalanobis Distance: 12.6

Mahalanobis Distance p-value: 2.07e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Chenodiol	Amsacrine	Ochratoxin a
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.631	0.637	0.644
Reference	Arch Int Pharm 246:149-158; 1980	Fundam Appl Toxicol 7(2):214-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976

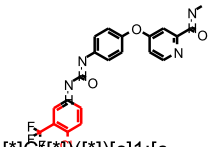
Model Applicability

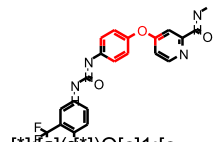
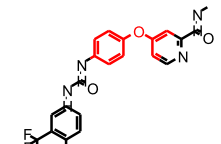
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

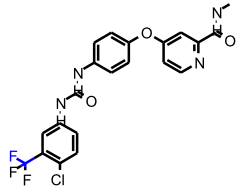
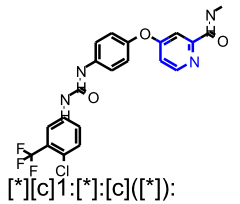
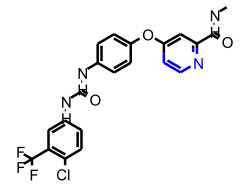
Feature Contribution

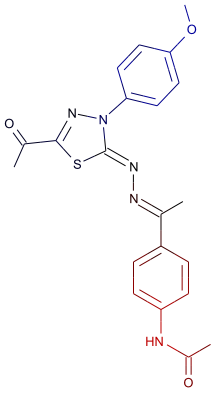
Top features for positive contribution

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

SCFP_6	-488587948	 [*]:[c](:[*])O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.381	2 out of 2
SCFP_6	-975241316	 [*][c]1:[cH]:[cH]:[c](O[c](:[cH]:[*]):[cH]:[*]):[cH]:[cH]:1	0.381	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1794974220	 [*]C([*])([*])F	-0.55	2 out of 8
SCFP_6	-937094999	 [*][c]1:[*]:[c]([*]):n:[cH]:[cH]:1	-0.358	3 out of 9
SCFP_6	-496201075	 [*]:[cH]:[cH]:n:[*]	-0.289	8 out of 21



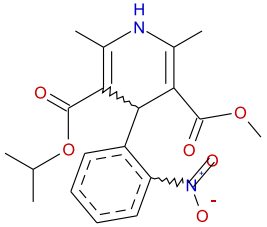
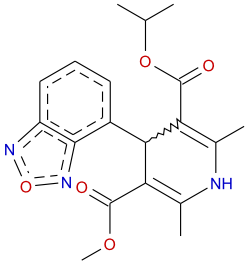
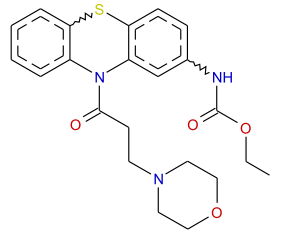
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.225
Enrichment: 0.701
Bayesian Score: -2.52
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 0.0223

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Nisoldipine	Isradipine	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.605	0.608	0.624
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

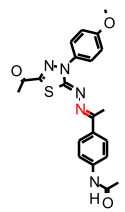
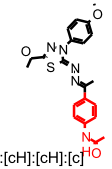
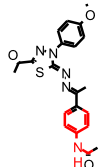
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC4 out of range. Value: 4.791. Training min, max, SD, explained variance: -4.7116, 4.7287, 2.103, 0.0450.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

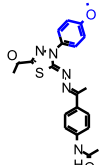
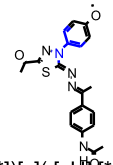
Feature Contribution

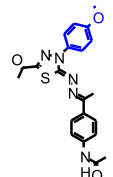
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

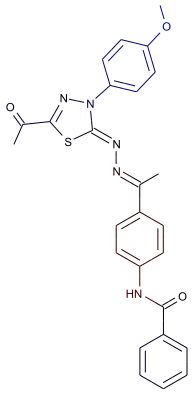
ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14
ECFP_6	-847011520	 [*][c]1:[cH]:[cH]:[c](NC(=O)C):[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1271104377	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.805	0 out of 4
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.805	0 out of 4

ECFP_6	693720869	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.805	0 out of 4

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



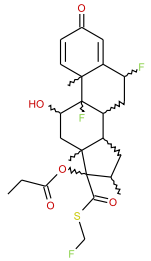
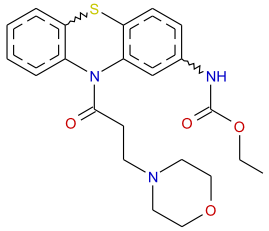
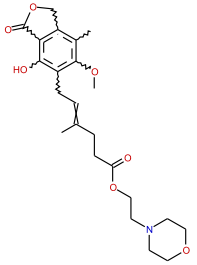
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.206
Enrichment: 0.641
Bayesian Score: -5.52
Mahalanobis Distance: 12.8
Mahalanobis Distance p-value: 0.00101

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Fluticasone	Moricizine	Mycophenolate
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.672	0.696	0.721
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

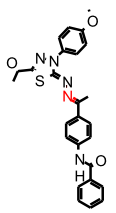
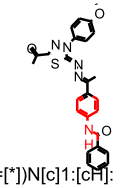
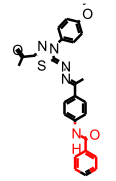
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

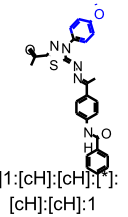
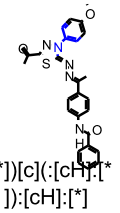
Feature Contribution

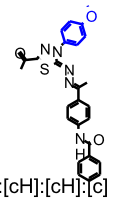
Top features for positive contribution

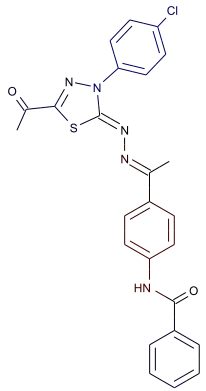
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.442	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1271104377	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.805	0 out of 4
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]:[*]):[cH]:[*]	-0.805	0 out of 4

ECFP_6	693720869	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.805	0 out of 4



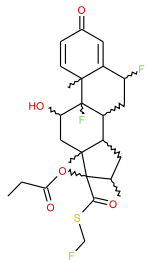
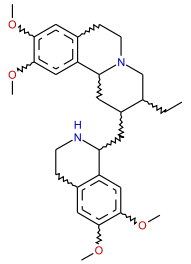
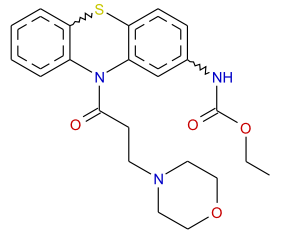
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.21
Enrichment: 0.655
Bayesian Score: -4.37
Mahalanobis Distance: 14.6
Mahalanobis Distance p-value: 1.1e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Emetine	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.632	0.692	0.698
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

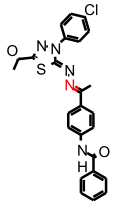
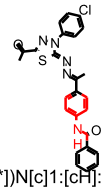
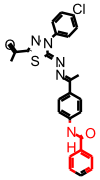
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

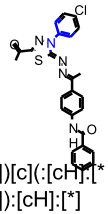
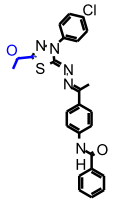
Feature Contribution

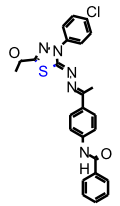
Top features for positive contribution

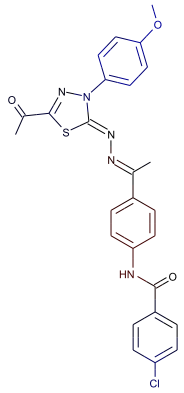
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.442	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.805	0 out of 4
ECFP_6	129482634	 [*]C(=[*])C(=O)C	-0.657	0 out of 3

ECFP_6	912478223	 [*]S[*]	-0.638	1 out of 9



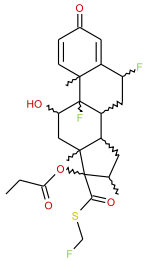
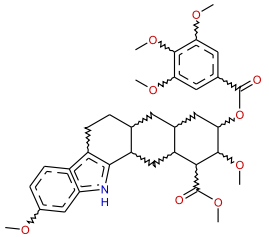
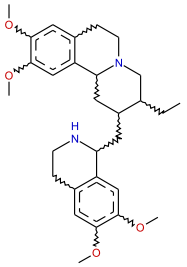
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.205
Enrichment: 0.639
Bayesian Score: -6.56
Mahalanobis Distance: 14.1
Mahalanobis Distance p-value: 9.92e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Reserpine	Emetine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.697	0.717	0.730
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

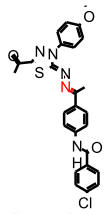
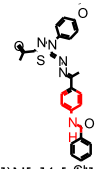
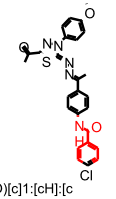
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC4 out of range. Value: 5.0431. Training min, max, SD, explained variance: -4.7116, 4.7287, 2.103, 0.0450.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

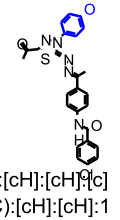
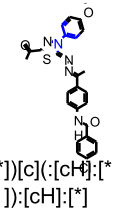
Feature Contribution

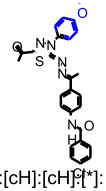
Top features for positive contribution

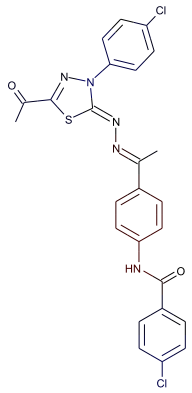
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.442	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	693720869	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.805	0 out of 4
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.805	0 out of 4

ECFP_6	-1271104377	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.805	0 out of 4



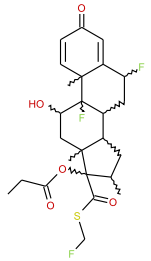
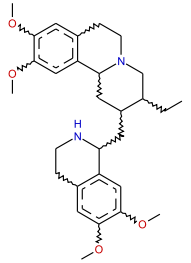
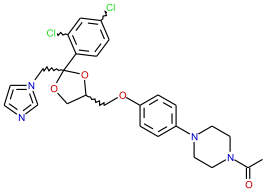
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.212
Enrichment: 0.662
Bayesian Score: -3.98
Mahalanobis Distance: 14.2
Mahalanobis Distance p-value: 6.17e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Emetine	Ketoconazole
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.686	0.714	0.757
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

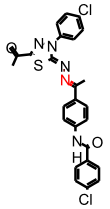
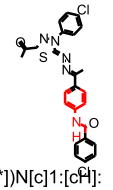
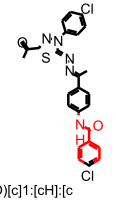
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

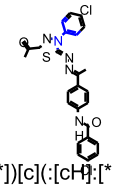
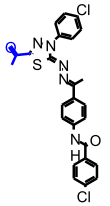
Feature Contribution

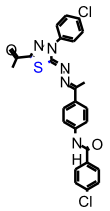
Top features for positive contribution

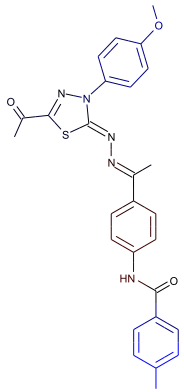
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.442	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[*]):[cH]:[*]	-0.805	0 out of 4
ECFP_6	129482634	 [*]C(=[*])C(=O)C	-0.657	0 out of 3

ECFP_6	912478223	 [*]S[*]	-0.638	1 out of 9



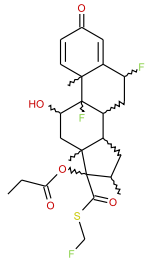
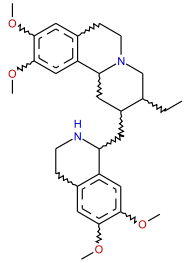
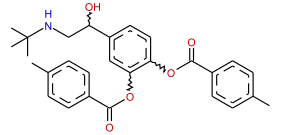
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.224
Enrichment: 0.698
Bayesian Score: -10
Mahalanobis Distance: 12.6
Mahalanobis Distance p-value: 0.00212

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Fluticasone	Emetine	Bitolterol
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.686	0.720	0.737
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

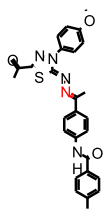
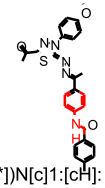
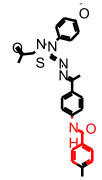
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC4 out of range. Value: 5.1129. Training min, max, SD, explained variance: -4.7116, 4.7287, 2.103, 0.0450.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

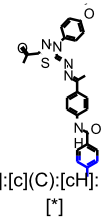
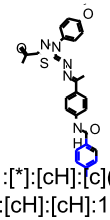
Feature Contribution

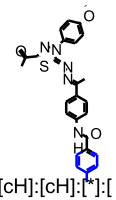
Top features for positive contribution

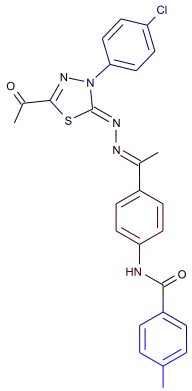
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.442	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-179515162	 [*]:[cH]:[c](C):[cH]:[*]	-1.41	0 out of 10
ECFP_6	-210573707	 [*][c]1:[*]:[cH]:[c](C):[cH]:[cH]:1	-1.25	0 out of 8

ECFP_6	-533780882	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.05	0 out of 6



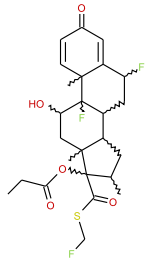
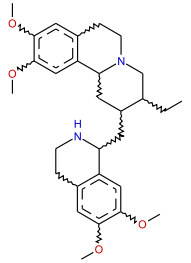
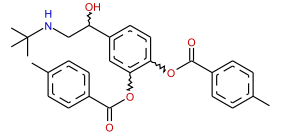
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.216
Enrichment: 0.674
Bayesian Score: -9.2
Mahalanobis Distance: 14.2
Mahalanobis Distance p-value: 7.52e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Emetine	Bitolterol
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.668	0.699	0.745
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

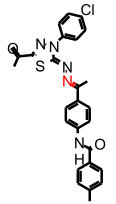
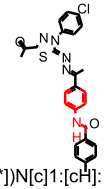
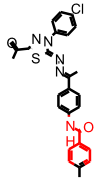
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

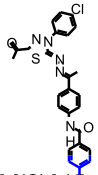
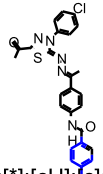
Feature Contribution

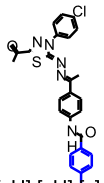
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.442	2 out of 3

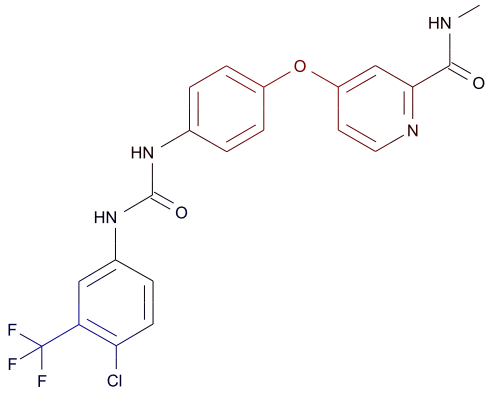
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-179515162	 [*]:[cH]:[c](C):[cH]:[*]	-1.41	0 out of 10
ECFP_6	-210573707	 [*][c]1:[*]:[cH]:[c](C):[cH]:[cH]:1	-1.25	0 out of 8

ECFP_6	-1926229349	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.05	0 out of 6

Sorafenib

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.257

Enrichment: 0.801

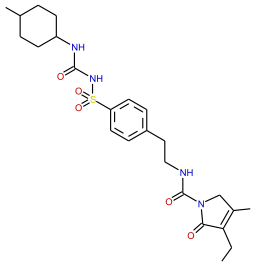
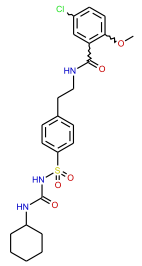
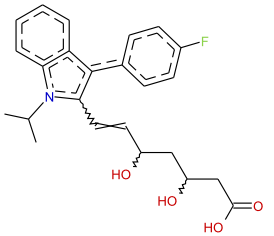
Bayesian Score: -0.321

Mahalanobis Distance: 14.9

Mahalanobis Distance p-value: 4.21e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Glimepride	Glyburide	Fluvastatin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.605	0.615	0.625
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

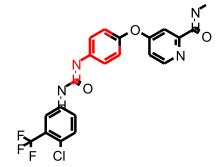
Model Applicability

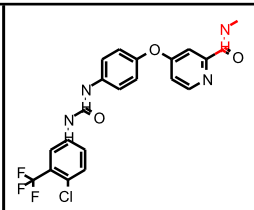
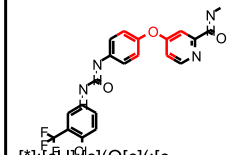
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

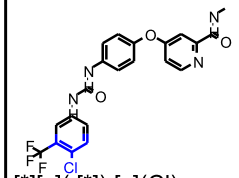
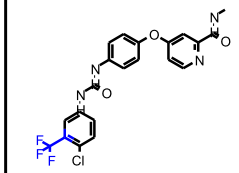
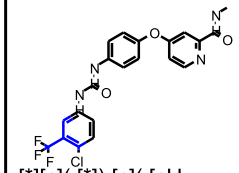
Feature Contribution

Top features for positive contribution

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

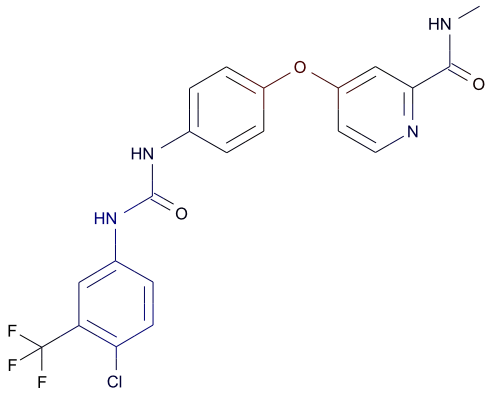
ECFP_6	1338334141	 [*]C(=[*])NC	0.442	2 out of 3
ECFP_6	-834094296	 [*]:[cH]:[c](O[c](:[cH]:[*]):[cH]:[*]):[cH]:[*]	0.424	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.669	3 out of 22
ECFP_6	-1952889961	 [*]:[c](:[*])C(F)(F)F	-0.657	0 out of 3
ECFP_6	1336678434	 [*][c](:[*]):[c](:[cH]:[*])C([*])([*])[*]	-0.657	0 out of 3

Sorafenib

TOPKAT_Mouse_Female_FDA_Single_vs_Multiple



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.283

Enrichment: 0.691

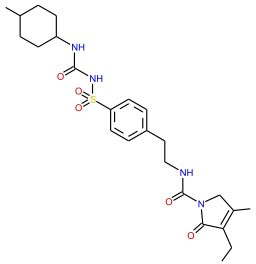
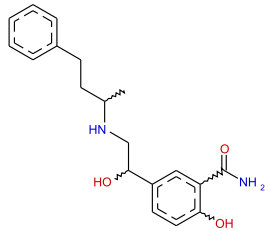
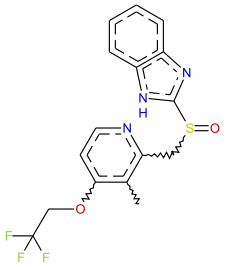
Bayesian Score: -3.89

Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.00221

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Glimepride	Labetalol	Lansoprazole
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.599	0.808	0.820
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

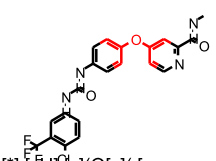
Model Applicability

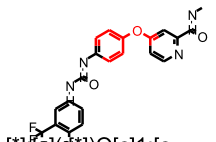
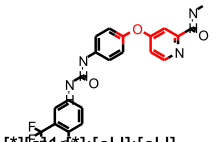
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

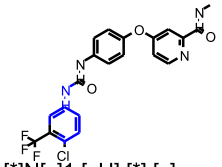
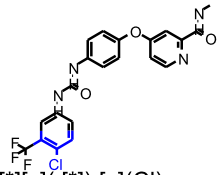
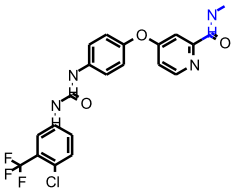
Feature Contribution

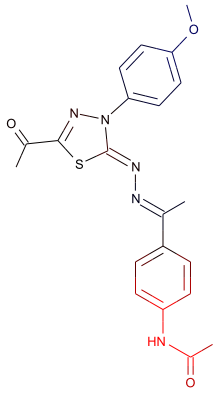
Top features for positive contribution

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

ECFP_4	1407472008	 [*]:[c](:[*])O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.351	1 out of 1
ECFP_4	143734695	 [*][c]1:[*]:[cH]:[cH]:[c](O[c](:[*]):[*]):[cH]:1	0.351	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.8	0 out of 3
ECFP_4	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.8	0 out of 3
ECFP_4	1338334141	 [*]C(=[*])NC	-0.597	0 out of 2



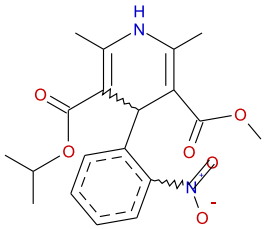
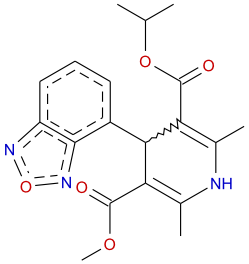
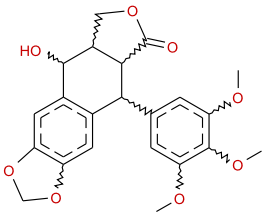
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.408
Enrichment: 1.39
Bayesian Score: 3.23
Mahalanobis Distance: 14.3
Mahalanobis Distance p-value: 1.62e-006

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Nisoldipine	Isradipine	Podofilox
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.575	0.587	0.593
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

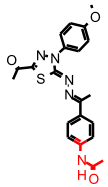
Model Applicability

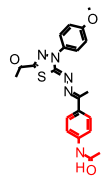
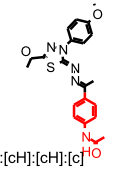
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

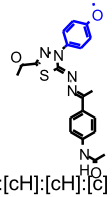
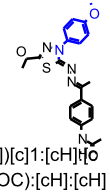
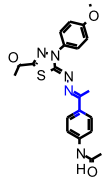
Feature Contribution

Top features for positive contribution

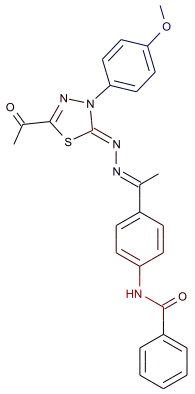
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1944671191	 [*]:[c](:[*])NC(=O)C	0.891	4 out of 4

FCFP_6	1907952166	 CC(=O)N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.805	3 out of 3
FCFP_6	-451043714	 [*][c]1:[cH]:[cH]:[c](NC(=O)C):[cH]:[cH]:1	0.676	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-9847677	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.719	0 out of 4
FCFP_6	356782498	 [*]N([*])[c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



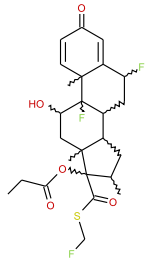
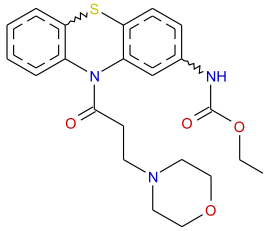
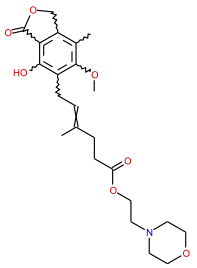
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.283
Enrichment: 0.963
Bayesian Score: -0.858
Mahalanobis Distance: 16.9
Mahalanobis Distance p-value: 9.73e-012

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Moricizine	Mycophenolate
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.661	0.684	0.704
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

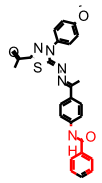
Model Applicability

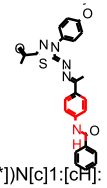
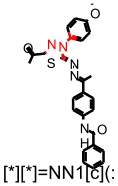
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

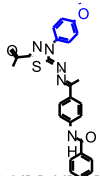
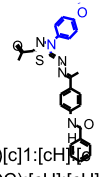
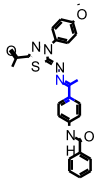
Feature Contribution

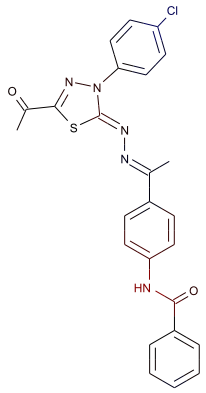
Top features for positive contribution

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

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-9847677	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.719	0 out of 4
FCFP_6	356782498	 [*]N([*])[c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21



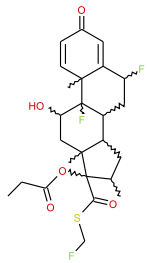
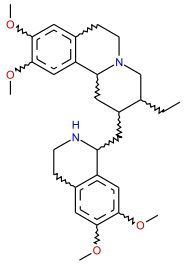
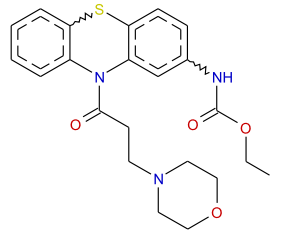
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.301
Enrichment: 1.02
Bayesian Score: -0.185
Mahalanobis Distance: 17.5
Mahalanobis Distance p-value: 4.16e-013

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Emetine	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.615	0.672	0.684
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

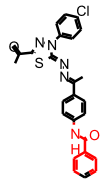
Model Applicability

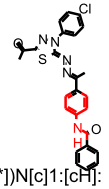
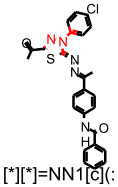
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

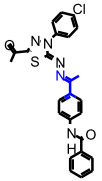
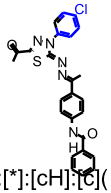
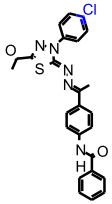
Feature Contribution

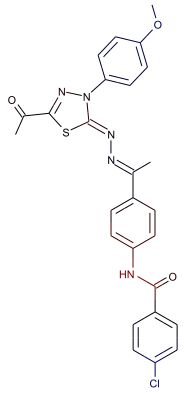
Top features for positive contribution

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

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21
FCFP_6	551850122	 [*][c]1:[*]:[cH]:[c](Cl):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59



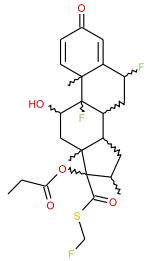
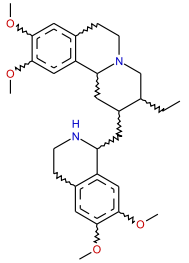
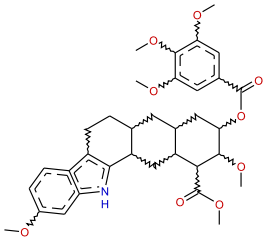
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.262
Enrichment: 0.892
Bayesian Score: -1.73
Mahalanobis Distance: 16.7
Mahalanobis Distance p-value: 3e-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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Emetine	Reserpine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.684	0.708	0.709
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

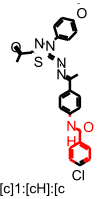
Model Applicability

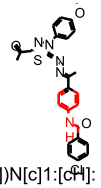
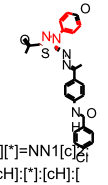
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

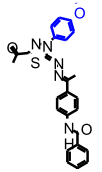
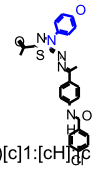
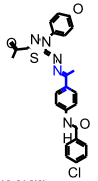
Feature Contribution

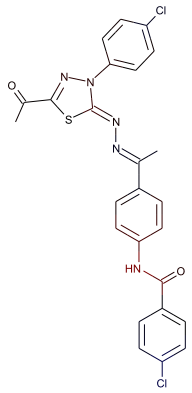
Top features for positive contribution

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

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	-319371573	 [*]=C1[*][*]=NN1[c]2:[cH]:[cH]:[*]:[cH]:[cH]:2	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-9847677	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.719	0 out of 4
FCFP_6	356782498	 [*]N([*])[c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21



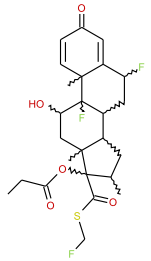
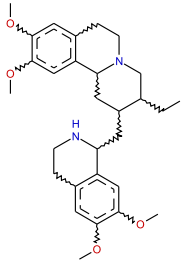
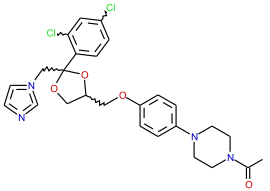
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.325
Enrichment: 1.11
Bayesian Score: 0.689
Mahalanobis Distance: 17.6
Mahalanobis Distance p-value: 3.41e-013

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Emetine	Ketoconazole
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.669	0.695	0.737
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

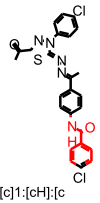
Model Applicability

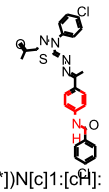
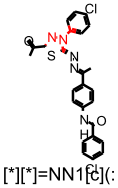
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

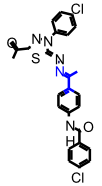
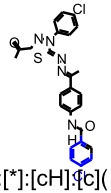
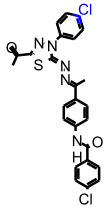
Feature Contribution

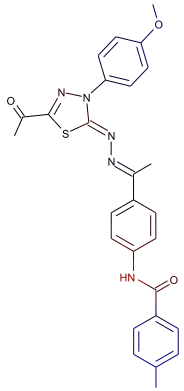
Top features for positive contribution

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

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21
FCFP_6	551850122	 [*][c]1:[*]:[cH]:[c](Cl):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59



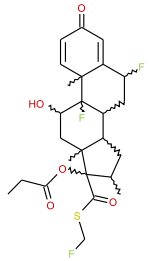
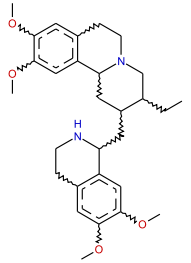
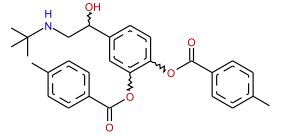
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.24
Enrichment: 0.814
Bayesian Score: -2.78
Mahalanobis Distance: 17.2
Mahalanobis Distance p-value: 2.63e-012

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Emetine	Bitolterol
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.676	0.698	0.724
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

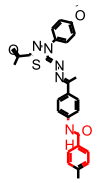
Model Applicability

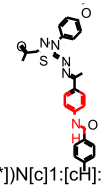
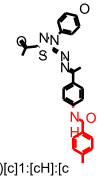
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

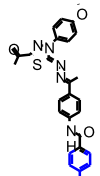
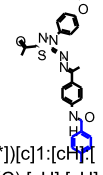
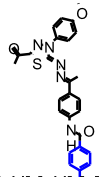
Feature Contribution

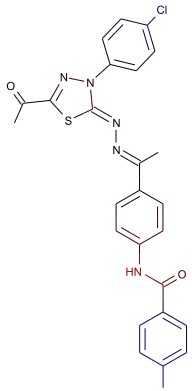
Top features for positive contribution

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

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	1790572653	 [*]NC(=O)[c]1:[cH]:[cH]:[c]:[C]:[cH]:[cH]:1	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 10
FCFP_6	632767364	 [*]C(=[*])[c]1:[cH]:[cH]:[c]:[C]:[cH]:[cH]:1	-1.04	0 out of 7
FCFP_6	2109043264	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.947	0 out of 6



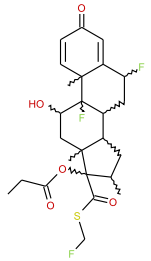
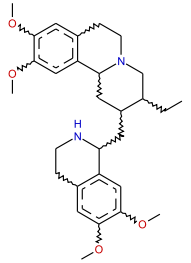
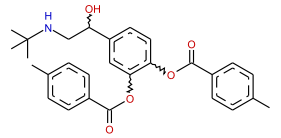
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.259
Enrichment: 0.879
Bayesian Score: -1.9
Mahalanobis Distance: 18.6
Mahalanobis Distance p-value: 1.44e-015

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Fluticasone	Emetine	Bitolterol
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.651	0.679	0.735
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

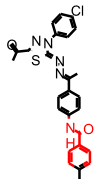
Model Applicability

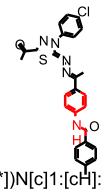
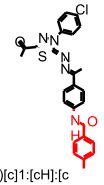
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

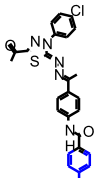
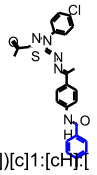
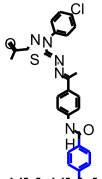
Feature Contribution

Top features for positive contribution

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

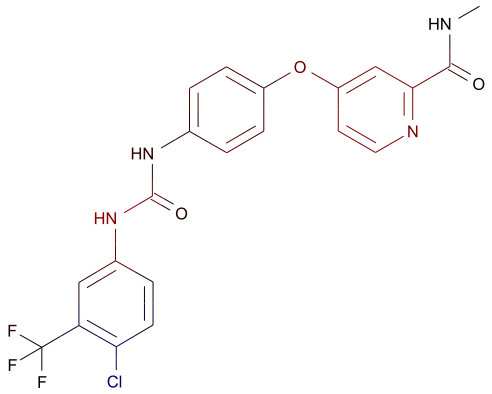
FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	1790572653	 [*]NC(=O)[c]1:[cH]:[cH]:[c]([C]):[cH]:[cH]:1	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 10
FCFP_6	632767364	 [*]C(=[*])[c]1:[cH]:[cH]:[c]([C]):[cH]:[cH]:1	-1.04	0 out of 7
FCFP_6	2109043264	 [*][c]1:[cH]:[cH]:[c]([C]):[cH]:[cH]:1	-0.947	0 out of 6

Sorafenib

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.444

Enrichment: 1.51

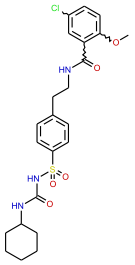
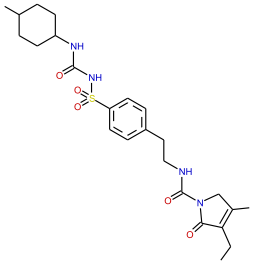
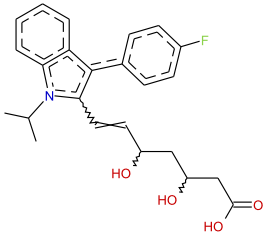
Bayesian Score: 4.21

Mahalanobis Distance: 20.3

Mahalanobis Distance p-value: 1.28e-019

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Glyburide	Glimepride	Fluvastatin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.594	0.599	0.603
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

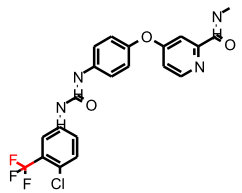
Model Applicability

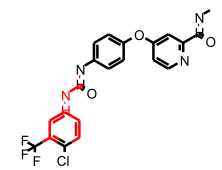
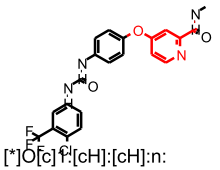
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

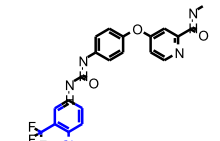
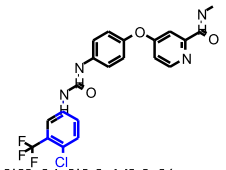
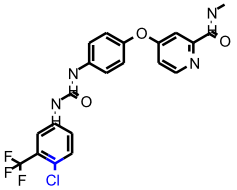
Feature Contribution

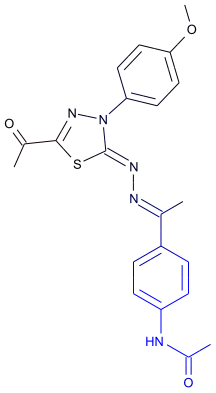
Top features for positive contribution

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

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	140656626	 [*]O[c]1:[cH]:[cH]:n:[c](:[cH]:1)C(=[*])[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	2104062943	 [*]C([*])([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	-1.01	1 out of 17
FCFP_6	551850122	 [*][c]1:[*]:[cH]:[c](Cl):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59



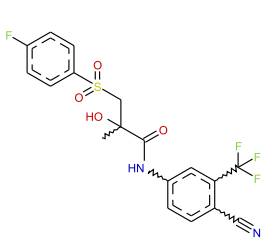
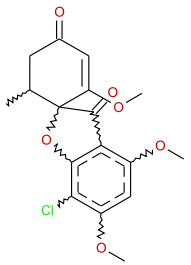
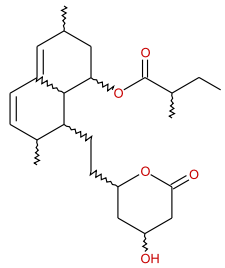
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.148
Enrichment: 0.493
Bayesian Score: -13.7
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.00212

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Bicalutamide	Griseofulvin	Lovastatin
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.807	0.828	0.832
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

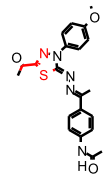
Model Applicability

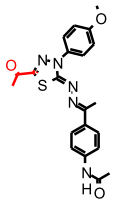
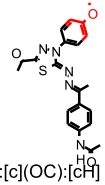
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC15 out of range. Value: -2.6395. Training min, max, SD, explained variance: -2.4461, 3.3002, 1.005, 0.0188.

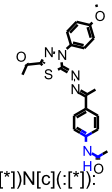
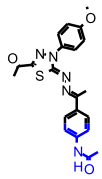
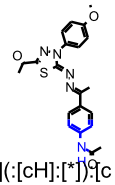
Feature Contribution

Top features for positive contribution

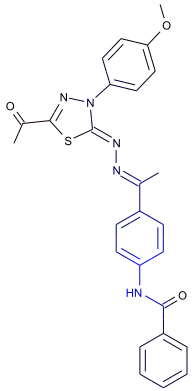
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14

FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.168	3 out of 7
FCFP_12	-1977641857	 [*]:[cH]:[c](OC):[cH]:[*]	0.105	2 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-1.63	0 out of 12
FCFP_12	1175665944	 [*]:[cH]:[c](NC(=O)C):[cH]:[*]	-1.22	0 out of 7
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.998	1 out of 13

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



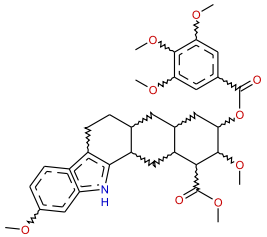
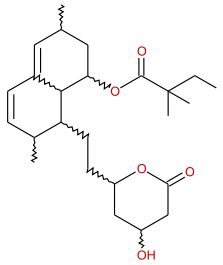
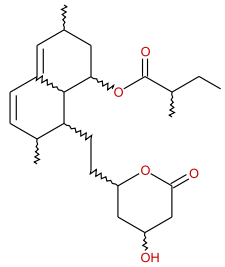
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.146
Enrichment: 0.484
Bayesian Score: -9.06
Mahalanobis Distance: 12.6
Mahalanobis Distance p-value: 0.000192

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Reserpine	Simvastatin	Lovastatin
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.799	0.826	0.834
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

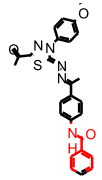
Model Applicability

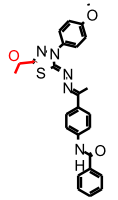
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

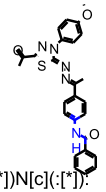
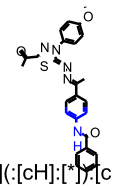
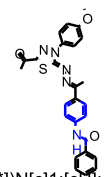
Feature Contribution

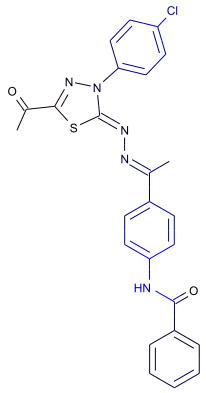
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.239	2 out of 4

FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.168	3 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-1.63	0 out of 12
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.998	1 out of 13
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.859	0 out of 4



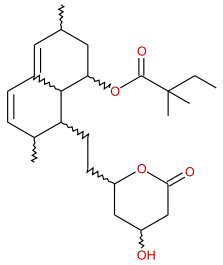
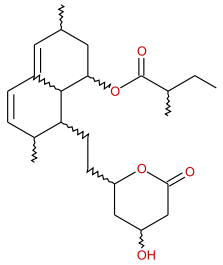
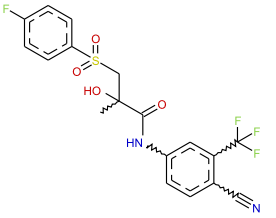
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.156
Enrichment: 0.519
Bayesian Score: -12.8
Mahalanobis Distance: 13.1
Mahalanobis Distance p-value: 7.77e-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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Simvastatin	Lovastatin	Bicalutamide
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.754	0.773	0.807
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

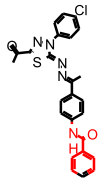
Model Applicability

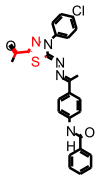
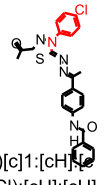
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC2 out of range. Value: 4.6015. Training min, max, SD, explained variance: -5.2888, 4.2744, 2.566, 0.1229.

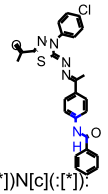
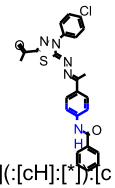
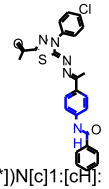
Feature Contribution

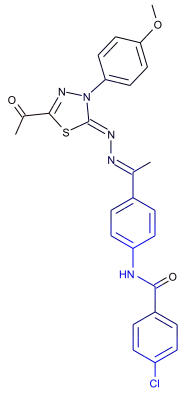
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.239	2 out of 4

FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14
FCFP_12	567484887	 [*]N([*])[c]1:[cH]:[c H]:[c](Cl):[cH]:[cH] :1	0.174	1 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-1.63	0 out of 12
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.859	0 out of 4



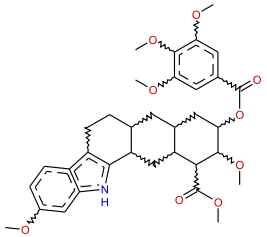
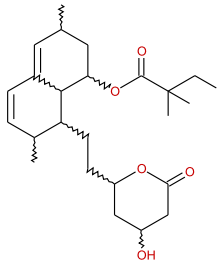
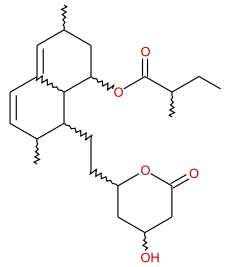
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.156
Enrichment: 0.518
Bayesian Score: -12.9
Mahalanobis Distance: 13.2
Mahalanobis Distance p-value: 6.49e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Reserpine	Simvastatin	Lovastatin
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.752	0.855	0.872
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

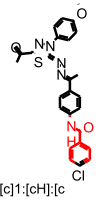
Model Applicability

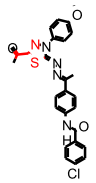
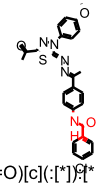
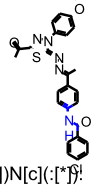
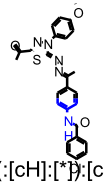
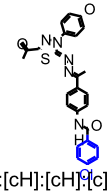
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

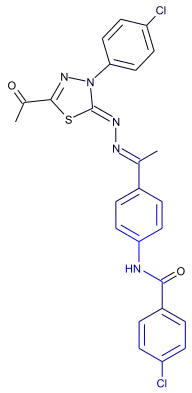
1. OPS PC5 out of range. Value: 3.8956. Training min, max, SD, explained variance: -3.5268, 3.8048, 1.733, 0.0560.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.239	2 out of 4

FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.168	3 out of 7
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Multiple-Carcinogen in training set**
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-1.63	0 out of 12
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.998	1 out of 13
FCFP_12	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	-0.859	0 out of 4



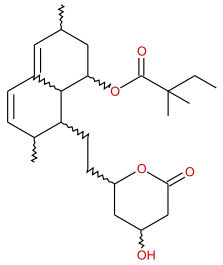
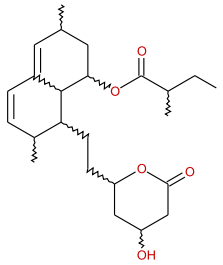
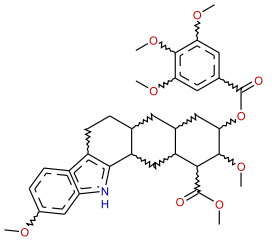
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.155
Enrichment: 0.514
Bayesian Score: -12.4
Mahalanobis Distance: 13.1
Mahalanobis Distance p-value: 8.92e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Simvastatin	Lovastatin	Reserpine
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.802	0.828	0.851
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

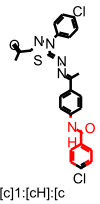
Model Applicability

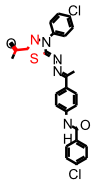
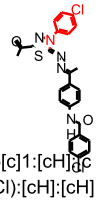
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

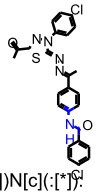
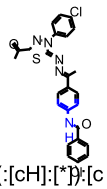
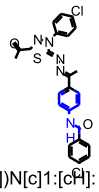
Feature Contribution

Top features for positive contribution

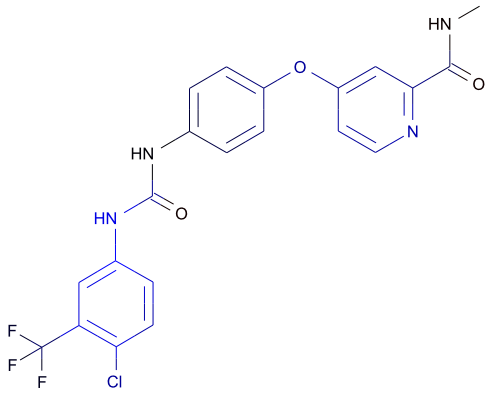
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.239	2 out of 4

FCFP_12	565998553		0.194	6 out of 14
FCFP_12	567484887		0.174	1 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210		-1.63	0 out of 12
FCFP_12	590925877		-0.998	1 out of 13
FCFP_12	-1838187238		-0.859	0 out of 4

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.139

Enrichment: 0.461

Bayesian Score: -14.7

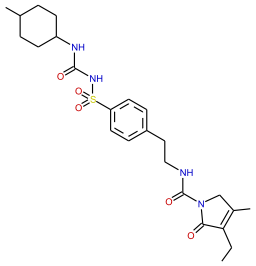
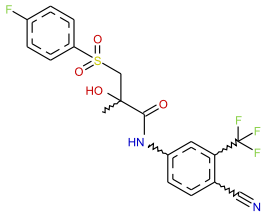
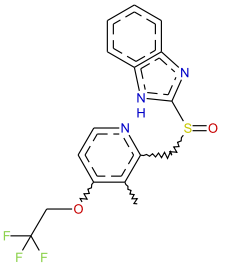
Mahalanobis Distance: 21.3

Mahalanobis Distance p-value: 4.93e-011

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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.

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

Structural Similar Compounds

Name	Glimepride	Bicalutamide	Lansoprazole
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.626	0.700	0.866
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

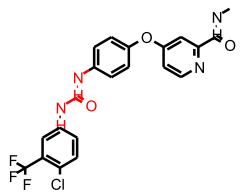
Model Applicability

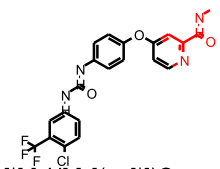
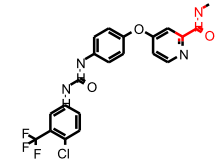
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

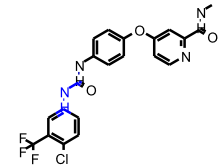
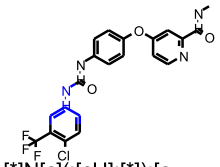
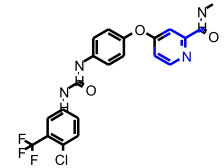
Feature Contribution

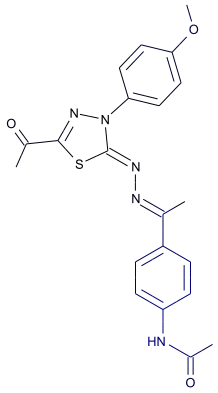
Top features for positive contribution

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

FCFP_12	-904785030	 [*]:[cH]:[c](:n:[*])C (=O)NC	0.174	1 out of 2
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.168	3 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-1.63	0 out of 12
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	-1462709112	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:n:1	-0.994	0 out of 5



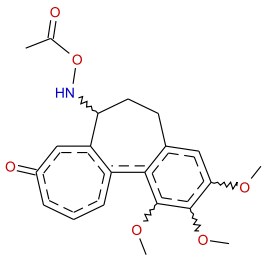
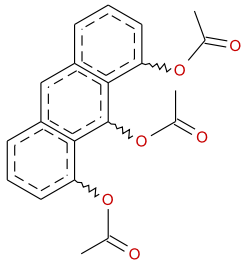
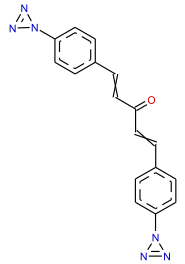
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.6
Enrichment: 0.871
Bayesian Score: -5.01
Mahalanobis Distance: 10.5
Mahalanobis Distance p-value: 0.0274

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild
Distance	0.622	0.783	0.801
Reference	AJOPAA 31;837;48	BJOPAL 53;819;69	28ZPAK-;123;72

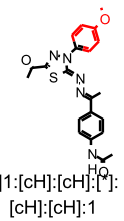
Model Applicability

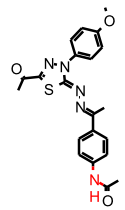
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

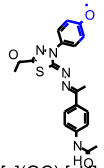
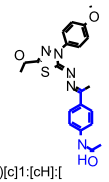
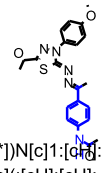
1. All properties and OPS components are within expected ranges.

Feature Contribution

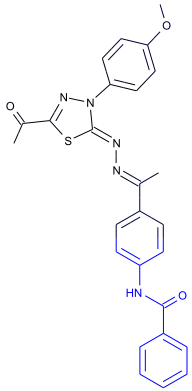
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	346218766	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.197	30 out of 37

FCFP_10	3	 [*]N[*]	0.165	383 out of 491

Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Moderate_Severe in training set**
FCFP_10	-1977641857	 [*]:[cH]:[c](OC):[cH] :[*]	-0.78	4 out of 15
FCFP_10	-178394671	 [*]C(=[*])[c]1:[cH]:[cH]:[c](NC(=O)C):[cH]:[cH]:1	-0.507	0 out of 1
FCFP_10	-790336137	 [*]C(=[*])N[c]1:[cH]: [cH]:[c](:[cH]:[cH]: 1)C(=[*])[*]	-0.507	0 out of 1

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



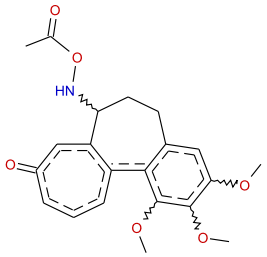
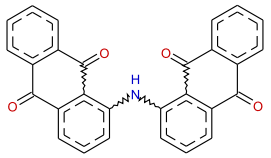
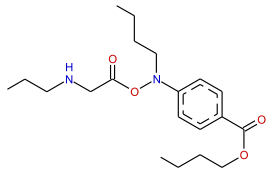
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.0156
Enrichment: 0.0227
Bayesian Score: -13.9
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.00761

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	ANTHRAQUINONE; 1;1'-IMINODI-	Benzoic acid; p-(N-butyl-2-(butylamino)acetamido)-; butyl ester;
Structure			
Actual Endpoint	Moderate_Severe	Mild	Moderate_Severe
Predicted Endpoint	Moderate_Severe	Mild	Moderate_Severe
Distance	0.779	0.791	0.833
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	Arzneimittel-Forschung 8;609;58

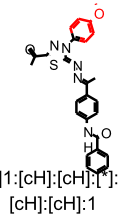
Model Applicability

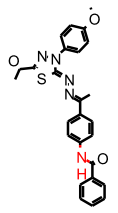
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

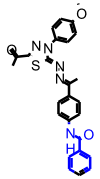
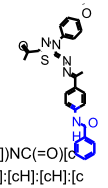
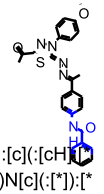
1. All properties and OPS components are within expected ranges.

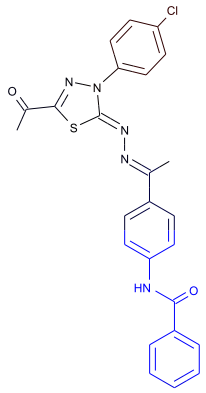
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	346218766	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.197	30 out of 37

FCFP_10	3	 [*]N[*]	0.165	383 out of 491

Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Moderate_Severe in training set**
FCFP_10	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 4
FCFP_10	384221478	 [*]:[c](:[*])NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.29	0 out of 4
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4



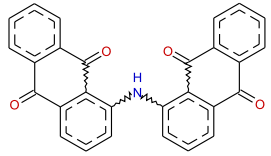
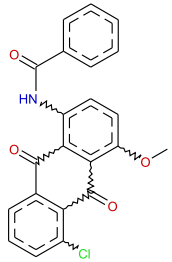
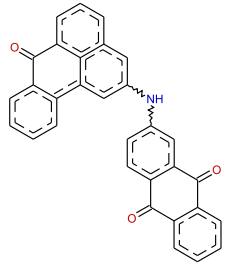
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.0614
Enrichment: 0.0892
Bayesian Score: -11.7
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.00264

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	1-BENZOYLAMINO-4-METHOXY-5-CHLORANTHRAQUINONE	2-(1'-ANTHRAQUINONYL)-AMINOBENZANTHRONE
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.697	0.762	0.807
Reference	28ZPAK-;125;72	28ZPAK-;90;72	28ZPAK-;126;72

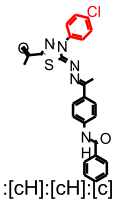
Model Applicability

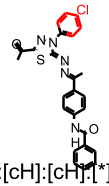
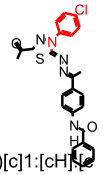
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

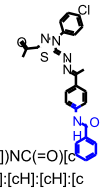
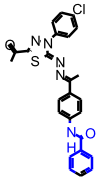
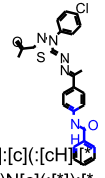
Feature Contribution

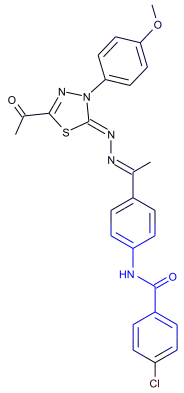
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.329	16 out of 17

FCFP_10	-745491832	 Cl[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.304	29 out of 32
FCFP_10	567484887	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.273	9 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	384221478	 [*]:[c](:[*])NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.29	0 out of 4
FCFP_10	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 4
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4



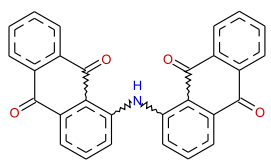
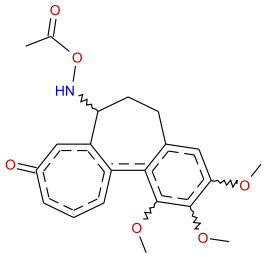
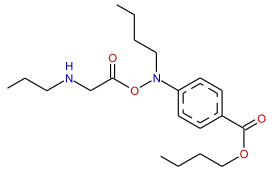
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.169
Enrichment: 0.245
Bayesian Score: -9.72
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 0.000456

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	COLCHICINE	Benzoic acid; p-(N-butyl-2-(butylamino)acetamido)-; butyl ester;
Structure			
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe
Distance	0.815	0.864	0.876
Reference	28ZPAK-;125;72	AJOPAA 31;837;48	Arzneimittel-Forschung 8;609;58

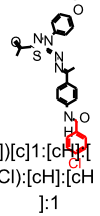
Model Applicability

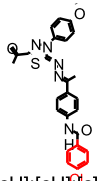
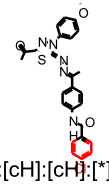
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

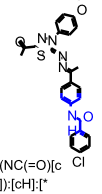
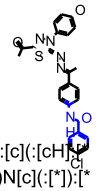
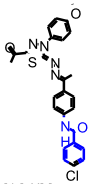
Feature Contribution

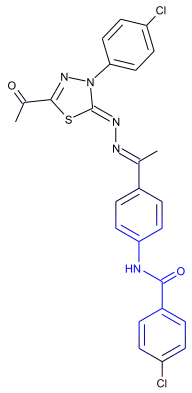
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-149636017	 [*]C(=[*])[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.352	7 out of 7

FCFP_10	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.329	16 out of 17
FCFP_10	-745491832	 Cl[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.304	29 out of 32

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-1.29	0 out of 4
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4
FCFP_10	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 4



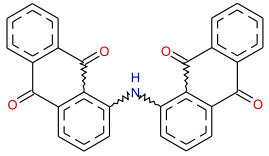
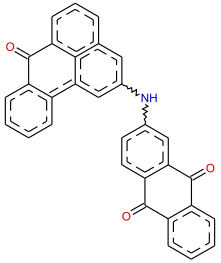
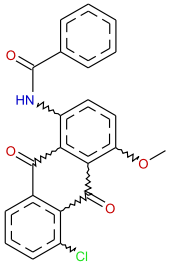
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.297
Enrichment: 0.431
Bayesian Score: -8.23
Mahalanobis Distance: 11.4
Mahalanobis Distance p-value: 0.000955

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	2-(1'-ANTHRAQUINONYL)-AMINOBENZANTHRONE	1-BENZOYLAMINO-4-METHOXY-5-CHLORANTHRAQUINONE
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.745	0.813	0.838
Reference	28ZPAK-;125;72	28ZPAK-;126;72	28ZPAK-;90;72

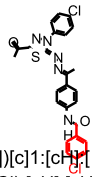
Model Applicability

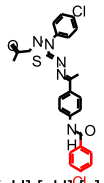
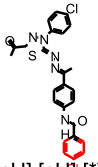
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

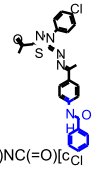
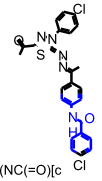
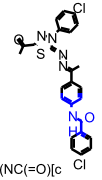
Feature Contribution

Top features for positive contribution

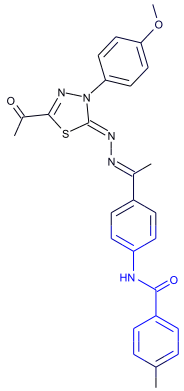
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-149636017	 [*]C(=[*])[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.352	7 out of 7

FCFP_10	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.329	16 out of 17
FCFP_10	-745491832	 Cl[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.304	29 out of 32

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	384221478	 [*]:[c](:[*])NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.29	0 out of 4
FCFP_10	241406177	 [*]:[cH]:[c](NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1):[cH]:[*]	-1.29	0 out of 4
FCFP_10	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-1.29	0 out of 4

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



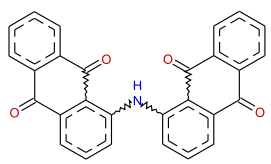
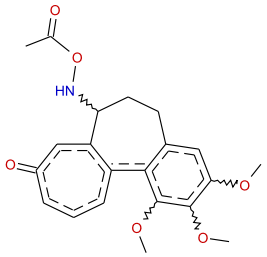
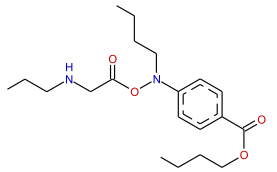
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.0898
Enrichment: 0.13
Bayesian Score: -11
Mahalanobis Distance: 11
Mahalanobis Distance p-value: 0.00424

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	COLCHICINE	Benzoic acid; p-(N-butyl-2-(butylamino)acetamido)-; butyl ester;
Structure			
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe
Distance	0.800	0.829	0.850
Reference	28ZPAK-;125;72	AJOPAA 31;837;48	Arzneimittel-Forschung 8;609;58

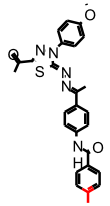
Model Applicability

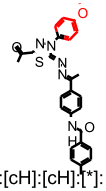
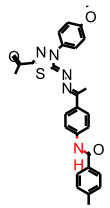
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

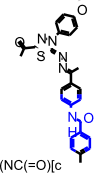
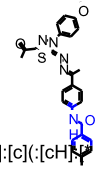
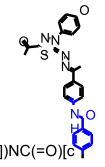
Feature Contribution

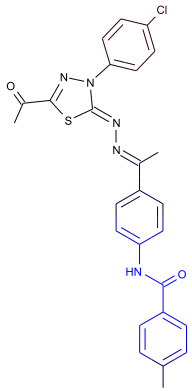
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	136120670	 [*]:[c](:[*])C	0.206	53 out of 65

FCFP_10	346218766	 CO[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.197	30 out of 37
FCFP_10	3	 [*]N[*]	0.165	383 out of 491

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	241406177	 [*]:[cH]:[c](NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1):[cH]:[*]	-1.29	0 out of 4
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4
FCFP_10	384221478	 [*]:[c](:[*])NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.29	0 out of 4



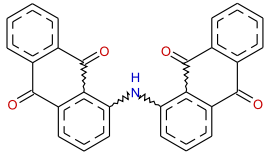
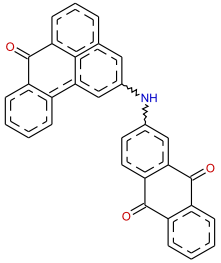
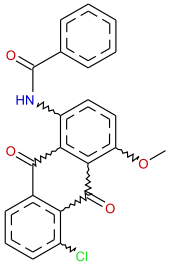
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.241
Enrichment: 0.35
Bayesian Score: -8.83
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.00155

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	2-(1'-ANTHRAQUINONYL)-AMINOBENZANTHRONE	1-BENZOYLAMINO-4-METHOXY-5-CHLORANTHRAQUINONE
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.723	0.804	0.813
Reference	28ZPAK-;125;72	28ZPAK-;126;72	28ZPAK-;90;72

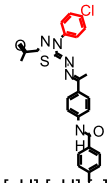
Model Applicability

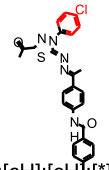
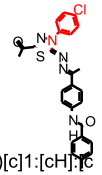
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

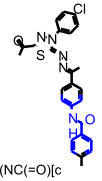
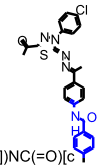
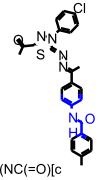
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.329	16 out of 17

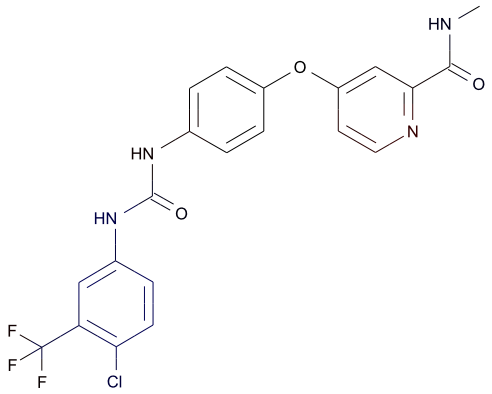
FCFP_10	-745491832	 Cl[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.304	29 out of 32
FCFP_10	567484887	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.273	9 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	241406177	 [*]:[cH]:[c](NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1):[cH]:[*]	-1.29	0 out of 4
FCFP_10	384221478	 [*]:[c](:[*])NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-1.29	0 out of 4
FCFP_10	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-1.29	0 out of 4

Sorafenib

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.776

Enrichment: 1.13

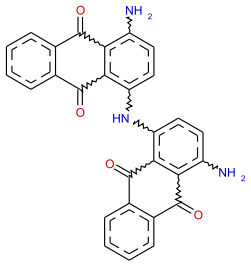
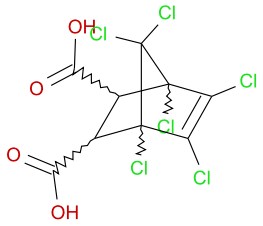
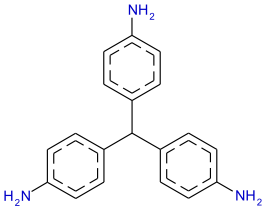
Bayesian Score: -1.8

Mahalanobis Distance: 8.95

Mahalanobis Distance p-value: 0.537

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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;4'-DIAMINO-1;1'-DIANTHRIMIDE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	METHANE;TRIS(4-AMINOPHENYL)-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe
Distance	0.799	0.816	0.827
Reference	28ZPAK-;125;72	28ZPAK-;92;72	28ZPAK-;73;72

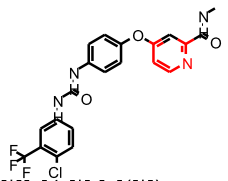
Model Applicability

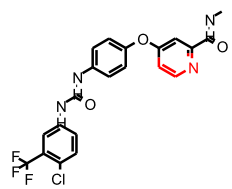
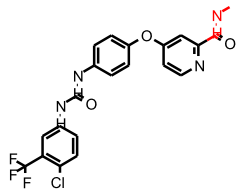
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

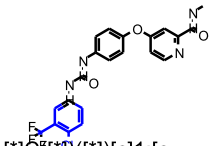
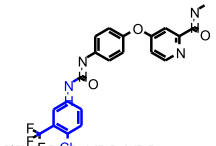
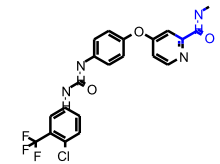
Feature Contribution

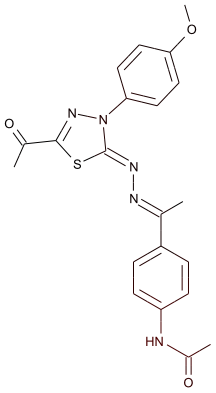
Top features for positive contribution

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

FCFP_10	-124655670	 [*]:[cH]:[cH]:n:[*]	0.259	14 out of 16
FCFP_10	-885550502	 [*]C(=[*])NC	0.239	54 out of 64

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	2104062943	 [*]C([*])([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	-0.745	7 out of 24
FCFP_10	-174293376	 [*]N[c]1:[cH]:[cH]:[c](Cl):[c](:[cH]:1)C([*])([*])[*]	-0.507	0 out of 1
FCFP_10	-1549103449	 [*]NC(=O)[c](:[*]):[*]	-0.504	2 out of 6



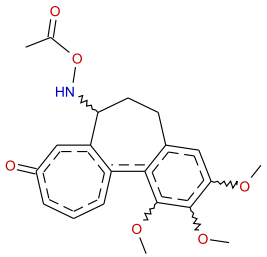
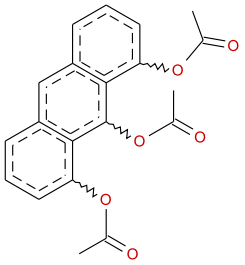
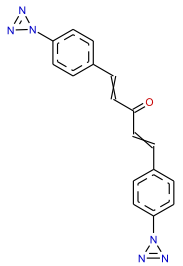
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 2.56
Mahalanobis Distance: 8.04
Mahalanobis Distance p-value: 0.917

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly 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-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.619	0.781	0.792
Reference	AJOPAA 31;837;48	BJOPAL 53;819;69	28ZPAK-;123;72

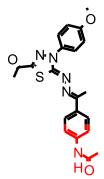
Model Applicability

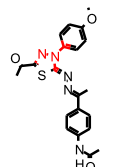
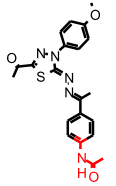
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

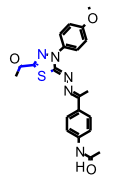
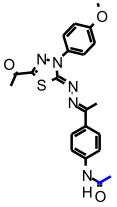
Feature Contribution

Top features for positive contribution

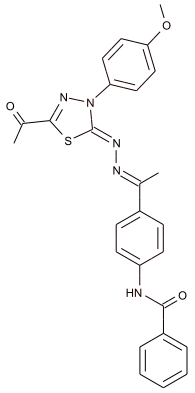
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]:[cH]:[c](NC(=O)C):[cH]:[*]	0.198	14 out of 14

FCFP_12	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.184	7 out of 7
FCFP_12	-1944671191	 [*]:[c](:[*])NC(=O)C	0.184	7 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.0662	198 out of 262
FCFP_12	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382
FCFP_12	136597326	 [*]C(=[*])C	0	612 out of 753

TOPKAT_Ocular_Irritancy_None_vs_Irritant



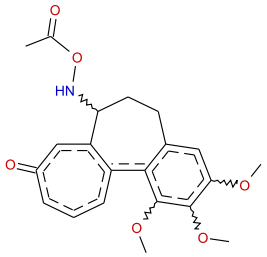
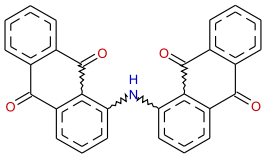
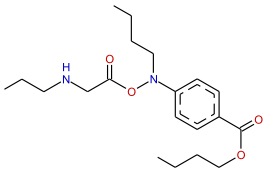
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.48
Mahalanobis Distance: 8.25
Mahalanobis Distance p-value: 0.861

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	ANTHRAQUINONE; 1;1'-IMINODI-	Benzoic acid; p-(N-butyl-2-(butylamino)acetamido)-; butyl ester;
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.753	0.779	0.817
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	Arzneimittel-Forschung 8;609;58

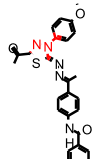
Model Applicability

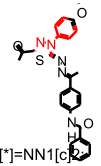
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

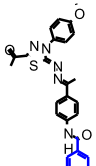
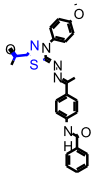
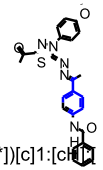
Feature Contribution

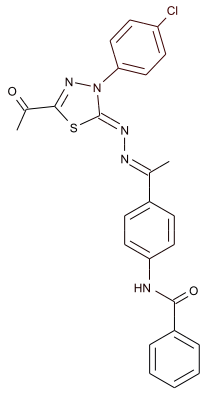
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	675799546	 [*]=C1[*][*]=NN1[C](:[*]):[*]	0.184	7 out of 7

FCFP_12	-319371573		0.18	6 out of 6
FCFP_12	580960234		0.18	6 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694		-0.0964	107 out of 146
FCFP_12	565998553		-0.0662	198 out of 262
FCFP_12	-453677277		0	264 out of 323



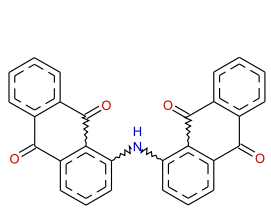
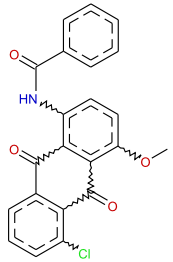
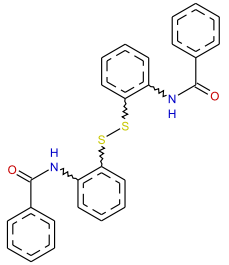
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 2.09
Mahalanobis Distance: 7.96
Mahalanobis Distance p-value: 0.933

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	1-BENZOYLAMINO-4-METHOXY-5-CHLORANTHRAQUINONE	BENZANILIDE;2';2'''-DITHIOBIS-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.690	0.753	0.773
Reference	28ZPAK-;125;72	28ZPAK-;90;72	28ZPAK-;173;72

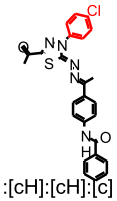
Model Applicability

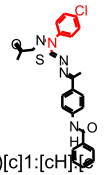
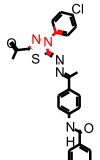
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

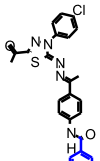
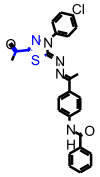
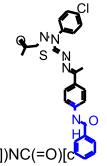
Feature Contribution

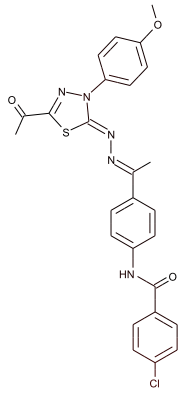
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.2	17 out of 17

FCFP_12	567484887	 [*]N([*])[c]1:[cH]:[c H]:[c](Cl):[cH]:[cH] :1	0.192	10 out of 10
FCFP_12	675799546	 [*]=C1[*][*]=NN1[c](: [*]):[*]	0.184	7 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.0662	198 out of 262
FCFP_12	384221478	 [*]:[c](:[*])NC(=O)[c]1:[cH]:[cH]:[cH]:[c H]:[cH]:1	0	4 out of 5



$C_{26}H_{22}ClN_5O_3S$

Molecular Weight: 520.00258

ALogP: 5.392

Rotatable Bonds: 7

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

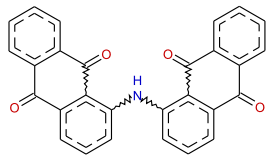
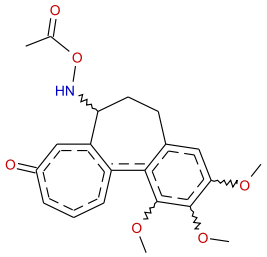
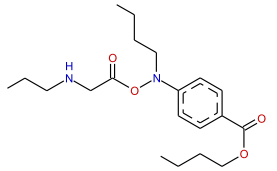
Bayesian Score: 2.43

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

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	COLCHICINE	Benzoic acid; p-(N-butyl-2-(butylamino)acetamido)-; butyl ester;
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.803	0.829	0.856
Reference	28ZPAK-;125;72	AJOPAA 31;837;48	Arzneimittel-Forschung 8;609;58

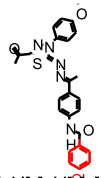
Model Applicability

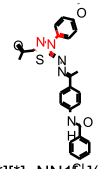
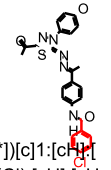
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

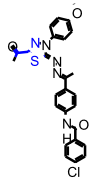
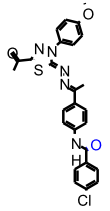
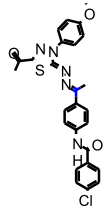
Feature Contribution

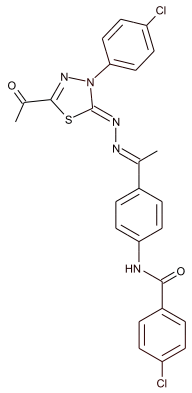
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.2	17 out of 17

FCFP_12	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.184	7 out of 7
FCFP_12	-149636017	 [*]C(=[*])[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.184	7 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.0662	198 out of 262
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	0	 [*]C(=[*])[*]	0	1184 out of 1397



$C_{25}H_{19}Cl_2N_5O_2S$

Molecular Weight: 524.42166

ALogP: 6.073

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

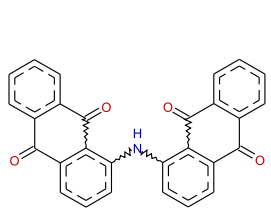
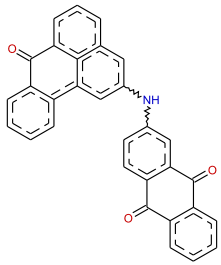
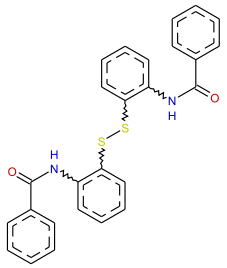
Bayesian Score: 2.37

Mahalanobis Distance: 8.2

Mahalanobis Distance p-value: 0.875

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	2-(1'-ANTHRAQUINONYL)-AMINOBENZANTHRONE	BENZANILIDE;2';2'''-DITHIOBIS-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.733	0.802	0.803
Reference	28ZPAK-;125;72	28ZPAK-;126;72	28ZPAK-;173;72

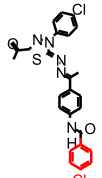
Model Applicability

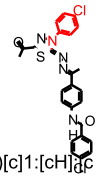
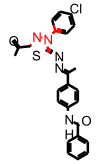
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

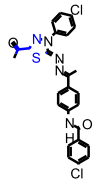
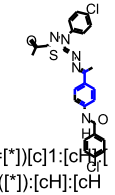
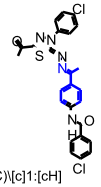
Feature Contribution

Top features for positive contribution

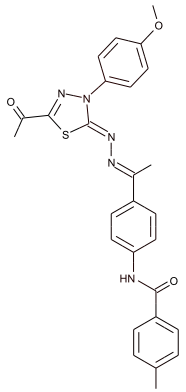
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.2	17 out of 17

FCFP_12	567484887		0.192	10 out of 10
FCFP_12	675799546		0.184	7 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553		-0.0662	198 out of 262
FCFP_12	-453677277		0	264 out of 323
FCFP_12	-581162801		0	7 out of 9

TOPKAT_Ocular_Irritancy_None_vs_Irritant



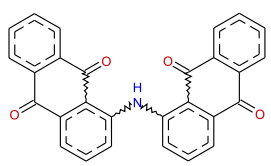
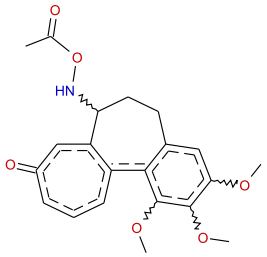
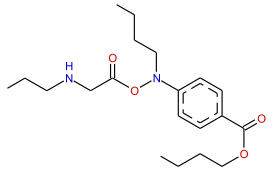
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.54
Mahalanobis Distance: 8.36
Mahalanobis Distance p-value: 0.823

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	COLCHICINE	Benzoic acid; p-(N-butyl-2-(butylamino)acetamido)-; butyl ester;
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.789	0.797	0.831
Reference	28ZPAK-;125;72	AJOPAA 31;837;48	Arzneimittel-Forschung 8;609;58

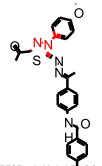
Model Applicability

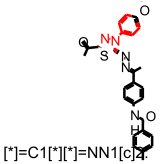
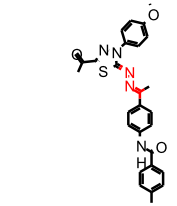
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

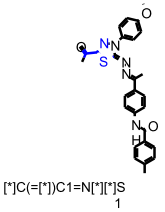
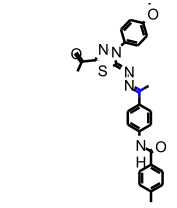
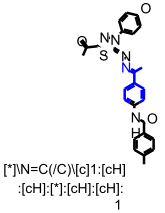
Feature Contribution

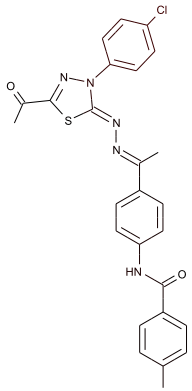
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	675799546	 [*]=C1[*][*]=NN1[c](: [*]):[*]	0.184	7 out of 7

FCFP_12	-319371573		0.18	6 out of 6
FCFP_12	580960234		0.18	6 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553		-0.0662	198 out of 262
FCFP_12	0		0	1184 out of 1397
FCFP_12	-581162801		0	7 out of 9



$C_{26}H_{22}ClN_5O_2S$

Molecular Weight: 504.00318

ALogP: 5.895

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

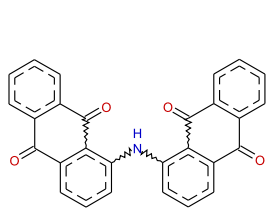
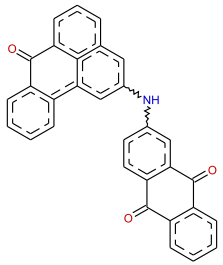
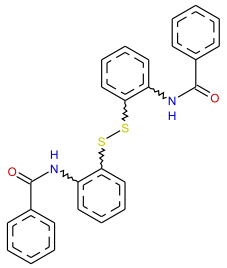
Bayesian Score: 2.52

Mahalanobis Distance: 8.08

Mahalanobis Distance p-value: 0.908

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ANTHRAQUINONE; 1;1'-IMINODI-	2-(1'-ANTHRAQUINONYL)-AMINOBENZANTHRONE	BENZANILIDE;2';2'''-DITHIOBIS-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.712	0.793	0.796
Reference	28ZPAK-;125;72	28ZPAK-;126;72	28ZPAK-;173;72

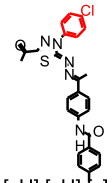
Model Applicability

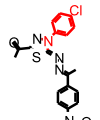
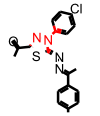
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

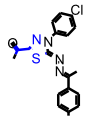
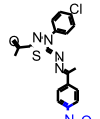
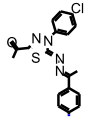
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.2	17 out of 17

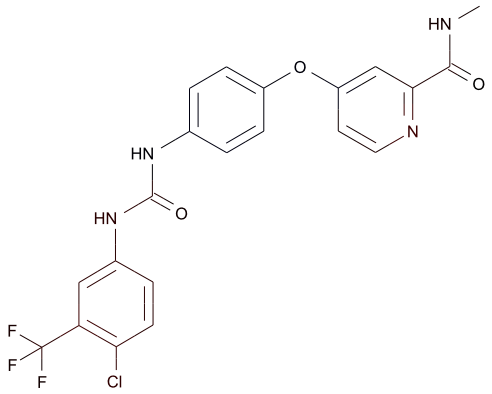
FCFP_12	567484887	 [*]N([*])[c]1:[cH]:[cH]:[c]([Cl]):[cH]:[cH]:1	0.192	10 out of 10
FCFP_12	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.184	7 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.0662	198 out of 262
FCFP_12	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	0	4 out of 5
FCFP_12	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[c]:[*]:[cH]:[cH]:1	0	4 out of 5

Sorafenib

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

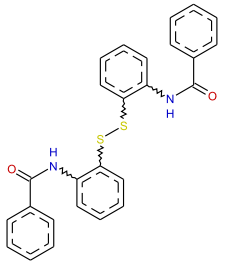
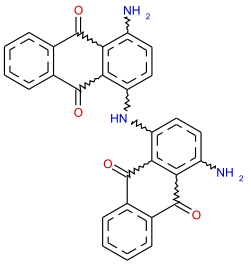
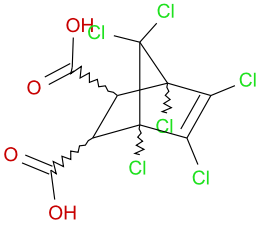
Bayesian Score: 3.04

Mahalanobis Distance: 6.28

Mahalanobis Distance p-value: 1

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	BENZANILIDE;2';2'''-DITHIOBIS-	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Non-Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.743	0.791	0.801
Reference	28ZPAK-;173;72	28ZPAK-;125;72	28ZPAK-;92;72

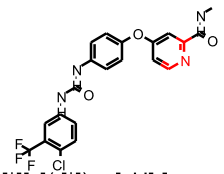
Model Applicability

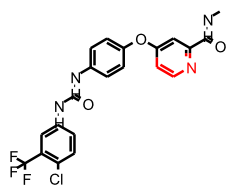
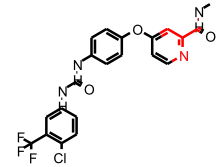
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

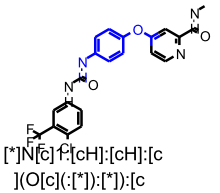
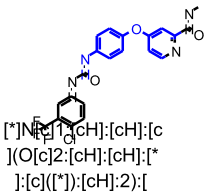
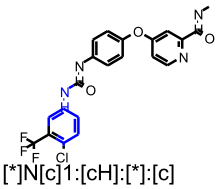
Feature Contribution

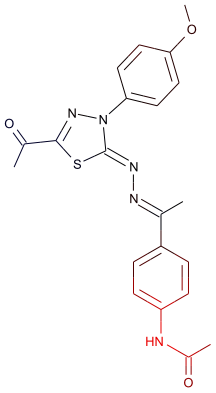
Top features for positive contribution

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

FCFP_12	-124655670	 [*]:[cH]:[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1539132615	 [*]C(=[*])[c](:[cH]:[*]):n:[*]	0.197	13 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-747629521	 [*]N[c]1:[cH]:[cH]:[c](O[c](:[*]):[*]):[cH]:[cH]:1	-0.268	1 out of 2
FCFP_12	702861189	 [*]N[c]1:[cH]:[cH]:[c](O[c]2:[cH]:[cH]:[*]:[c]([*]):[cH]:2):[cH]:[cH]:1	-0.268	1 out of 2
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0	102 out of 121



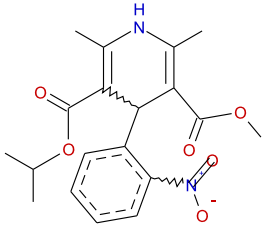
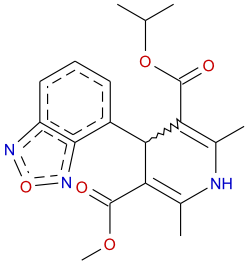
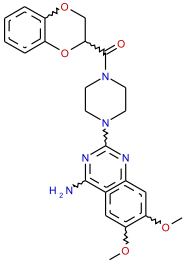
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.404
Enrichment: 1.26
Bayesian Score: 4.24
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.0488

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Nisoldipine	Isradipine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.594	0.618	0.644
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

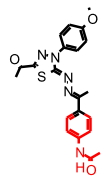
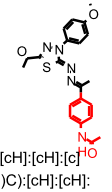
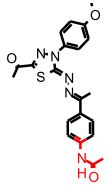
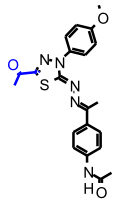
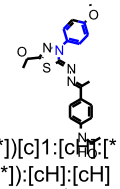
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

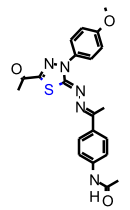
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

Feature Contribution

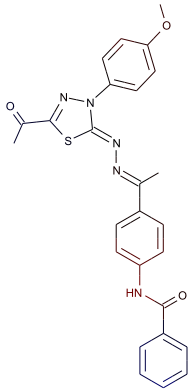
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	1776488	 CC(=O)N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.613	2 out of 2
ECFP_12	-847011520	 [*][c]1:[cH]:[cH]:[c](NC(=O)C):[cH]:[cH]:1	0.613	2 out of 2
ECFP_12	-1923054811	 [*]:[c](:[*])NC(=O)C	0.575	3 out of 4
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4
ECFP_12	-2137232509	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.485	0 out of 2

ECFP_12	912478223	 [*]S[*]	-0.318	2 out of 10

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



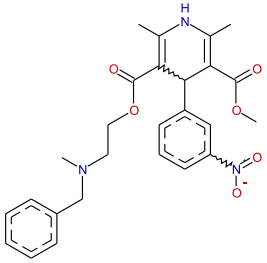
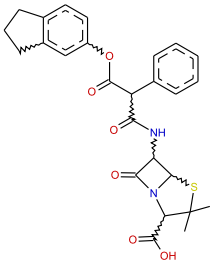
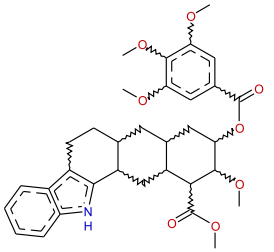
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.289
Enrichment: 0.897
Bayesian Score: -0.487
Mahalanobis Distance: 12.2
Mahalanobis Distance p-value: 0.00281

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Nicardipine	Carbenicillin	Deserpidine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.667	0.669	0.707
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

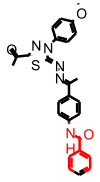
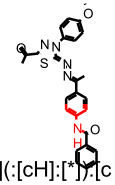
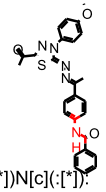
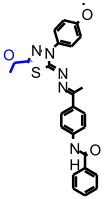
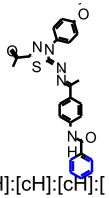
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

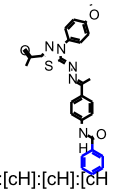
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

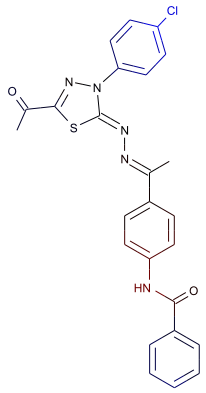
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.613	2 out of 2
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4
ECFP_12	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64

ECFP_12	-281505363	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64



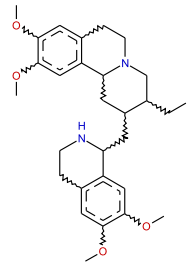
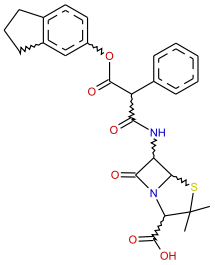
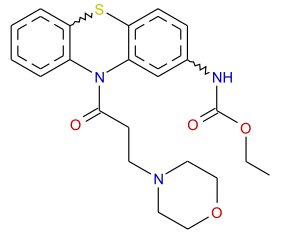
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.218
Enrichment: 0.678
Bayesian Score: -5.27
Mahalanobis Distance: 11.5
Mahalanobis Distance p-value: 0.0198

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Emetine	Carbenicillin	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.698	0.710	0.725
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

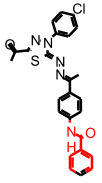
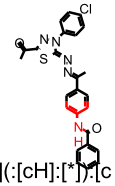
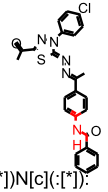
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

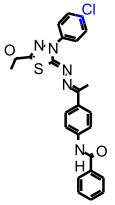
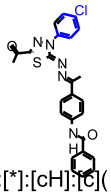
Feature Contribution

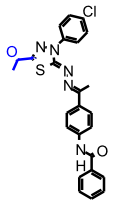
Top features for positive contribution

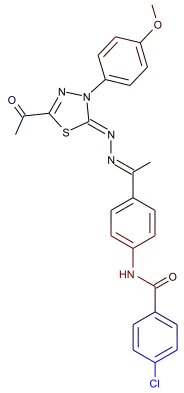
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-223149939		0.613	2 out of 2
ECFP_12	-177077903		0.529	6 out of 10
ECFP_12	-1236483485		0.46	9 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	99947387		-0.817	8 out of 62
ECFP_12	1854732111		-0.816	4 out of 33

ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4



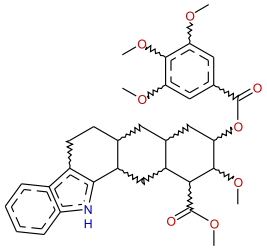
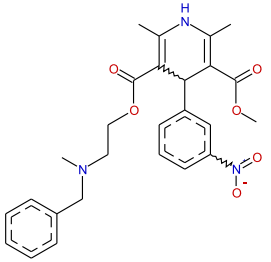
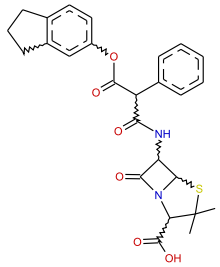
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.25
Enrichment: 0.776
Bayesian Score: -2.79
Mahalanobis Distance: 11.9
Mahalanobis Distance p-value: 0.00555

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Deserpidine	Nicardipine	Carbenicillin
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.653	0.715	0.724
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

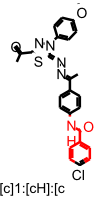
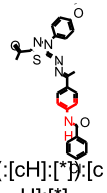
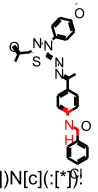
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

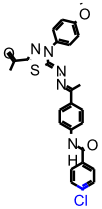
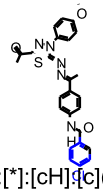
Feature Contribution

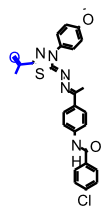
Top features for positive contribution

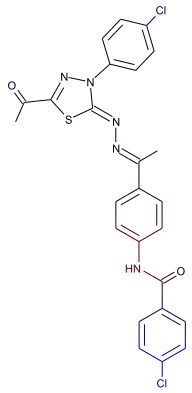
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-223149939		0.613	2 out of 2
ECFP_12	-177077903		0.529	6 out of 10
ECFP_12	-1236483485		0.46	9 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	99947387		-0.817	8 out of 62
ECFP_12	1854732111		-0.816	4 out of 33

ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4



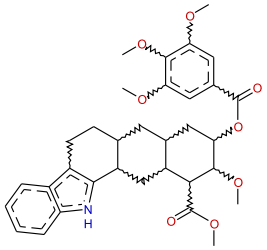
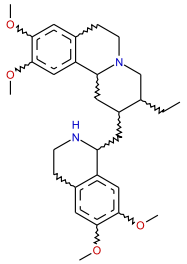
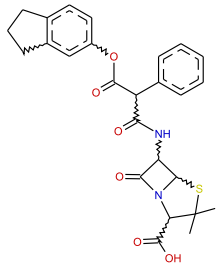
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.246
Enrichment: 0.764
Bayesian Score: -3.05
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.0364

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Deserpidine	Emetine	Carbenicillin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.720	0.723	0.775
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

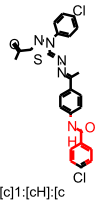
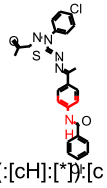
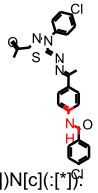
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

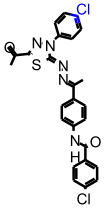
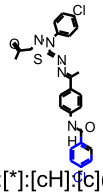
Feature Contribution

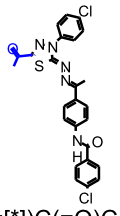
Top features for positive contribution

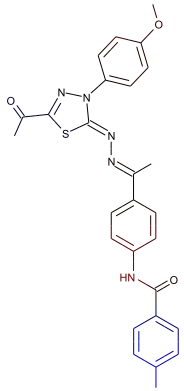
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.613	2 out of 2
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	99947387	 [*]:[c](:[*])Cl	-0.817	8 out of 62
ECFP_12	1854732111	 [*][c]1:[*]:[cH]:[c](Cl):[cH]:[cH]:1	-0.816	4 out of 33

ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4



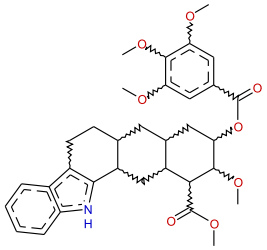
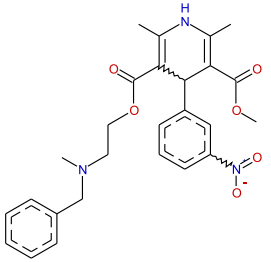
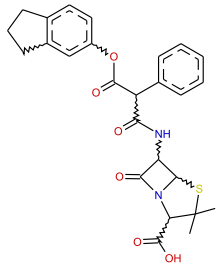
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.262
Enrichment: 0.815
Bayesian Score: -1.98
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 0.0111

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Deserpidine	Nicardipine	Carbenicillin
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.678	0.695	0.709
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

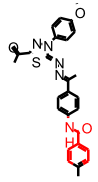
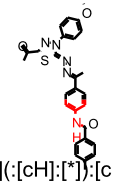
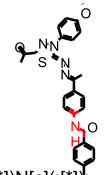
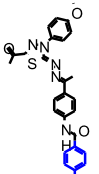
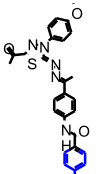
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

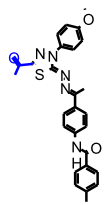
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

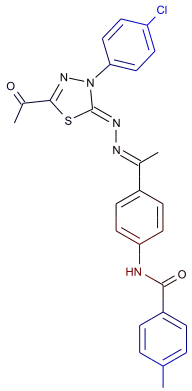
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.613	2 out of 2
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_12	-1926229349	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.06	0 out of 6
ECFP_12	-533780882	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.06	0 out of 6

ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4



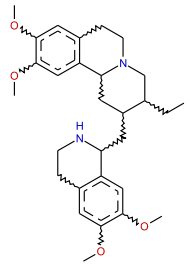
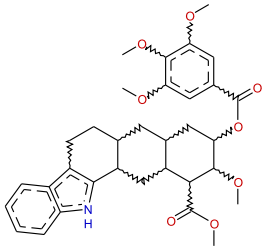
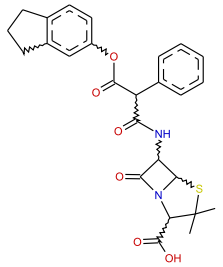
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.208
Enrichment: 0.646
Bayesian Score: -6.27
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.0321

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Emetine	Deserpidine	Carbenicillin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.707	0.734	0.758
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

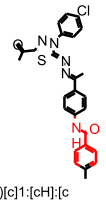
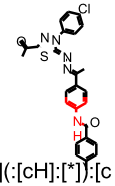
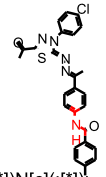
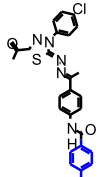
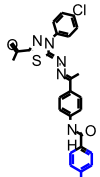
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

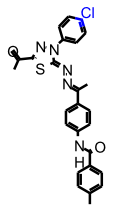
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

Feature Contribution

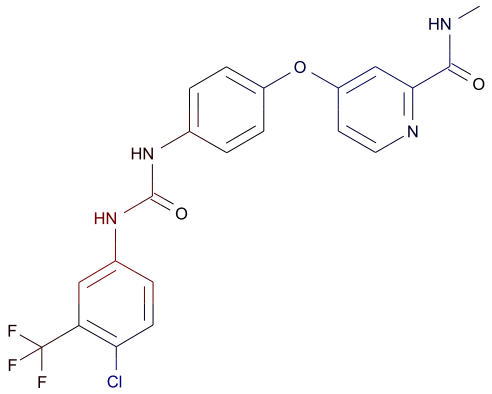
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-223149939	 [*]NC(=O)[c]1:[cH]:[cH]:[c]H]:[*]:[cH]:[cH]:1	0.613	2 out of 2
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_12	-1926229349	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.06	0 out of 6
ECFP_12	-533780882	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.06	0 out of 6

ECFP_12	99947387	 [*]:[c](:[*])Cl	-0.817	8 out of 62

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.236

Enrichment: 0.734

Bayesian Score: -3.76

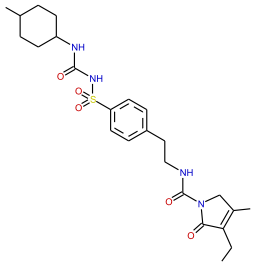
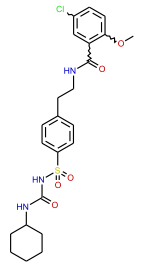
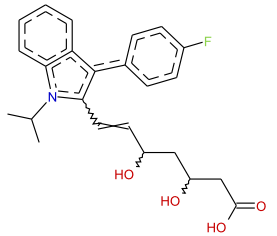
Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.00229

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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.

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

Structural Similar Compounds

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

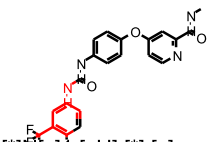
Model Applicability

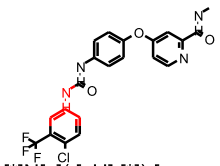
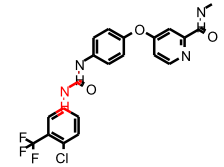
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

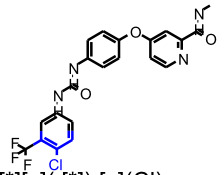
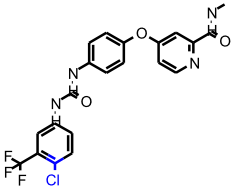
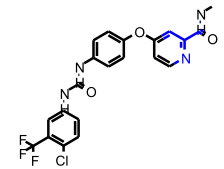
Feature Contribution

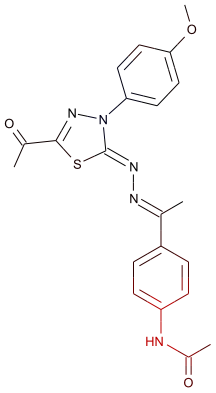
Top features for positive contribution

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

ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-1.11	2 out of 26
ECFP_12	99947387	 [*]:[c](:[*])Cl	-0.817	8 out of 62
ECFP_12	1413420509	 [*]C(=[*])[c](:[cH]:[*]):n:[*]	-0.661	0 out of 3



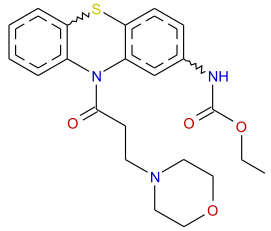
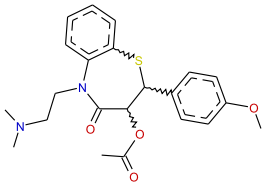
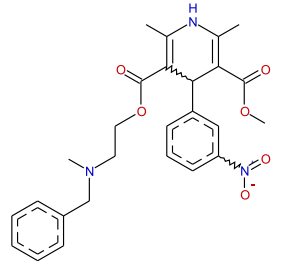
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.692
Enrichment: 1.85
Bayesian Score: 5.3
Mahalanobis Distance: 15.8
Mahalanobis Distance p-value: 4.69e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Moricizine	Diltiazem	Nicardipine
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.620	0.671	0.702
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

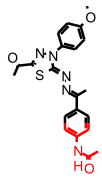
Model Applicability

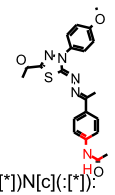
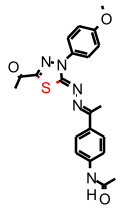
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC15 out of range. Value: 2.8947. Training min, max, SD, explained variance: -2.9572, 2.6953, 1.089, 0.0176.

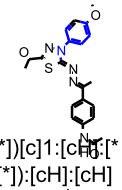
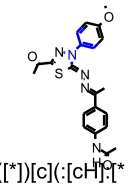
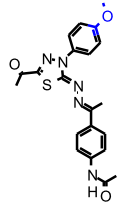
Feature Contribution

Top features for positive contribution

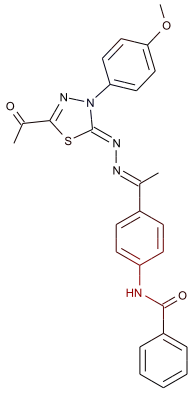
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	2097618059	 [*]:[cH]:[c](NC(=O)C):[cH]:[*]	0.73	5 out of 6

SCFP_4	1631845520	 [*]C(=[*])N[c](:[*]):[*]	0.601	6 out of 9
SCFP_4	17	 [*]S[*]	0.548	10 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.413	3 out of 16
SCFP_4	1334669481	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.24	4 out of 17
SCFP_4	276283342	 [*]:[c](:[*])OC	-0.106	5 out of 18

TOPKAT_Rat_Female_FDA_Single_vs_Multiple



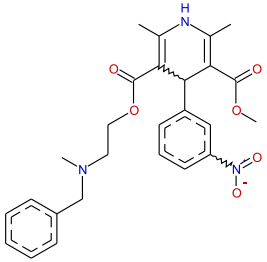
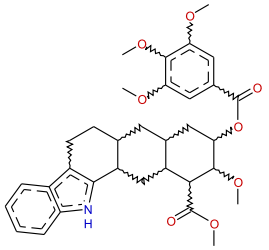
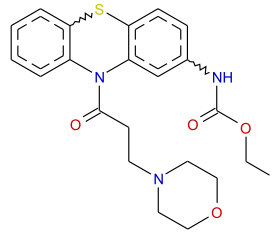
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.635
Enrichment: 1.7
Bayesian Score: 3.98
Mahalanobis Distance: 18.9
Mahalanobis Distance p-value: 5.91e-010

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Nicardipine	Deserpidine	Moricizine
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.612	0.656	0.689
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

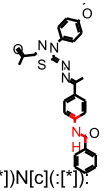
Model Applicability

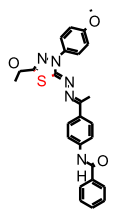
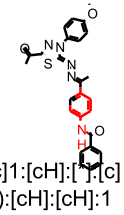
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

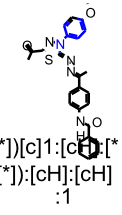
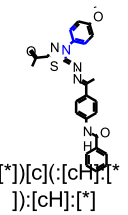
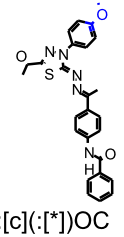
Feature Contribution

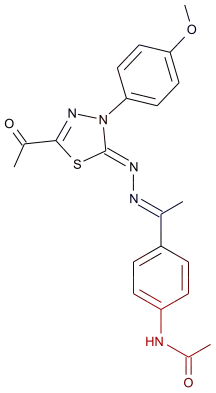
Top features for positive contribution

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

SCFP_4	17	 [*]S[*]	0.548	10 out of 17
SCFP_4	-1375926917	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.522	6 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.413	3 out of 16
SCFP_4	1334669481	 [*]N([*])[c](:[cH]):[*]):[cH]:[*]	-0.24	4 out of 17
SCFP_4	276283342	 [*]:[c](:[*])OC	-0.106	5 out of 18



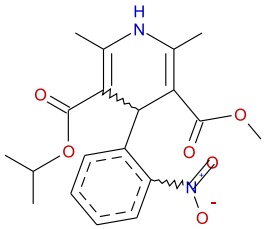
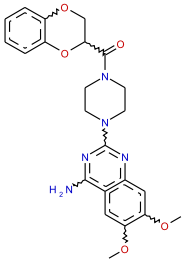
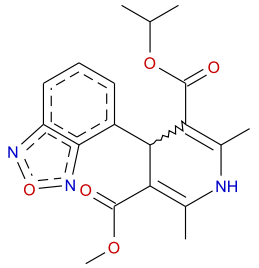
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.456
Enrichment: 1.36
Bayesian Score: 3.11
Mahalanobis Distance: 14.2
Mahalanobis Distance p-value: 2.25e-005

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Nisoldipine	Doxazosin	Isradipine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.589	0.608	0.612
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

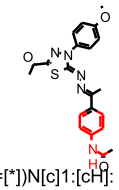
Model Applicability

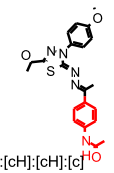
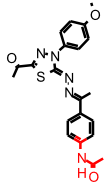
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

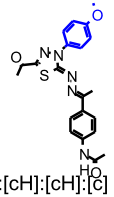
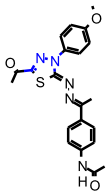
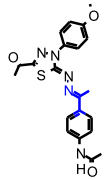
Feature Contribution

Top features for positive contribution

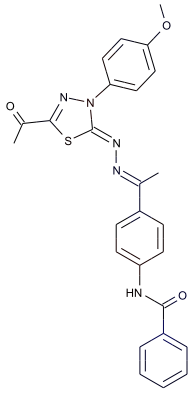
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.615	5 out of 7

SCFP_6	814408713	 [*][c]1:[cH]:[cH]:[c](NC(=O)C):[cH]:[cH]:1	0.603	2 out of 2
SCFP_6	1626825020	 [*]:[c](:[*])NC(=O)C	0.561	3 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1287669168	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.38	1 out of 6
SCFP_6	-1325991669	 [*]N1[*][*]C(=N1)[*]	-0.278	0 out of 1
SCFP_6	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.278	0 out of 1

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



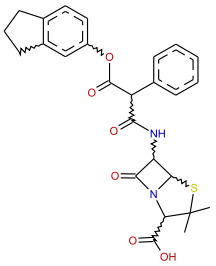
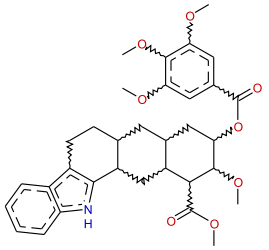
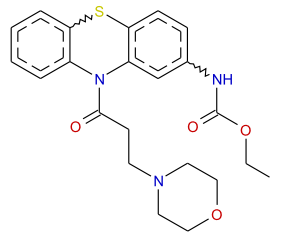
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.322
Enrichment: 0.962
Bayesian Score: -1.27
Mahalanobis Distance: 16.4
Mahalanobis Distance p-value: 8.15e-010

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Carbenicillin	Deserpidine	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.654	0.688	0.707
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

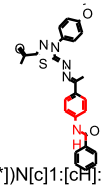
Model Applicability

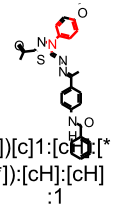
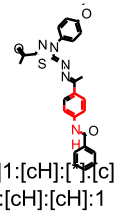
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

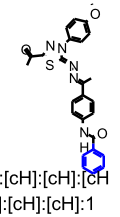
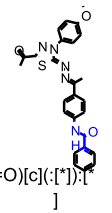
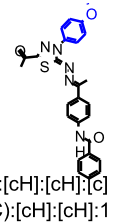
Feature Contribution

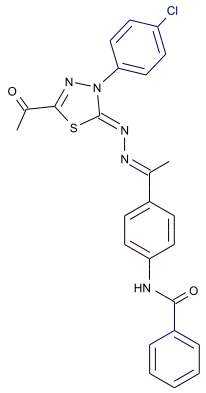
Top features for positive contribution

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

SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[cH]:[c]([*])[c]([*]):[cH]:[cH]:1	0.287	17 out of 39
SCFP_6	-1375926917	 [*]N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	0.251	11 out of 26

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.504	12 out of 64
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	-0.436	4 out of 21
SCFP_6	1287669168	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.38	1 out of 6



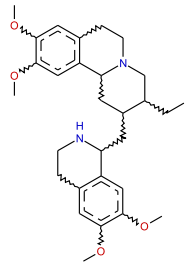
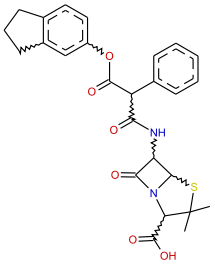
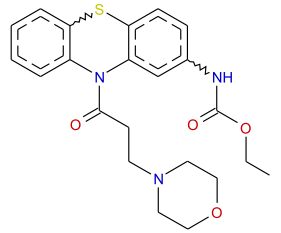
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.286
Enrichment: 0.857
Bayesian Score: -2.7
Mahalanobis Distance: 15.2
Mahalanobis Distance p-value: 2.64e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Emetine	Carbenicillin	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.694	0.701	0.708
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

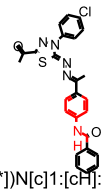
Model Applicability

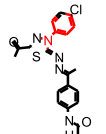
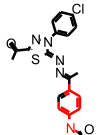
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC1 out of range. Value: -5.7239. Training min, max, SD, explained variance: -5.694, 8.348, 2.949, 0.0811.

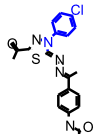
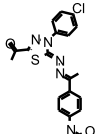
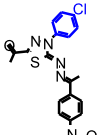
Feature Contribution

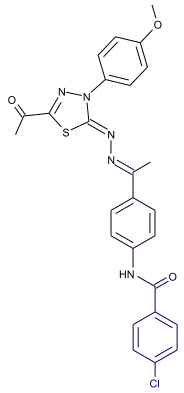
Top features for positive contribution

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

SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH] :1	0.287	17 out of 39
SCFP_6	-1375926917	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.251	11 out of 26

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	341480432	 [*]N([*])[c]1:[cH]:[c H]:[c](Cl):[cH]:[cH] :1	-0.674	0 out of 3
SCFP_6	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.504	12 out of 64
SCFP_6	1905487031	 [*][c]1:[cH]:[cH]:[c] (Cl):[cH]:[cH]:1	-0.48	2 out of 12



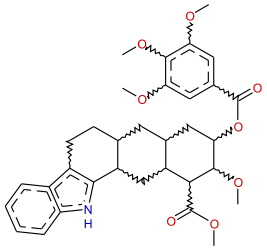
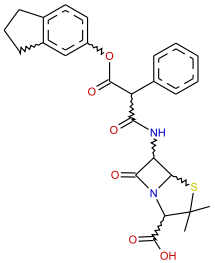
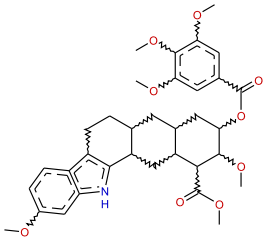
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.277
Enrichment: 0.829
Bayesian Score: -3.11
Mahalanobis Distance: 15.8
Mahalanobis Distance p-value: 1.38e-008

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Deserpidine	Carbenicillin	Reserpine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.634	0.709	0.716
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

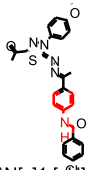
Model Applicability

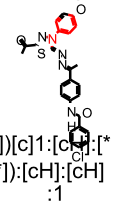
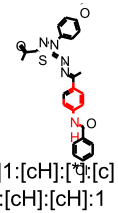
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

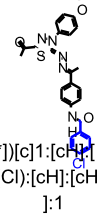
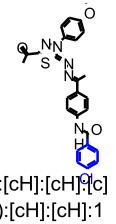
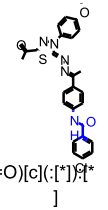
Feature Contribution

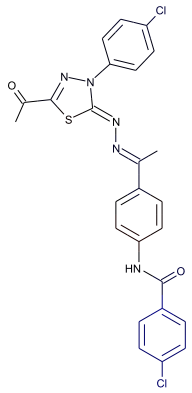
Top features for positive contribution

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

SCFP_6	-1380909229		0.287	17 out of 39
SCFP_6	-1375926917		0.251	11 out of 26

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1915307678		-0.496	0 out of 2
SCFP_6	1905487031		-0.48	2 out of 12
SCFP_6	1257084377		-0.436	4 out of 21



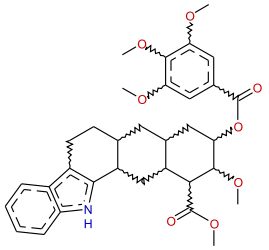
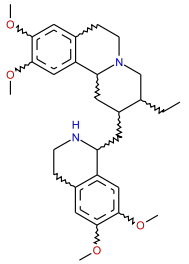
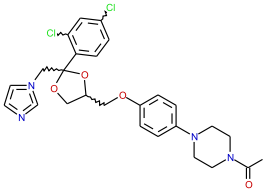
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.291
Enrichment: 0.869
Bayesian Score: -2.52
Mahalanobis Distance: 15.3
Mahalanobis Distance p-value: 1.48e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Deserpidine	Emetine	Ketoconazole
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.711	0.718	0.754
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

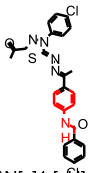
Model Applicability

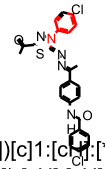
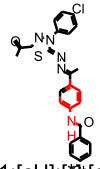
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

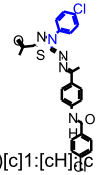
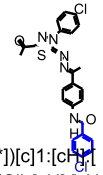
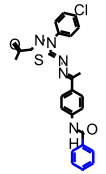
Feature Contribution

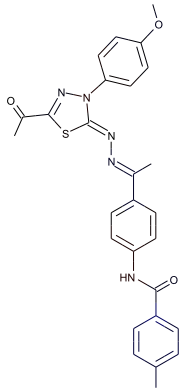
Top features for positive contribution

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

SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.287	17 out of 39
SCFP_6	-1375926917	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.251	11 out of 26

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	341480432	 [*]N([*])[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	-0.674	0 out of 3
SCFP_6	1915307678	 [*]C(=[*])[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	-0.496	0 out of 2
SCFP_6	1905487031	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	-0.48	2 out of 12



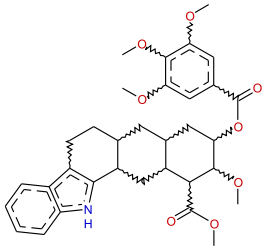
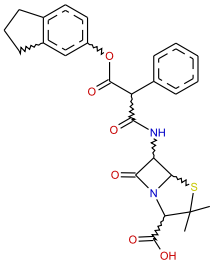
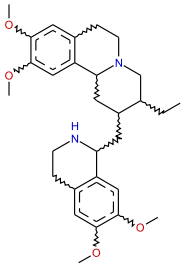
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.319
Enrichment: 0.953
Bayesian Score: -1.38
Mahalanobis Distance: 16.2
Mahalanobis Distance p-value: 1.65e-009

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Deserpidine	Carbenicillin	Emetine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.655	0.685	0.711
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

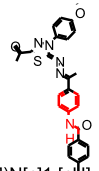
Model Applicability

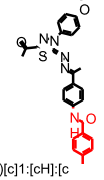
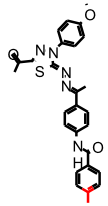
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

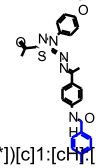
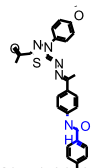
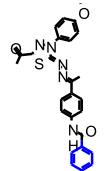
Feature Contribution

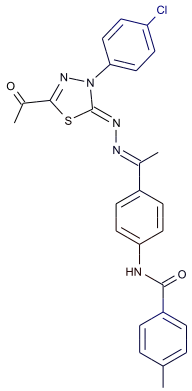
Top features for positive contribution

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

SCFP_6	-1555568408		0.415	1 out of 1
SCFP_6	136686699		0.287	17 out of 39

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	2048398673		-0.674	0 out of 3
SCFP_6	1257084377		-0.436	4 out of 21
SCFP_6	795925860		-0.38	1 out of 6



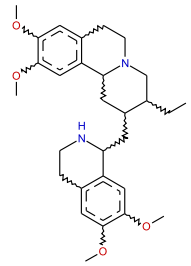
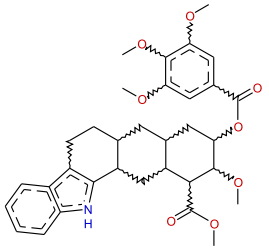
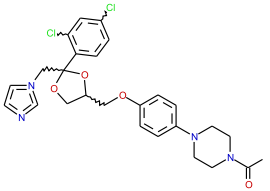
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.289
Enrichment: 0.865
Bayesian Score: -2.58
Mahalanobis Distance: 15.6
Mahalanobis Distance p-value: 3.38e-008

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Emetine	Deserpidine	Ketoconazole
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.697	0.722	0.740
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

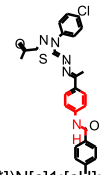
Model Applicability

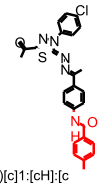
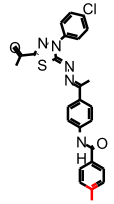
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

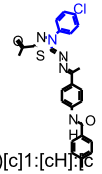
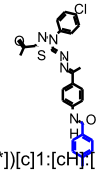
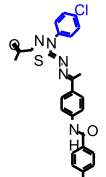
Feature Contribution

Top features for positive contribution

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

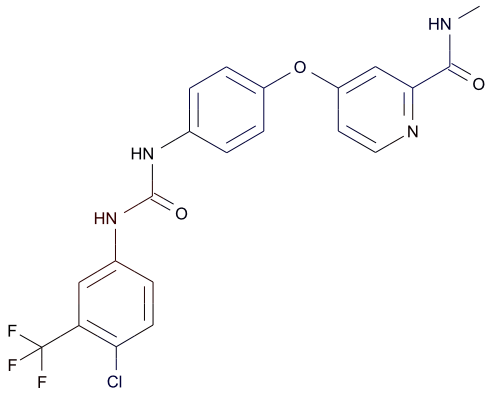
SCFP_6	-1555568408	 [*]NC(=O)[c]1:[cH]:[cH]:[c]:[c](C):[cH]:[cH]:1	0.415	1 out of 1
SCFP_6	136686699	 [*]:[c](:[*])C	0.287	17 out of 39

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	341480432	 [*]N([*])[c]1:[cH]:[cH]:[c]:[c](Cl):[cH]:[cH]:1	-0.674	0 out of 3
SCFP_6	2048398673	 [*]C(=[*])[c]1:[cH]:[cH]:[c]:[c](C):[cH]:[cH]:1	-0.674	0 out of 3
SCFP_6	1905487031	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	-0.48	2 out of 12

Sorafenib

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.293

Enrichment: 0.878

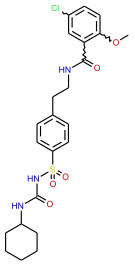
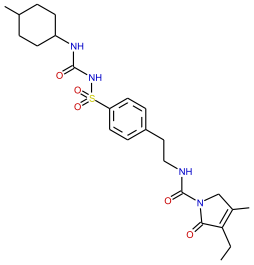
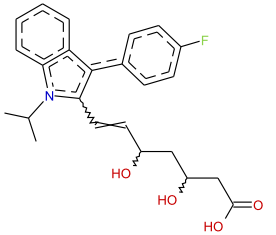
Bayesian Score: -2.4

Mahalanobis Distance: 17.6

Mahalanobis Distance p-value: 1.1e-012

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Glyburide	Glimepride	Fluvastatin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.593	0.600	0.615
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

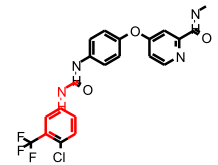
Model Applicability

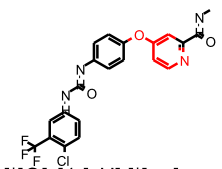
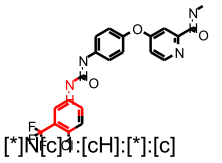
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

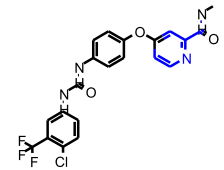
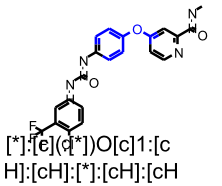
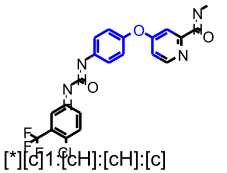
Feature Contribution

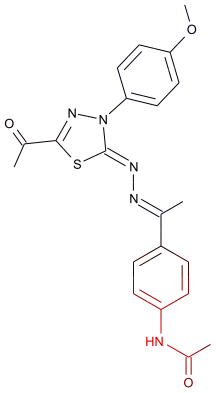
Top features for positive contribution

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

SCFP_6	-754059116	 [*]O[c]1:[cH]:[*]:n:[cH]:[cH]:1	0.415	1 out of 1
SCFP_6	-347281112	 [*]N[c]1:[cH]:[*]:[c]([*]):[c](:[cH]:1)C([*])([*])[*]	0.273	2 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-827073191	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:n:1	-0.674	0 out of 3
SCFP_6	-488587948	 [*]:[c](:[*])O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.496	0 out of 2
SCFP_6	-975241316	 [*][c]1:[cH]:[cH]:[c](O[c](:[cH]:[*]):[cH]:[*]):[cH]:[cH]:1	-0.496	0 out of 2



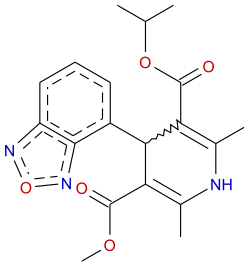
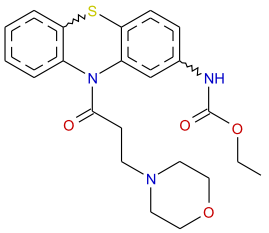
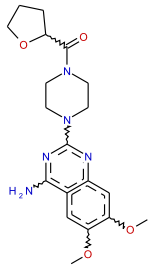
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.562
Enrichment: 1.36
Bayesian Score: 2.95
Mahalanobis Distance: 17.1
Mahalanobis Distance p-value: 1.38e-007

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Isradipine	Moricizine	Terazosin
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.629	0.650	0.681
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

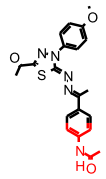
Model Applicability

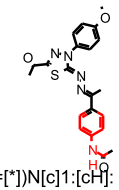
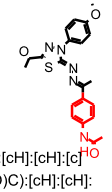
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC4 out of range. Value: 5.2707. Training min, max, SD, explained variance: -6.1092, 5.1042, 2.173, 0.0642.

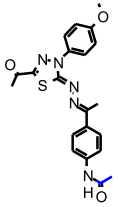
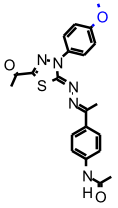
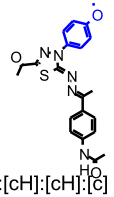
Feature Contribution

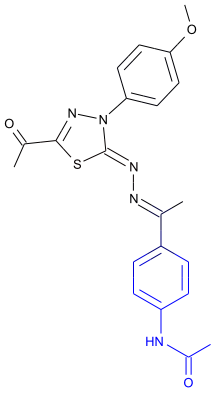
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	2097618059	 [*]:[cH]:[c](NC(=O)C):[cH]:[*]	0.681	6 out of 7

SCFP_8	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.574	4 out of 5
SCFP_8	814408713	 [*][c]1:[cH]:[cH]:[c](NC(=O)C):[cH]:[cH]:1	0.553	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	136627117	 [*]C(=[*])C	-0.41	4 out of 18
SCFP_8	136239834	 [*]OC	-0.358	3 out of 13
SCFP_8	1287669168	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	-0.31	0 out of 1



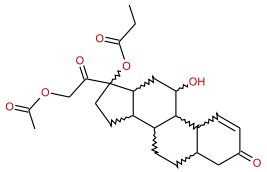
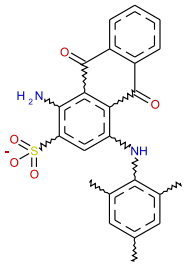
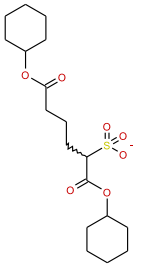
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 6.57e-005
Enrichment: 7.14e-005
Bayesian Score: -9.11
Mahalanobis Distance: 9.91
Mahalanobis Distance p-value: 0.0783

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	Butanedioic acid, sulfo-, 1,4-dicyclohexyl ester, sodium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.643	0.768	0.808
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: -,1327,1986	ATDAEI Acute Toxicity Data. Journal of the American College of Toxicology, Pa rt B. (Mary Ann Liebert, Inc., 1651 Third Ave., New York, NY 10128) V.1- 1990- Volume(issue)/page/year: 1,108,1990

Model Applicability

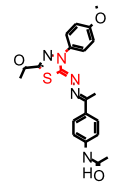
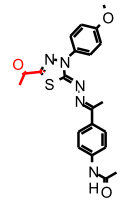
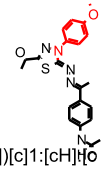
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

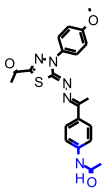
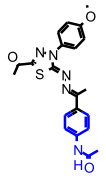
Feature Contribution

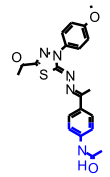
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

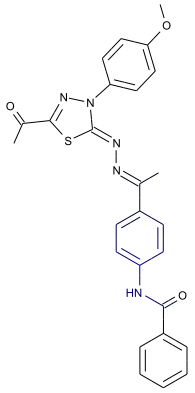
FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79
FCFP_12	356782498	 [*]N([*])[c]1:[cH]:[cH]:[c](OC):[cH]:[cH]:1	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1944671191	 [*]:[c](:[*])NC(=O)C	-1.87	0 out of 6
FCFP_12	1907952166	 CC(=O)N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.72	0 out of 5

FCFP_12	1175665944	 [*]:[cH]:[c](NC(=O)C) :[cH]:[*]	-1.02	2 out of 8

TOPKAT_Skin_Irritancy_None_vs_Irritant



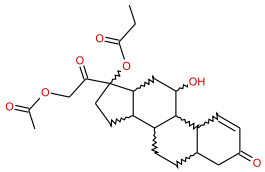
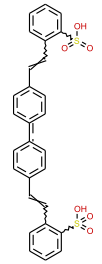
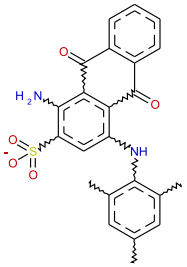
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 0.898
Enrichment: 0.975
Bayesian Score: -2.78
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 0.0322

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.778	0.780	0.878
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	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986

Model Applicability

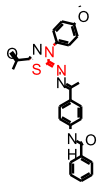
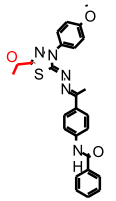
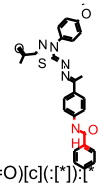
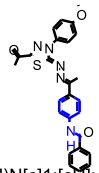
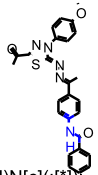
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

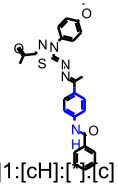
1. All properties and OPS components are within expected ranges.

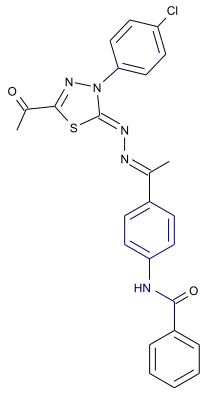
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.0734	5 out of 5
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Irritant in training set**
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-0.486	12 out of 22

FCFP_12	-773983804	 [*]N[c]1:[cH]:[cH]:[c] ([*]):[cH]:[cH]:1	-0.444	46 out of 79



$C_{25}H_{20}ClN_5O_2S$

Molecular Weight: 489.97659

ALogP: 5.409

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.816

Enrichment: 0.886

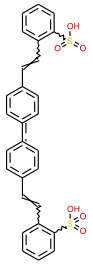
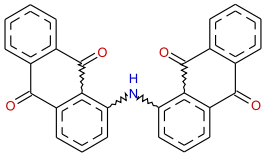
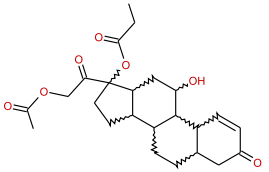
Bayesian Score: -3.42

Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 0.00706

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	Anthraquinone, 1,1'-iminodi-	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.721	0.796	0.817
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986	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

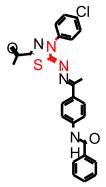
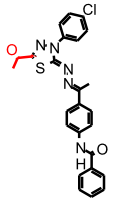
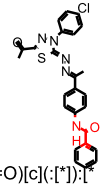
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

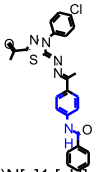
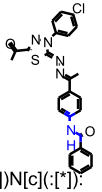
Feature Contribution

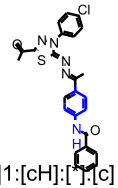
Top features for positive contribution

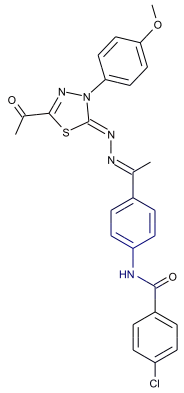
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.0734	5 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-0.486	12 out of 22

FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.444	46 out of 79



$C_{26}H_{22}ClN_5O_3S$

Molecular Weight: 520.00258

ALogP: 5.392

Rotatable Bonds: 7

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.873

Enrichment: 0.948

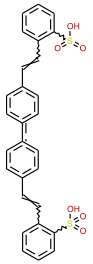
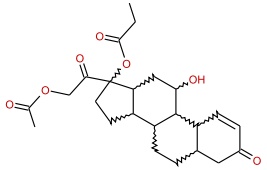
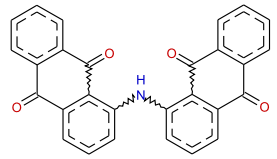
Bayesian Score: -3.02

Mahalanobis Distance: 10.8

Mahalanobis Distance p-value: 0.00565

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	Anthraquinone, 1,1'-iminodi-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.764	0.860	0.935
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	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: -,735,1986

Model Applicability

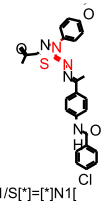
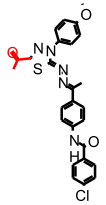
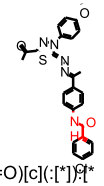
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

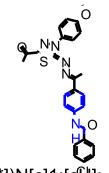
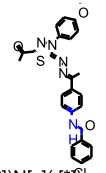
Feature Contribution

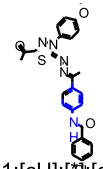
Top features for positive contribution

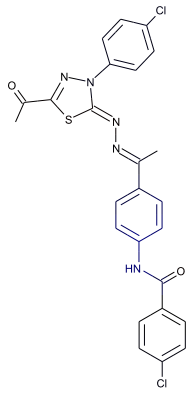
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.0734	5 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-0.486	12 out of 22

FCFP_12	-773983804	 [*]N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	-0.444	46 out of 79



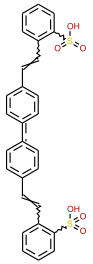
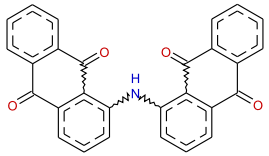
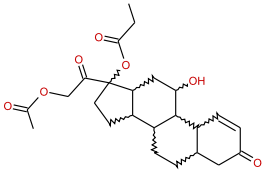
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 0.84
Enrichment: 0.912
Bayesian Score: -3.27
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.00383

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	Anthraquinone, 1,1'-iminodi-	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.719	0.842	0.903
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986	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

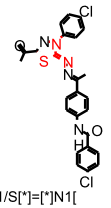
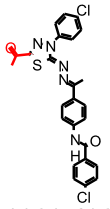
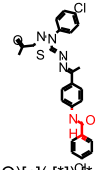
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

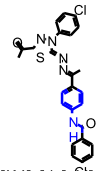
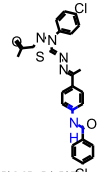
Feature Contribution

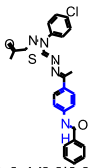
Top features for positive contribution

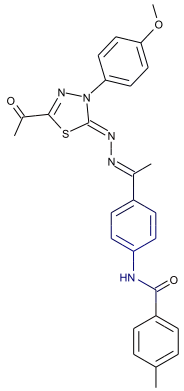
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.0734	5 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-0.486	12 out of 22

FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.444	46 out of 79



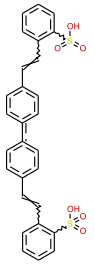
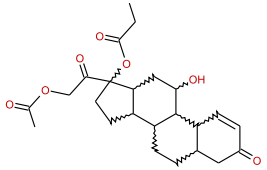
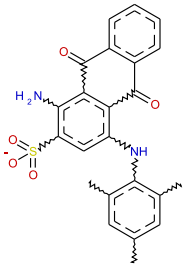
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 0.91
Enrichment: 0.988
Bayesian Score: -2.64
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.021

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.770	0.828	0.917
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	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: -,1327,1986

Model Applicability

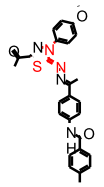
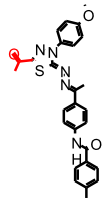
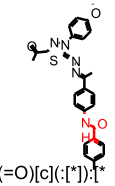
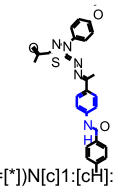
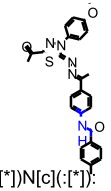
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

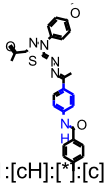
1. All properties and OPS components are within expected ranges.

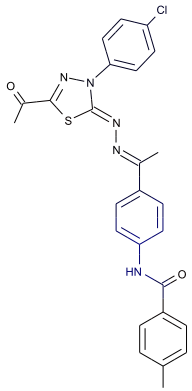
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.0734	5 out of 5
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Irritant in training set**
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-0.486	12 out of 22

FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.444	46 out of 79



$C_{26}H_{22}ClN_5O_2S$

Molecular Weight: 504.00318

ALogP: 5.895

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.829

Enrichment: 0.901

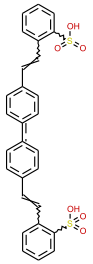
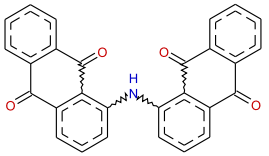
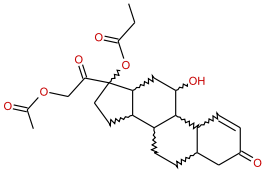
Bayesian Score: -3.34

Mahalanobis Distance: 10.8

Mahalanobis Distance p-value: 0.00516

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	Anthraquinone, 1,1'-iminodi-	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.721	0.819	0.871
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986	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

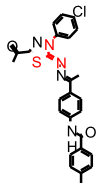
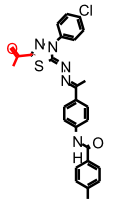
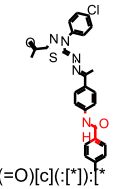
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

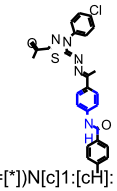
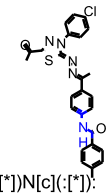
Feature Contribution

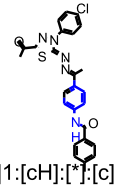
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.075	78 out of 79
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.0734	5 out of 5

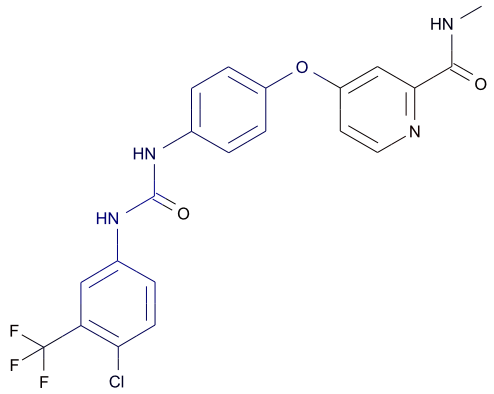
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-0.486	12 out of 22

FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.444	46 out of 79

Sorafenib

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Irritant

Probability: 0.264

Enrichment: 0.287

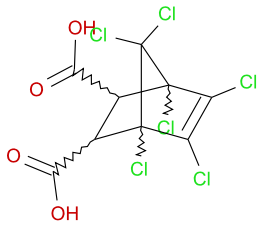
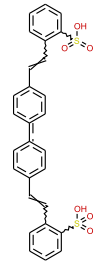
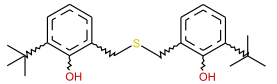
Bayesian Score: -5.23

Mahalanobis Distance: 8.27

Mahalanobis Distance p-value: 0.791

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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-	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	Sulfide, bis(4-t-butyl-m-cresyl)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.844	0.871	0.884
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	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

Model Applicability

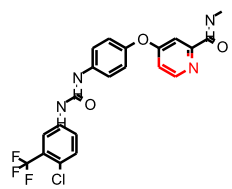
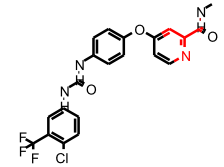
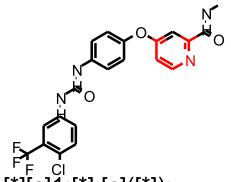
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

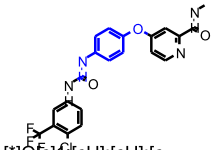
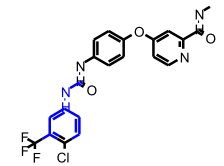
Feature Contribution

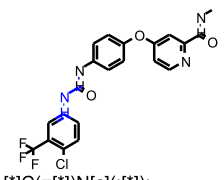
Top features for positive contribution

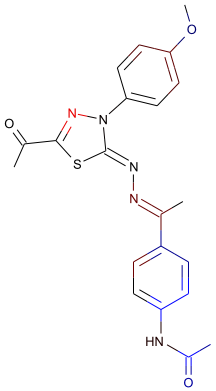
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-124655670	 [*]:[cH]:[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1539132615	 [*]C(=[*])[c](:[cH]:[*]):n:[*]	0.0795	9 out of 9
FCFP_12	-1695756380	 [*][c]1:[*]:[c]([*]):n:[cH]:[cH]:1	0.0772	7 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-789307649	 [*]O[c]1:[cH]:[cH]:[cH]:[c](NC(=[*])[*]):[cH]:[cH]:1	-1.54	0 out of 4
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12

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



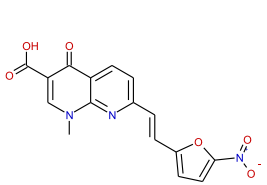
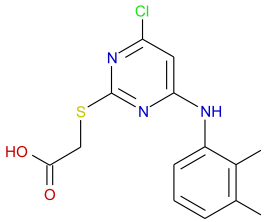
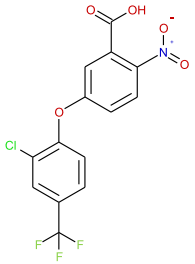
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 25.2
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 2.2e-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	470	[4-Chloro-6-(2,3-xylidino)-2-pyri-midinylthio]acetic acid s	Acifluorfen
Structure			
Actual Endpoint (-log C)	4.62839	4.47685	3.40908
Predicted Endpoint (-log C)	3.93264	3.8529	3.10974
Distance	0.674	0.738	0.744
Reference	CPDB	CPDB	CPDB

Model Applicability

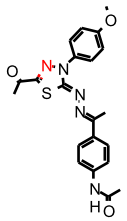
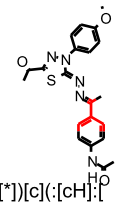
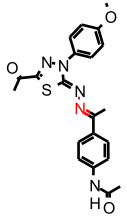
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

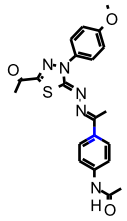
Feature Contribution

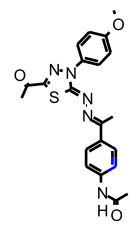
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

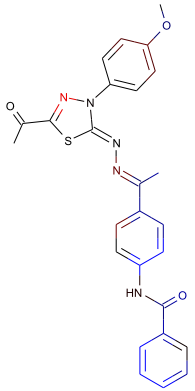
ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107
ECFP_6	-1087070950	 [*]N=[*]	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232

TOPKAT_Carcinogenic_Potency_TD50_Mouse



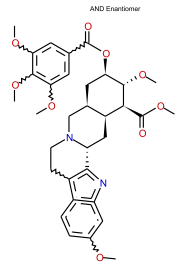
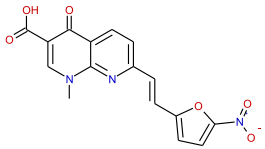
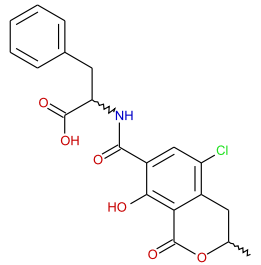
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 19.1
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.3
Mahalanobis Distance p-value: 5.86e-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	223	470	Ochratoxin A
Structure			
Actual Endpoint (-log C)	5.08368	4.62839	4.79932
Predicted Endpoint (-log C)	5.08273	3.93264	3.6353
Distance	0.789	0.837	0.845
Reference	CPDB	CPDB	CPDB

Model Applicability

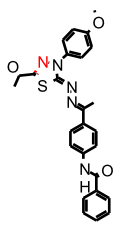
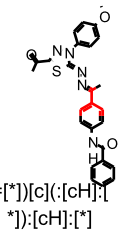
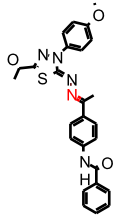
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

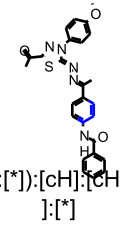
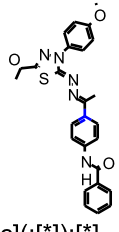
Feature Contribution

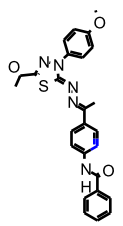
Top features for positive contribution

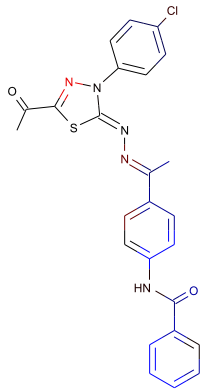
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107
ECFP_6	-1087070950	 [*]N=[*]	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



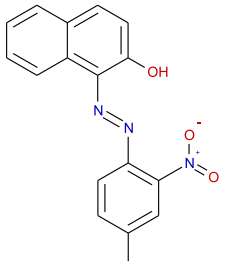
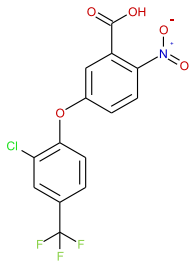
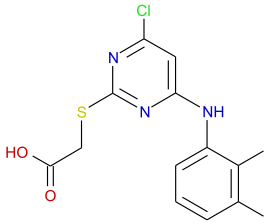
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 10.1
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 1.67e-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	646	Acifluorfen	[4-Chloro-6-(2,3-xylidino)-2-pyri-midinylthio]acetic acid s
Structure			
Actual Endpoint (-log C)	0.937339	3.40908	4.47685
Predicted Endpoint (-log C)	3.26294	3.10974	3.8529
Distance	0.788	0.826	0.828
Reference	CPDB	CPDB	CPDB

Model Applicability

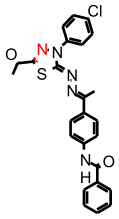
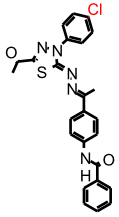
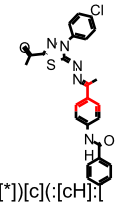
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

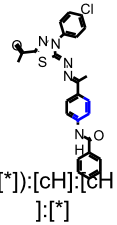
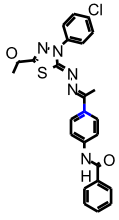
Feature Contribution

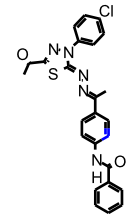
Top features for positive contribution

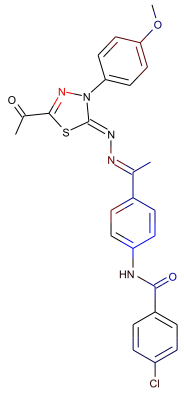
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



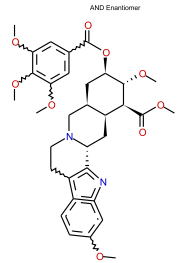
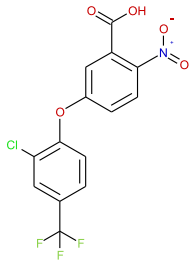
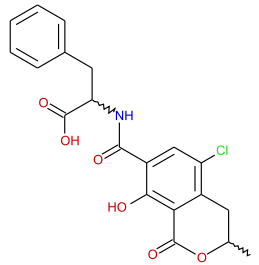
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 5.86
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.7
Mahalanobis Distance p-value: 2.56e-012

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	223	Acifluorfen	Ochratoxin A
Structure			
Actual Endpoint (-log C)	5.08368	3.40908	4.79932
Predicted Endpoint (-log C)	5.08273	3.10974	3.6353
Distance	0.781	0.871	0.878
Reference	CPDB	CPDB	CPDB

Model Applicability

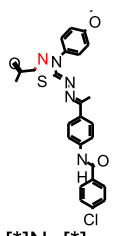
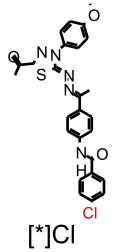
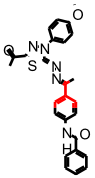
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC24 out of range. Value: -4.6195. Training min, max, SD, explained variance: -4.4826, 3.8729, 1.034, 0.0133.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

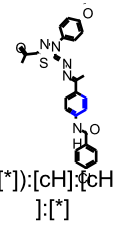
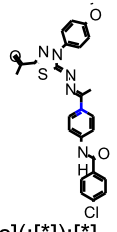
Feature Contribution

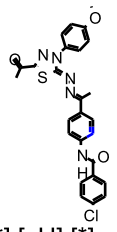
Top features for positive contribution

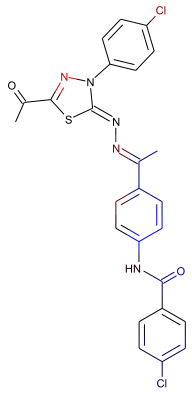
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



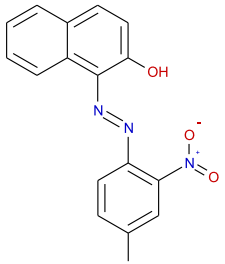
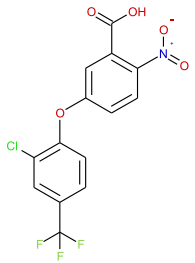
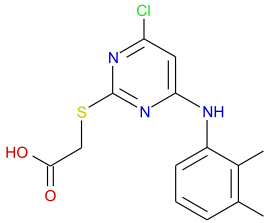
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 3.51
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.6
Mahalanobis Distance p-value: 1.36e-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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	646	Acifluorfen	[4-Chloro-6-(2,3-xylidino)-2-pyri-midinylthio]acetic acid s
Structure			
Actual Endpoint (-log C)	0.937339	3.40908	4.47685
Predicted Endpoint (-log C)	3.26294	3.10974	3.8529
Distance	0.829	0.854	0.863
Reference	CPDB	CPDB	CPDB

Model Applicability

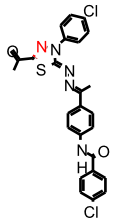
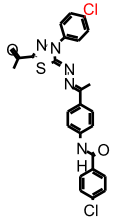
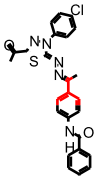
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC20 out of range. Value: 3.7769. Training min, max, SD, explained variance: -4.3384, 3.4394, 1.14, 0.0162.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

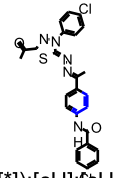
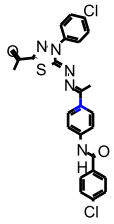
Feature Contribution

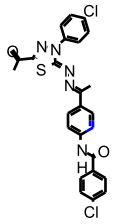
Top features for positive contribution

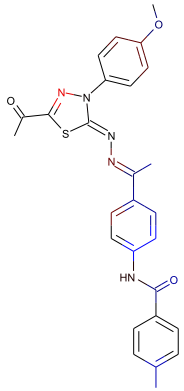
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



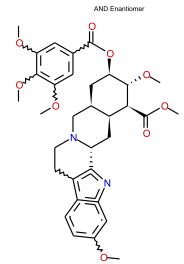
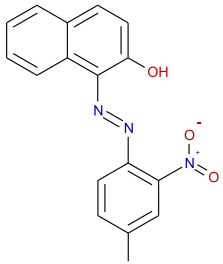
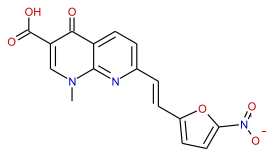
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 11
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.7
Mahalanobis Distance p-value: 4.27e-012

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	223	646	470
Structure			
Actual Endpoint (-log C)	5.08368	0.937339	4.62839
Predicted Endpoint (-log C)	5.08273	3.26294	3.93264
Distance	0.787	0.852	0.861
Reference	CPDB	CPDB	CPDB

Model Applicability

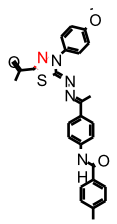
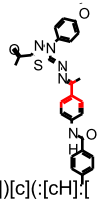
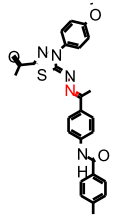
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

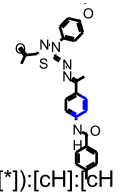
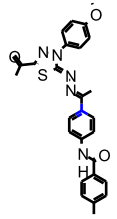
Feature Contribution

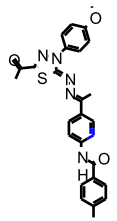
Top features for positive contribution

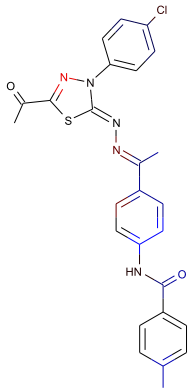
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107
ECFP_6	-1087070950	 [*]N=[*]	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232



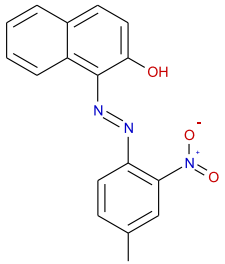
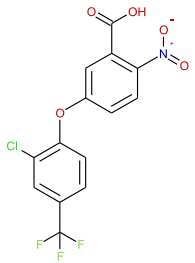
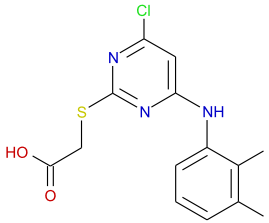
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 5.83
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.1
Mahalanobis Distance p-value: 8.89e-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	646	Acifluorfen	[4-Chloro-6-(2,3-xylidino)-2-pyri-midinylthio]acetic acid s
Structure			
Actual Endpoint (-log C)	0.937339	3.40908	4.47685
Predicted Endpoint (-log C)	3.26294	3.10974	3.8529
Distance	0.798	0.841	0.843
Reference	CPDB	CPDB	CPDB

Model Applicability

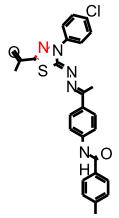
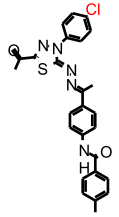
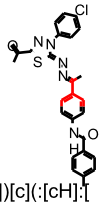
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC20 out of range. Value: 3.6466. Training min, max, SD, explained variance: -4.3384, 3.4394, 1.14, 0.0162.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

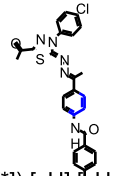
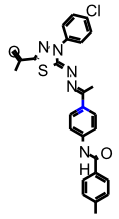
Feature Contribution

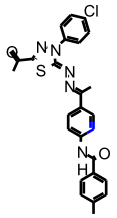
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

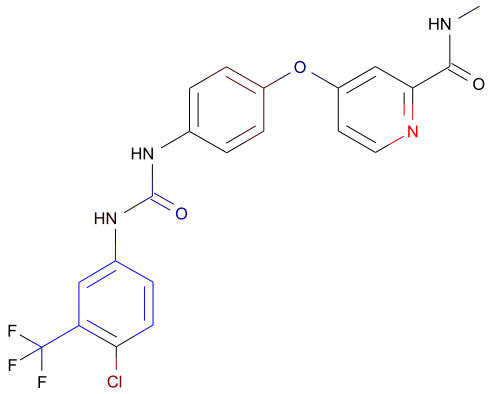
ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 19.2

Unit: mg/kg_body_weight/day

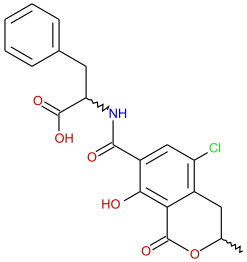
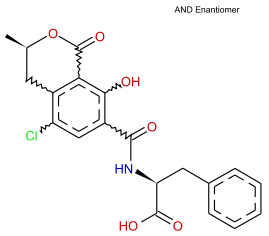
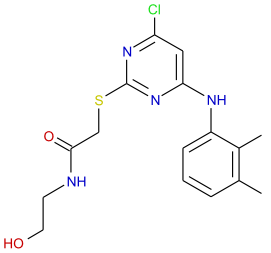
Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 2.94e-006

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

TOPKAT_Carcinogenic_Potency_TD50_Mouse

Structural Similar Compounds

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

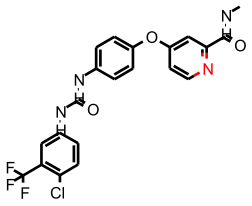
Model Applicability

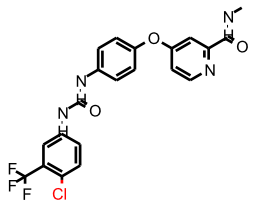
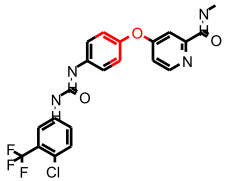
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

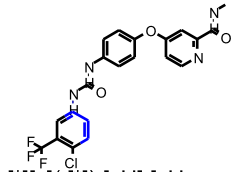
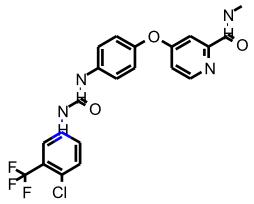
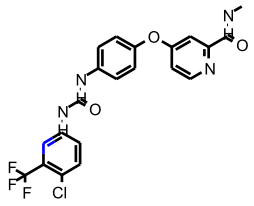
Feature Contribution

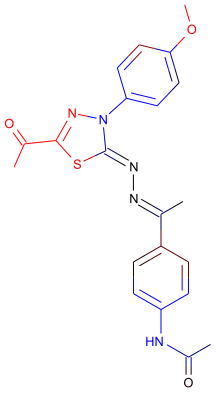
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]N=[*]	0.229

ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.0818

Top Features for negative contribution

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



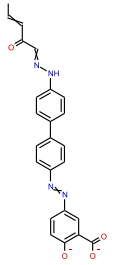
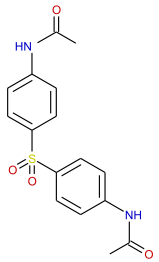
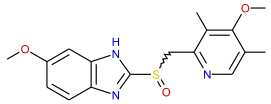
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 2.46
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.1
Mahalanobis Distance p-value: 1.52e-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. direct brown 95	4,4´-Sulfonylbisacetanilide	Omeprazole
Structure			
Actual Endpoint (-log C)	5.31387	3.77655	3.4628
Predicted Endpoint (-log C)	4.30266	3.55337	4.7324
Distance	0.616	0.697	0.697
Reference	CPDB	CPDB	CPDB

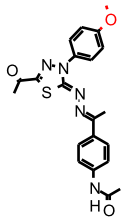
Model Applicability

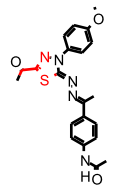
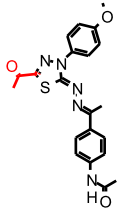
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

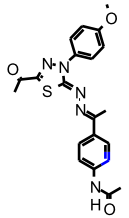
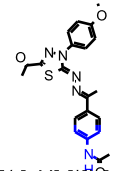
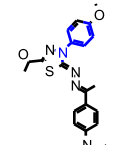
1. OPS PC7 out of range. Value: -5.4904. Training min, max, SD, explained variance: -5.0422, 6.1749, 1.868, 0.0335.

Feature Contribution

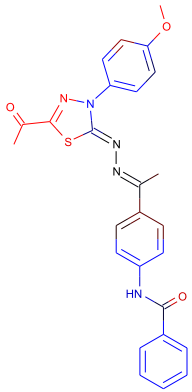
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357
FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	-0.323
FCFP_6	1674451008	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH] :1	-0.233

TOPKAT_Carcinogenic_Potency_TD50_Rat



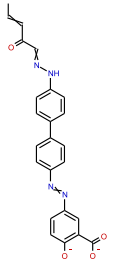
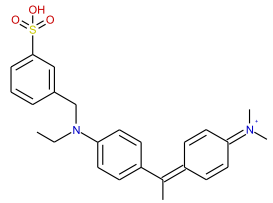
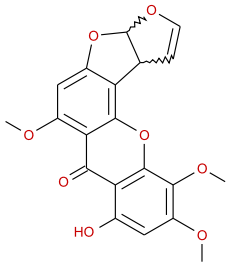
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 3.63
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.4
Mahalanobis Distance p-value: 1.69e-009

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the 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	FD & C violet no. 1	5,6-Dimethoxysterigmatocystin
Structure			
Actual Endpoint (-log C)	5.31387	2.8543	6.02361
Predicted Endpoint (-log C)	4.30266	3.40838	4.98771
Distance	0.606	0.779	0.780
Reference	CPDB	CPDB	CPDB

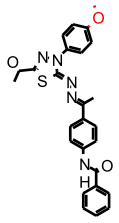
Model Applicability

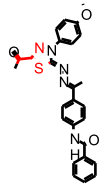
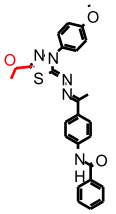
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

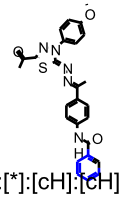
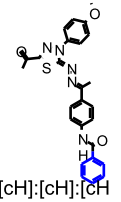
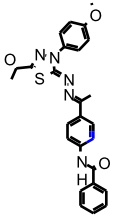
1. OPS PC7 out of range. Value: -5.7577. Training min, max, SD, explained variance: -5.0422, 6.1749, 1.868, 0.0335.

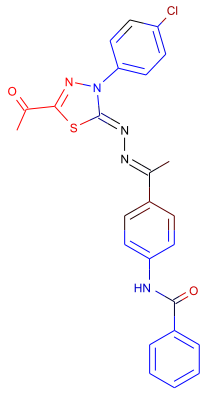
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	565998553		0.357
FCFP_6	565968762		0.266

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	991735244		-0.422
FCFP_6	-2093839777		-0.378
FCFP_6	16		-0.354



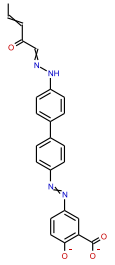
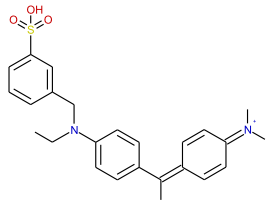
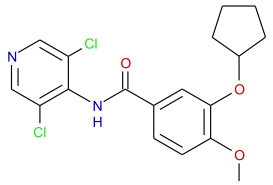
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 12
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.2
Mahalanobis Distance p-value: 7.46e-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	C.I. direct brown 95	FD & C violet no. 1	3-(Cyclopentyloxy)-N-(3,5-di-chloro-4-pyridyl)-4-methoxy-benzamide
Structure			
Actual Endpoint (-log C)	5.31387	2.8543	5.39369
Predicted Endpoint (-log C)	4.30266	3.40838	4.27874
Distance	0.659	0.719	0.748
Reference	CPDB	CPDB	CPDB

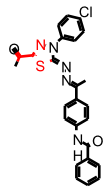
Model Applicability

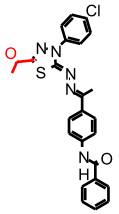
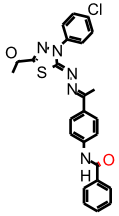
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

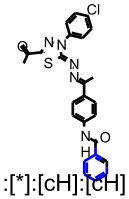
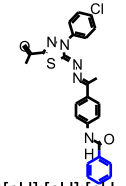
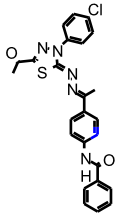
Feature Contribution

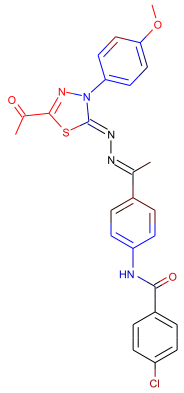
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	-2093839777	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.378
FCFP_6	16	 [*]:[cH]:[*]	-0.354



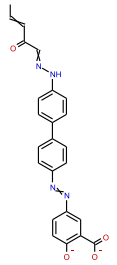
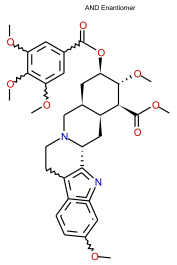
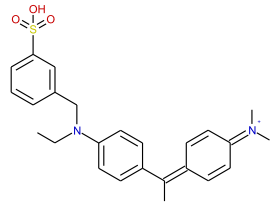
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.349
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.7
Mahalanobis Distance p-value: 2.92e-010

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	C.I. direct brown 95	223	FD & C violet no. 1
Structure			
Actual Endpoint (-log C)	5.31387	6.29867	2.8543
Predicted Endpoint (-log C)	4.30266	7.5657	3.40838
Distance	0.647	0.787	0.796
Reference	CPDB	CPDB	CPDB

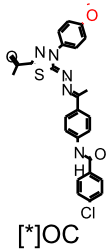
Model Applicability

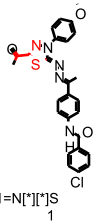
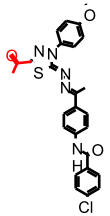
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

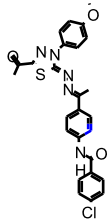
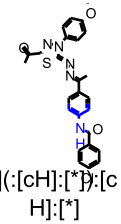
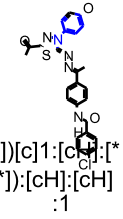
1. OPS PC7 out of range. Value: -5.3084. Training min, max, SD, explained variance: -5.0422, 6.1749, 1.868, 0.0335.

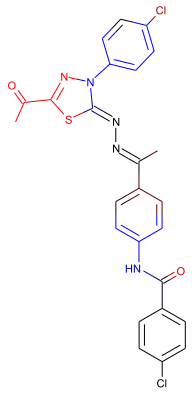
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	565998553		0.357
FCFP_6	565968762		0.266

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	16		-0.354
FCFP_6	590925877		-0.323
FCFP_6	1674451008		-0.233



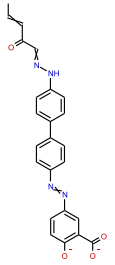
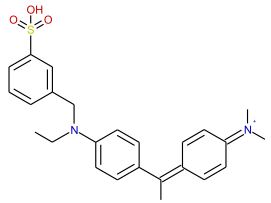
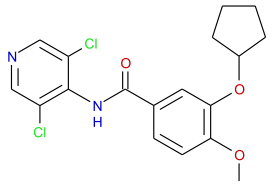
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 1.66
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.5
Mahalanobis Distance p-value: 3.62e-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	FD & C violet no. 1	3-(Cyclopentyloxy)-N-(3,5-di-chloro-4-pyridyl)-4-methoxy-benzamide
Structure			
Actual Endpoint (-log C)	5.31387	2.8543	5.39369
Predicted Endpoint (-log C)	4.30266	3.40838	4.27874
Distance	0.696	0.742	0.776
Reference	CPDB	CPDB	CPDB

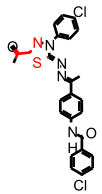
Model Applicability

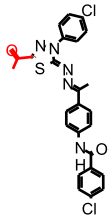
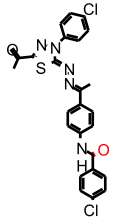
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

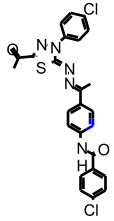
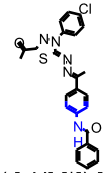
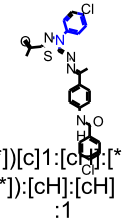
Feature Contribution

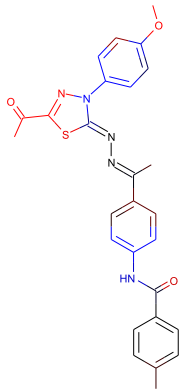
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.323
FCFP_6	1674451008	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.233



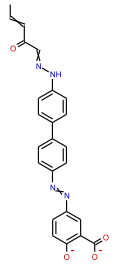
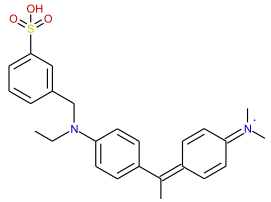
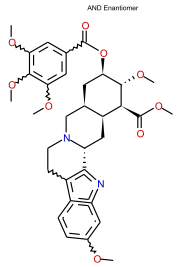
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.451
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15.1
Mahalanobis Distance p-value: 2.14e-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. direct brown 95	FD & C violet no. 1	223
Structure			
Actual Endpoint (-log C)	5.31387	2.8543	6.29867
Predicted Endpoint (-log C)	4.30266	3.40838	7.5657
Distance	0.631	0.790	0.790
Reference	CPDB	CPDB	CPDB

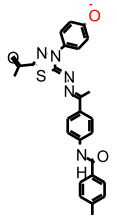
Model Applicability

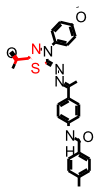
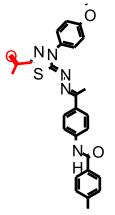
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

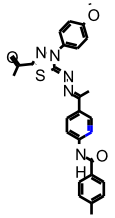
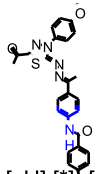
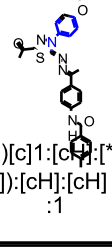
Feature Contribution

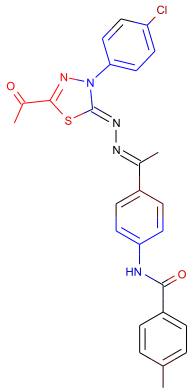
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	565998553		0.357
FCFP_6	565968762		0.266

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16		-0.354
FCFP_6	590925877		-0.323
FCFP_6	1674451008		-0.233



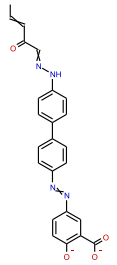
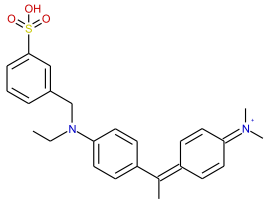
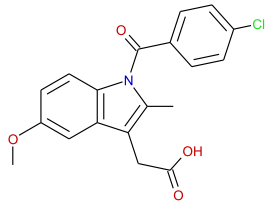
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 1.5
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.3
Mahalanobis Distance p-value: 3.32e-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	C.I. direct brown 95	FD & C violet no. 1	Indomethacin
Structure			
Actual Endpoint (-log C)	5.31387	2.8543	5.49293
Predicted Endpoint (-log C)	4.30266	3.40838	4.9569
Distance	0.684	0.735	0.768
Reference	CPDB	CPDB	CPDB

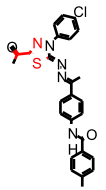
Model Applicability

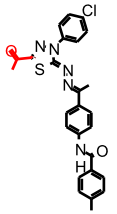
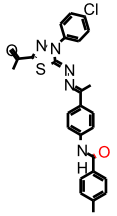
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

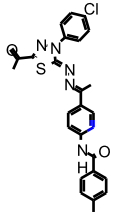
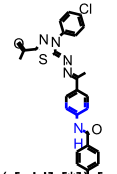
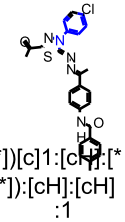
1. All properties and OPS components are within expected ranges.

Feature Contribution

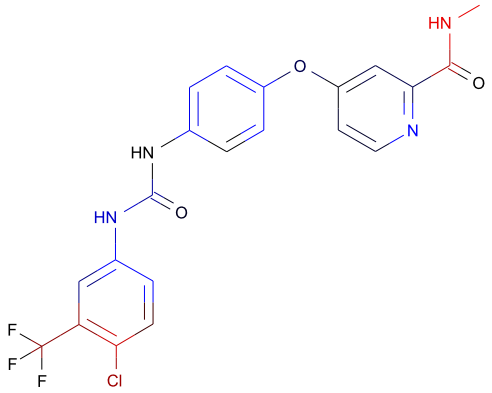
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.323
FCFP_6	1674451008	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.233

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 14.2

Unit: mg/kg_body_weight/day

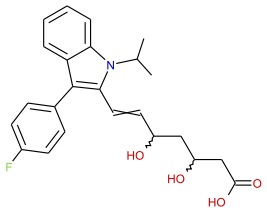
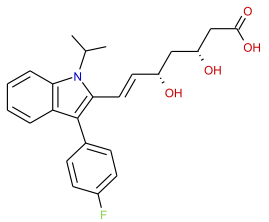
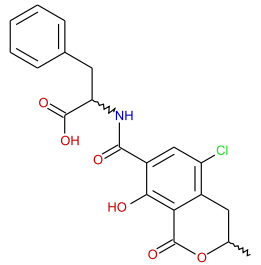
Mahalanobis Distance: 20.4

Mahalanobis Distance p-value: 9.56e-031

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

TOPKAT_Carcinogenic_Potency_TD50_Rat

Structural Similar Compounds

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

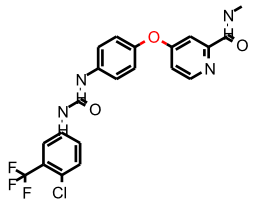
Model Applicability

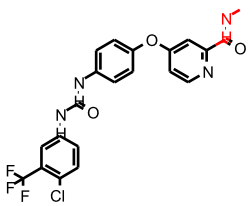
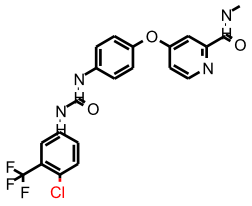
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

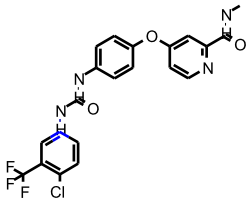
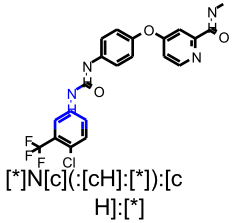
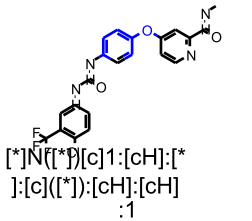
Feature Contribution

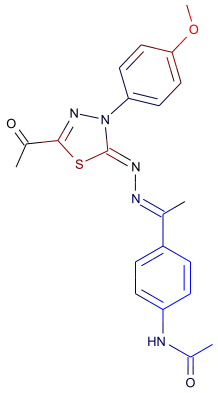
Top features for positive contribution

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

FCFP_6	-885550502	 [*]C(=[*])NC	0.229
FCFP_6	32	 [*]Cl	0.154

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.323
FCFP_6	1674451008	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.233



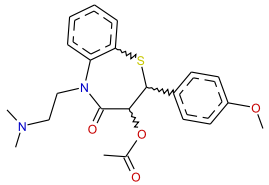
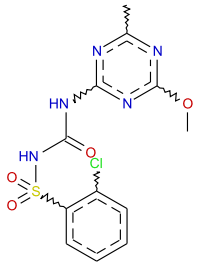
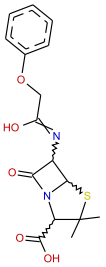
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0341
Unit: g/kg_body_weight
Mahalanobis Distance: 30.9
Mahalanobis Distance p-value: 2.16e-026

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

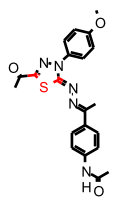
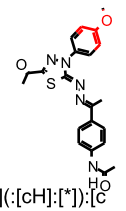
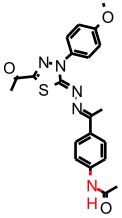
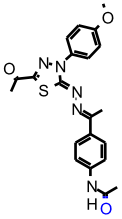
Name	DILTIAZEM	CHLORSULFURON	PENICILLIN VK
Structure			
Actual Endpoint (-log C)	4.21961	4.15566	2.99188
Predicted Endpoint (-log C)	4.005	3.79771	4.18433
Distance	0.596	0.659	0.677
Reference	NDA-18602	EPA COVER SHEET 0027;880301;(1)	NTP REPORT # 336

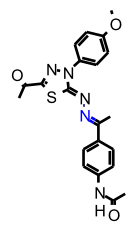
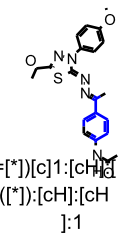
Model Applicability

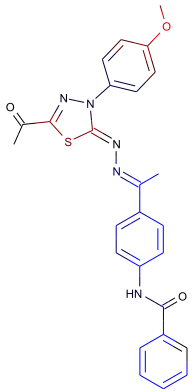
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC32 out of range. Value: 4.5485. Training min, max, SD, explained variance: -4.2021, 4.2975, 1.228, 0.0066.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown ECFP_6 feature: -474544785: [*]NC(=O)C
7. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
8. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
10. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
11. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
12. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
13. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
14. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
16. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

Feature Contribution

Top features for positive contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.106
FCFP_6	3	 [*]N[*]	0.0924
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	1	 [*]=O	-0.102

ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[c *]:[c]([*]):[cH]:[cH]:1	-0.0906



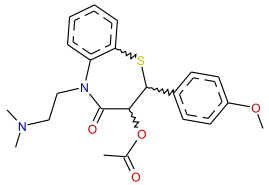
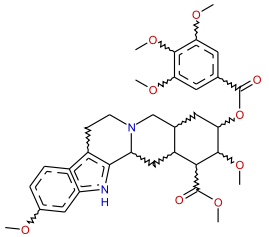
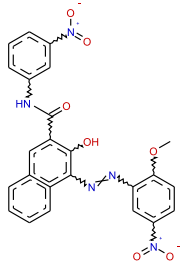
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0407
Unit: g/kg_body_weight
Mahalanobis Distance: 31.3
Mahalanobis Distance p-value: 5.06e-027

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

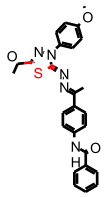
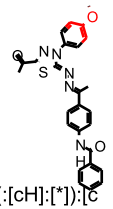
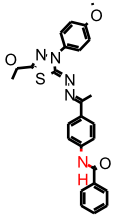
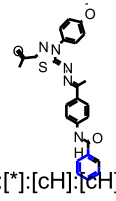
Name	DILTIAZEM	RESERPINE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	4.21961	6.38645	2.28997
Predicted Endpoint (-log C)	4.005	5.548	3.52921
Distance	0.706	0.710	0.746
Reference	NDA-18602	NTP 193 22	NTP 411 146

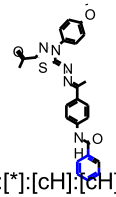
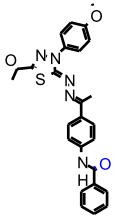
Model Applicability

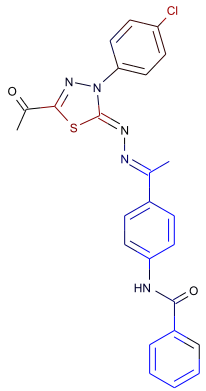
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC32 out of range. Value: 4.3704. Training min, max, SD, explained variance: -4.2021, 4.2975, 1.228, 0.0066.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
7. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
8. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
12. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
13. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
14. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
15. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

Feature Contribution

Top features for positive contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.106
FCFP_6	3	 [*]N[*]	0.0924
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134

ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133
FCFP_6	1	 [*]=O	-0.102



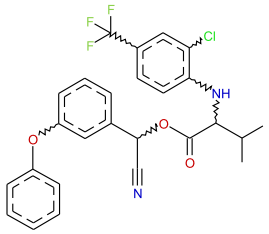
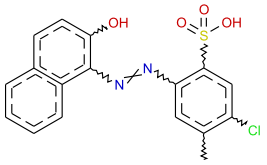
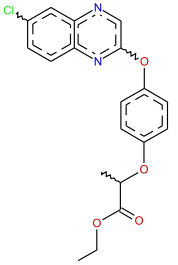
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0654
Unit: g/kg_body_weight
Mahalanobis Distance: 31.8
Mahalanobis Distance p-value: 7.4e-028

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	FLUVALINATE	D & C RED 9	ASSURE
Structure			
Actual Endpoint (-log C)	5.30356	3.87715	5.00328
Predicted Endpoint (-log C)	4.89944	3.6546	4.27671
Distance	0.707	0.720	0.726
Reference	EPA COVER SHEET 0281;880630;(1)	NTP REPORT # 225	EPA COVER SHEET 0335;891001;(1)

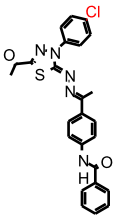
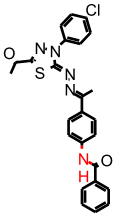
Model Applicability

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

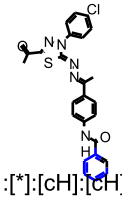
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
7. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
8. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
12. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
13. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
14. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
15. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
18. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

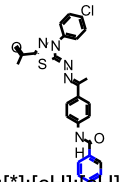
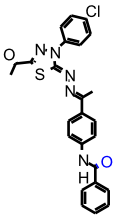
Feature Contribution

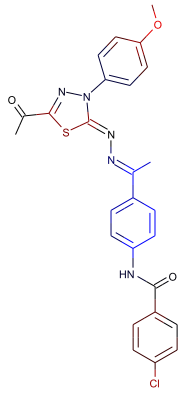
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134

ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133
FCFP_6	1	 [*]=O	-0.102



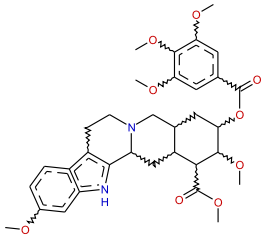
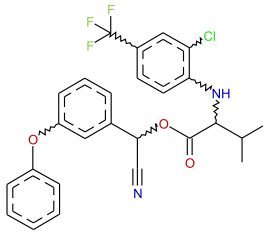
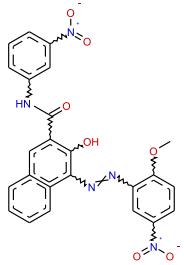
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.013
Unit: g/kg_body_weight
Mahalanobis Distance: 31.3
Mahalanobis Distance p-value: 4.19e-027

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	RESERPINE	FLUVALINATE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	6.38645	5.30356	2.28997
Predicted Endpoint (-log C)	5.548	4.89944	3.52921
Distance	0.695	0.729	0.755
Reference	NTP 193 22	EPA COVER SHEET 0281;880630;(1)	NTP 411 146

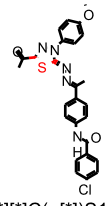
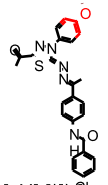
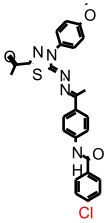
Model Applicability

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

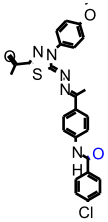
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
7. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
8. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
12. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
13. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
14. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
15. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
18. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

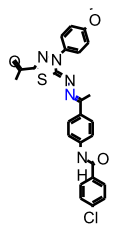
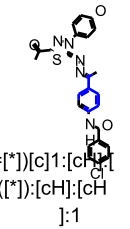
Feature Contribution

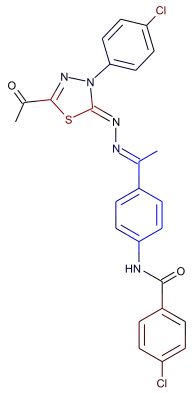
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.106
FCFP_6	32	 [*]Cl	0.101

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	-0.102

ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.0906



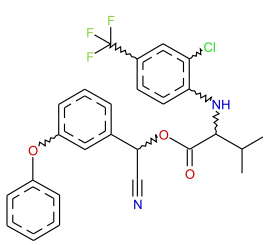
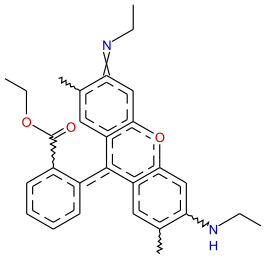
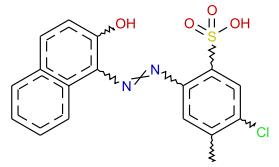
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0299
Unit: g/kg_body_weight
Mahalanobis Distance: 30.9
Mahalanobis Distance p-value: 3.23e-026

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

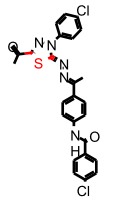
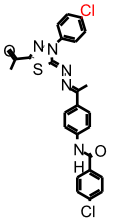
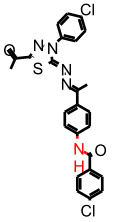
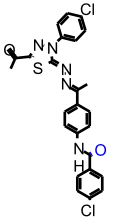
Name	FLUVALINATE	RHODAMINE 6G	D & C RED 9
Structure			
Actual Endpoint (-log C)	5.30356	4.54906	3.87715
Predicted Endpoint (-log C)	4.89944	4.6787	3.6546
Distance	0.693	0.747	0.763
Reference	EPA COVER SHEET 0281;880630;(1)	NTP 364 39	NTP REPORT # 225

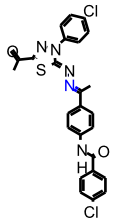
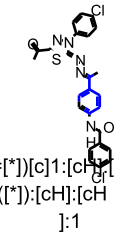
Model Applicability

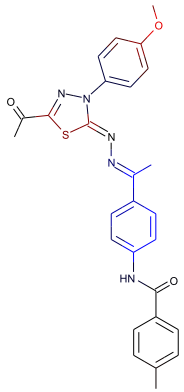
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
7. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
8. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
12. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
13. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
14. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
15. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

Feature Contribution

Top features for positive contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	1	 [*]=O	-0.102

ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[c]([*]):[c]([*]):[cH]:[cH]:1	-0.0906



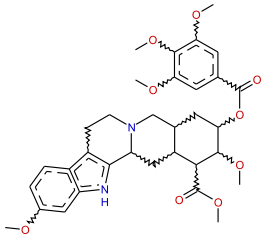
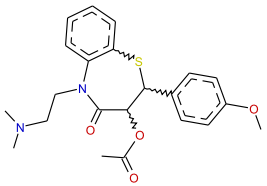
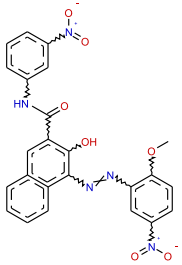
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.018
Unit: g/kg_body_weight
Mahalanobis Distance: 31
Mahalanobis Distance p-value: 2.12e-026

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

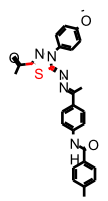
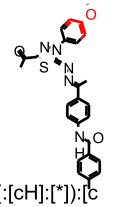
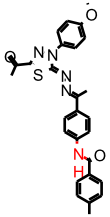
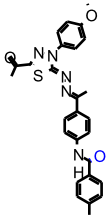
Name	RESERPINE	DILTIAZEM	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	6.38645	4.21961	2.28997
Predicted Endpoint (-log C)	5.548	4.005	3.52921
Distance	0.698	0.740	0.747
Reference	NTP 193 22	NDA-18602	NTP 411 146

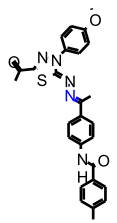
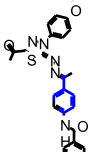
Model Applicability

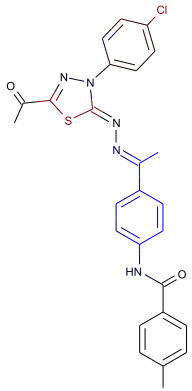
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown ECFP_6 feature: -179515162: [*]:[cH]:[c](C):[cH]:[*]
7. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
8. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
12. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
13. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
14. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
15. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

Feature Contribution

Top features for positive contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.106
FCFP_6	3	 [*]N[*]	0.0924
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	1	 [*]=O	-0.102

ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.0906



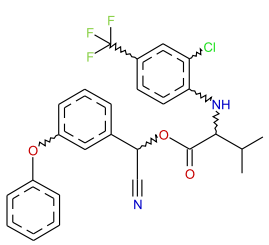
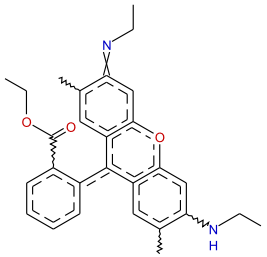
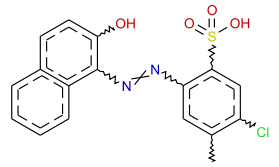
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0289
Unit: g/kg_body_weight
Mahalanobis Distance: 31
Mahalanobis Distance p-value: 1.86e-026

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	FLUVALINATE	RHODAMINE 6G	D & C RED 9
Structure			
Actual Endpoint (-log C)	5.30356	4.54906	3.87715
Predicted Endpoint (-log C)	4.89944	4.6787	3.6546
Distance	0.698	0.735	0.745
Reference	EPA COVER SHEET 0281;880630;(1)	NTP 364 39	NTP REPORT # 225

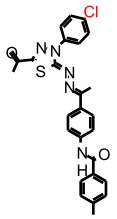
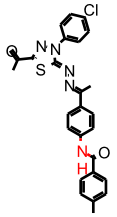
Model Applicability

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

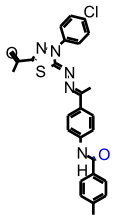
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown ECFP_6 feature: -179515162: [*]:[cH]:[c](C):[cH]:[*]
7. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
8. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
12. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
13. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
14. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
15. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
18. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

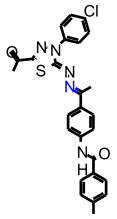
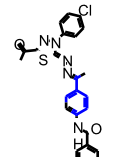
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

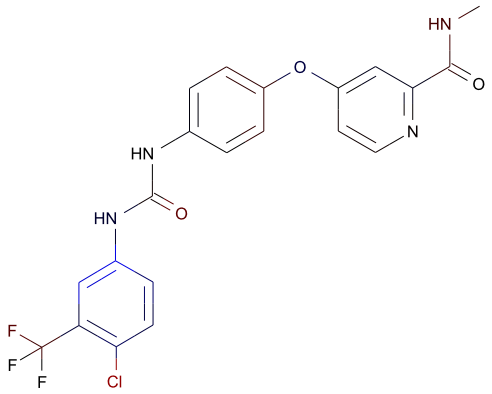
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	-0.102

ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.0906

Sorafenib

TOPKAT_Chronic_LOAEL



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.00483

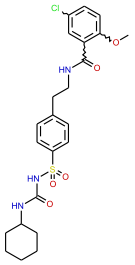
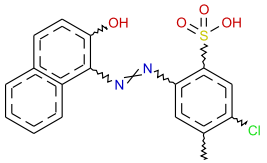
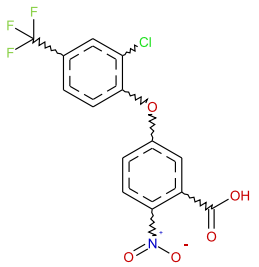
Unit: g/kg_body_weight

Mahalanobis Distance: 30

Mahalanobis Distance p-value: 1.21e-024

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

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

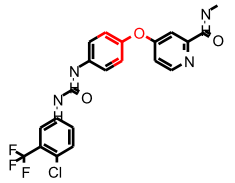
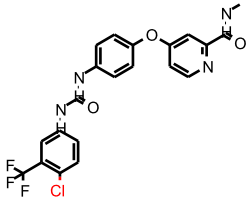
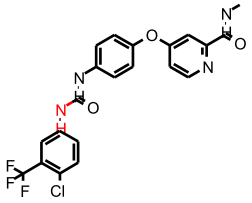
Model Applicability

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

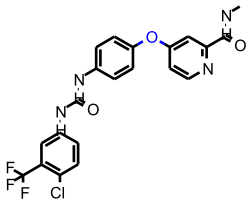
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -1046436026: [*]F
3. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
4. Unknown ECFP_6 feature: 226796801: [*]C([*])([*])F
5. Unknown ECFP_6 feature: 1305253718: [*]:[c](:[*])O[c](:[*]):[*]
6. Unknown ECFP_6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

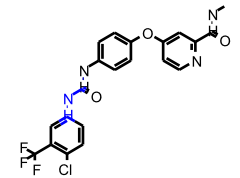
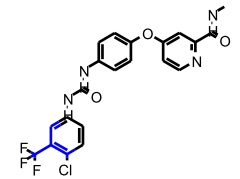
Feature Contribution

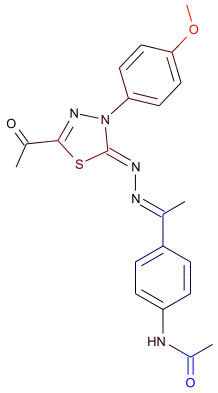
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.106
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	-0.102

ECFP_6	-1236483485	 [*]C(=[*])N[c](:[*]): [*]	-0.0747
FCFP_6	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.0713



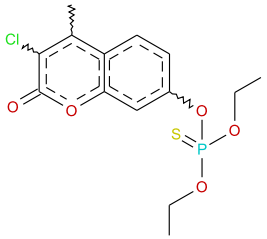
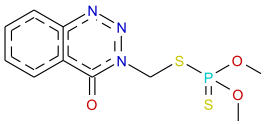
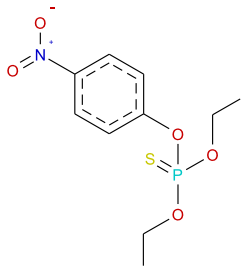
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0292
Unit: g/kg_body_weight
Mahalanobis Distance: 9.65
Mahalanobis Distance p-value: 0.000158

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	COUMAPHOS	AZINPHOSMETHYL	PARATHION
Structure			
Actual Endpoint (-log C)	5.60537	4.65515	5.01172
Predicted Endpoint (-log C)	4.15004	4.22281	3.88389
Distance	0.643	0.665	0.677
Reference	NCI/NTP TR-96	NCI/NTP TR-69	NCI/NTP TR-070

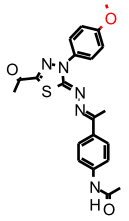
Model Applicability

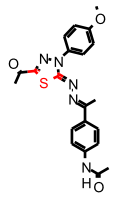
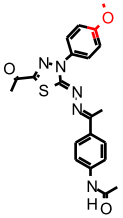
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC9 out of range. Value: 3.7179. Training min, max, SD, explained variance: -2.8548, 3.3954, 1.263, 0.0360.

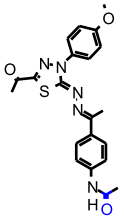
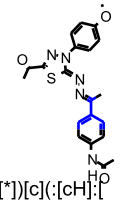
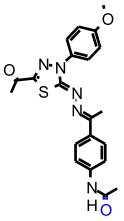
Feature Contribution

Top features for positive contribution

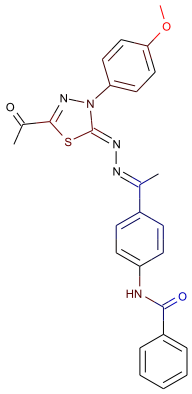
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.0829
FCFP_2	1	 [*]=O	-0.0796

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



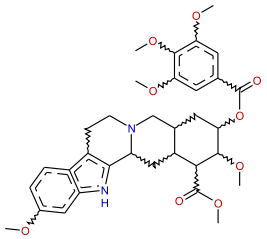
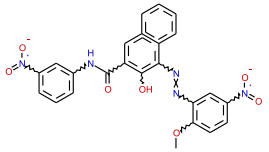
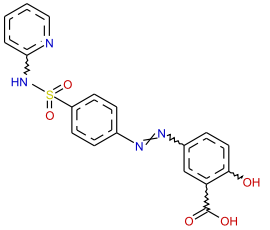
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0279
Unit: g/kg_body_weight
Mahalanobis Distance: 8.69
Mahalanobis Distance p-value: 0.00372

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	RESERPINE	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRIDINE
Structure			
Actual Endpoint (-log C)	6.13118	2.30052	3.375
Predicted Endpoint (-log C)	4.38304	3.55333	2.80292
Distance	0.658	0.748	0.797
Reference	NCI/NTP TR-193	NCI/NTP TR-411	NCI/NTP TR-457

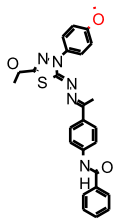
Model Applicability

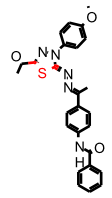
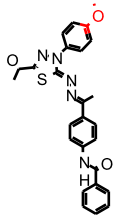
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

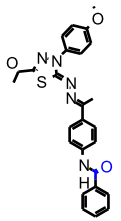
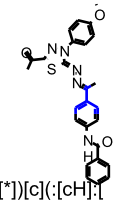
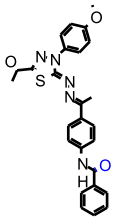
1. All properties and OPS components are within expected ranges.

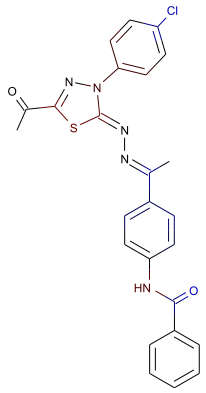
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

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



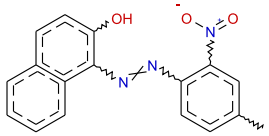
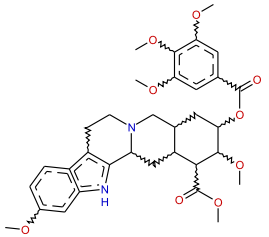
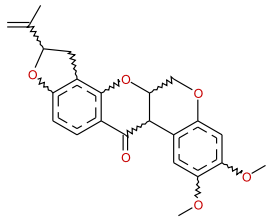
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0624
Unit: g/kg_body_weight
Mahalanobis Distance: 9.31
Mahalanobis Distance p-value: 0.000508

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	RESERPINE	ROTENONE
Structure			
Actual Endpoint (-log C)	2.65635	6.13118	5.06769
Predicted Endpoint (-log C)	2.97957	4.38304	4.11907
Distance	0.773	0.780	0.815
Reference	NCI/NTP TR-407	NCI/NTP TR-193	NCI/NTP TR-320

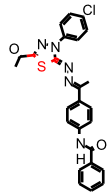
Model Applicability

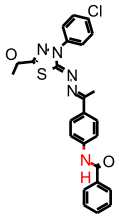
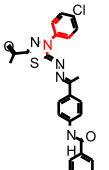
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 5.2711. Training min, max, SD, explained variance: -3.3892, 5.0834, 1.644, 0.0611.

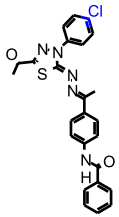
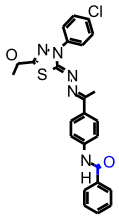
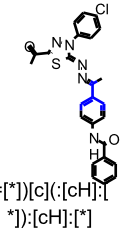
Feature Contribution

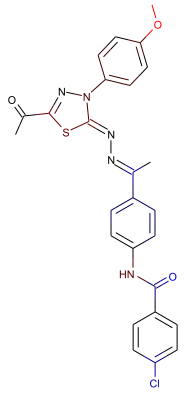
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.0611

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	 [*]:[c](:[*])Cl	-0.134
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



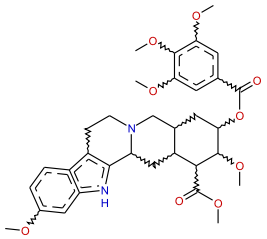
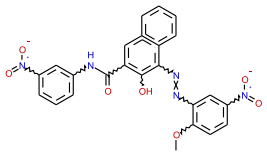
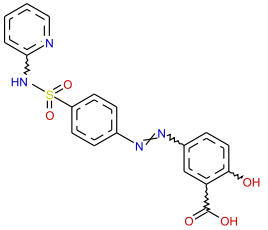
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0336
Unit: g/kg_body_weight
Mahalanobis Distance: 9.12
Mahalanobis Distance p-value: 0.000974

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	RESERPINE	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRIDINE
Structure			
Actual Endpoint (-log C)	6.13118	2.30052	3.375
Predicted Endpoint (-log C)	4.38304	3.55333	2.80292
Distance	0.651	0.771	0.849
Reference	NCI/NTP TR-193	NCI/NTP TR-411	NCI/NTP TR-457

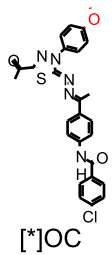
Model Applicability

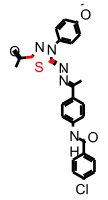
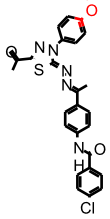
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 5.2806. Training min, max, SD, explained variance: -3.3892, 5.0834, 1.644, 0.0611.

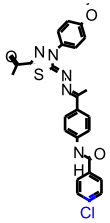
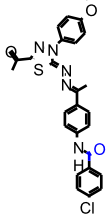
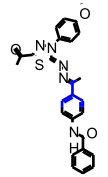
Feature Contribution

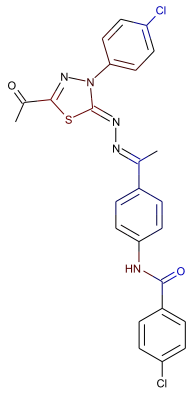
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	 [*]:[c](:[*])Cl	-0.134
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



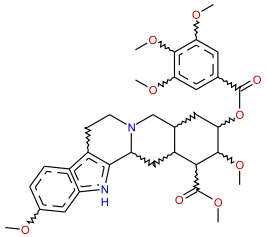
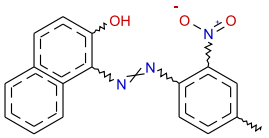
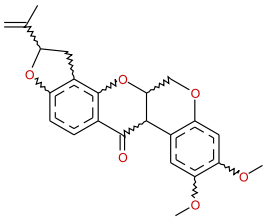
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0498
Unit: g/kg_body_weight
Mahalanobis Distance: 9.27
Mahalanobis Distance p-value: 0.000596

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	RESERPINE	C.I.PIGMENT RED 3	ROTENONE
Structure			
Actual Endpoint (-log C)	6.13118	2.65635	5.06769
Predicted Endpoint (-log C)	4.38304	2.97957	4.11907
Distance	0.776	0.825	0.857
Reference	NCI/NTP TR-193	NCI/NTP TR-407	NCI/NTP TR-320

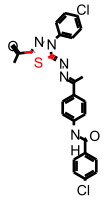
Model Applicability

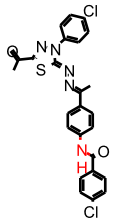
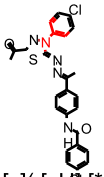
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 5.3889. Training min, max, SD, explained variance: -3.3892, 5.0834, 1.644, 0.0611.

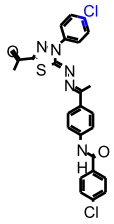
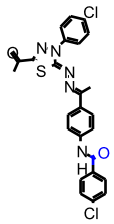
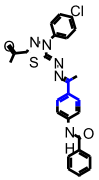
Feature Contribution

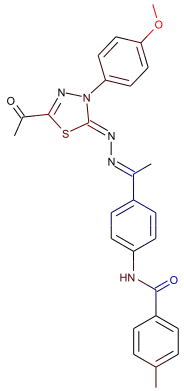
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.0611

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	 [*]:[c](:[*])Cl	-0.134
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



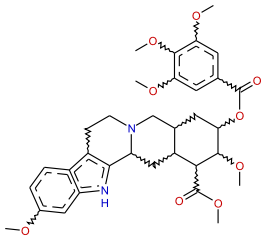
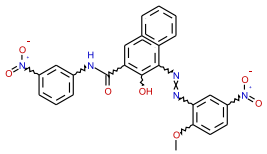
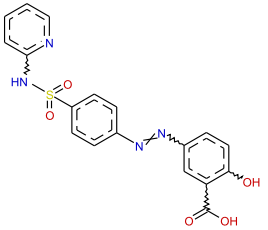
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0228
Unit: g/kg_body_weight
Mahalanobis Distance: 9.35
Mahalanobis Distance p-value: 0.000442

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	RESERPINE	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRIDINE
Structure			
Actual Endpoint (-log C)	6.13118	2.30052	3.375
Predicted Endpoint (-log C)	4.38304	3.55333	2.80292
Distance	0.657	0.754	0.822
Reference	NCI/NTP TR-193	NCI/NTP TR-411	NCI/NTP TR-457

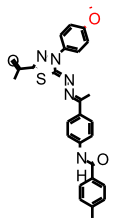
Model Applicability

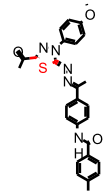
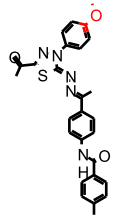
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

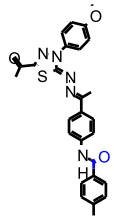
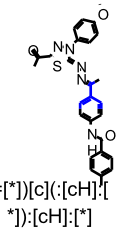
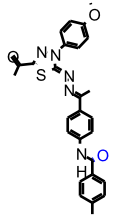
1. All properties and OPS components are within expected ranges.

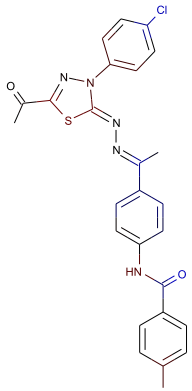
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

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



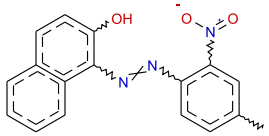
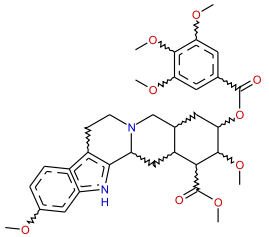
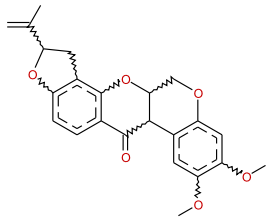
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0511
Unit: g/kg_body_weight
Mahalanobis Distance: 9.87
Mahalanobis Distance p-value: 7.21e-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	C.I.PIGMENT RED 3	RESERPINE	ROTENONE
Structure			
Actual Endpoint (-log C)	2.65635	6.13118	5.06769
Predicted Endpoint (-log C)	2.97957	4.38304	4.11907
Distance	0.778	0.786	0.842
Reference	NCI/NTP TR-407	NCI/NTP TR-193	NCI/NTP TR-320

Model Applicability

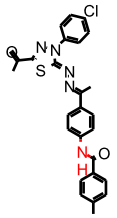
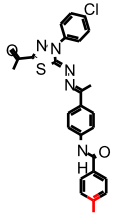
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

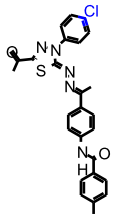
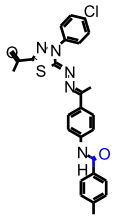
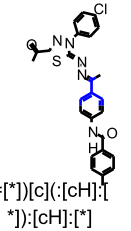
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

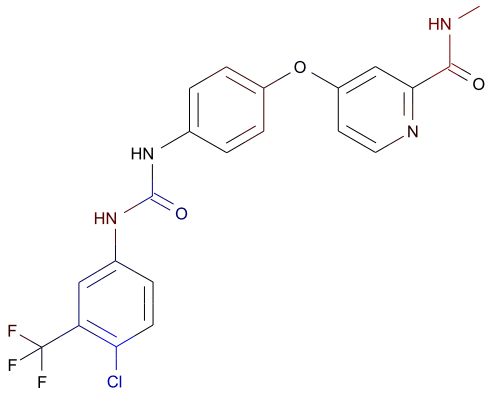
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	136120670	 [*]:[c](:[*])C	0.064

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

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.0885

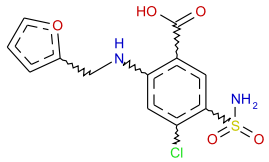
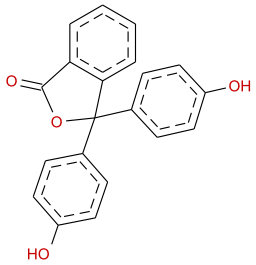
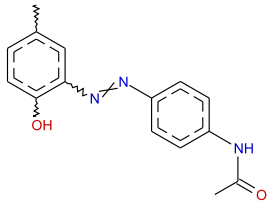
Unit: g/kg_body_weight

Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 1.76e-009

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

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

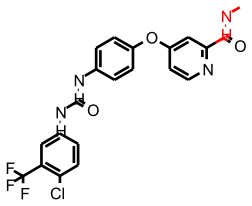
Model Applicability

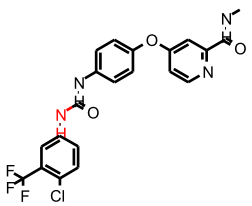
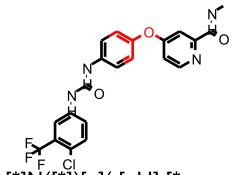
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

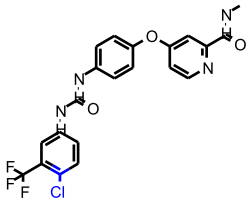
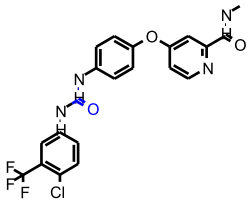
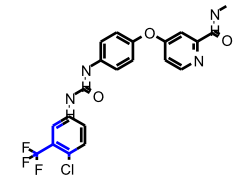
Feature Contribution

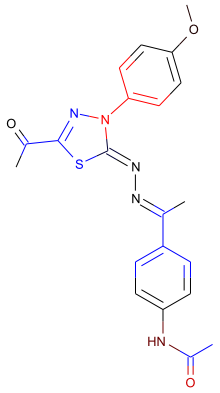
Top features for positive contribution

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

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.0611

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	 [*]:[c](:[*])Cl	-0.134
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.0829



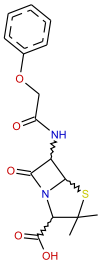
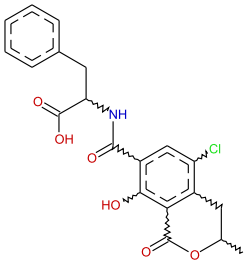
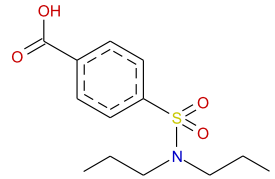
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.000796
Unit: g/kg_body_weight
Mahalanobis Distance: 10.5
Mahalanobis Distance p-value: 1.14e-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	PENICILLIN VK	OCHRATOXIN	PROBENECID
Structure			
Actual Endpoint (-log C)	2.54455	6.28396	2.85333
Predicted Endpoint (-log C)	3.9702	5.12358	2.4258
Distance	0.809	0.882	1.004
Reference	NCI/NTP TR-336	NCI/NTP TR-358	NCI/NTP TR-395

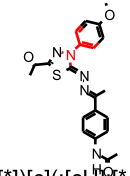
Model Applicability

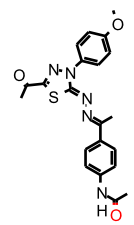
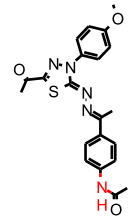
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
2. OPS PC6 out of range. Value: -2.6649. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
3. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
4. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
5. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

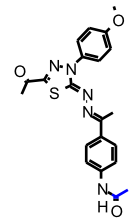
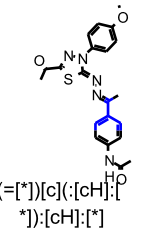
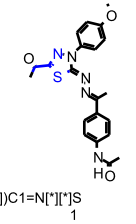
Feature Contribution

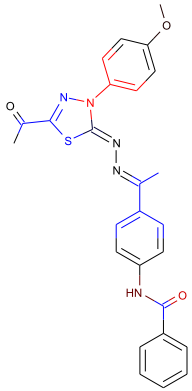
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.672

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.348



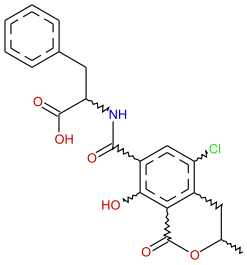
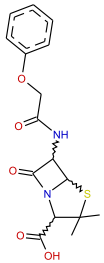
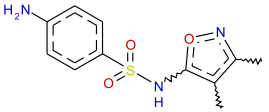
$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.000425
Unit: g/kg_body_weight
Mahalanobis Distance: 12.1
Mahalanobis Distance p-value: 6.28e-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	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.977	1.125	1.214
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

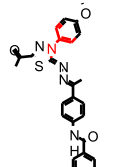
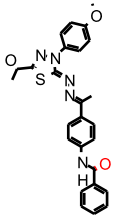
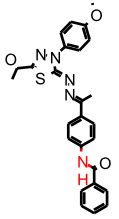
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

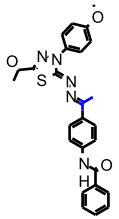
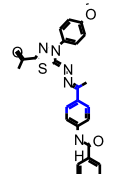
1. Molecular_Weight out of range. Value: 485.56. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC6 out of range. Value: -3.1548. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
5. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
7. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

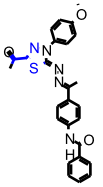
Feature Contribution

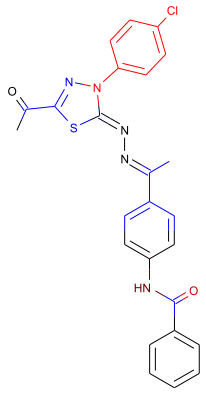
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	1	 [*]=O	0.511
FCFP_2	3	 [*]N[*]	0.104

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348



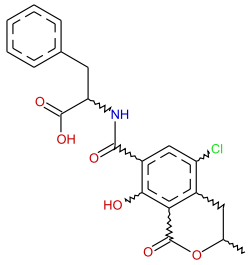
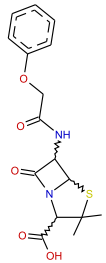
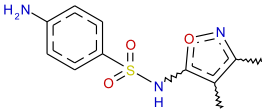
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 9.28e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 9.54
Mahalanobis Distance p-value: 2.35e-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	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.912	1.169	1.199
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

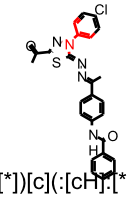
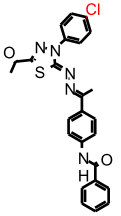
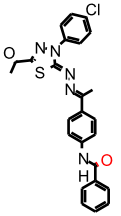
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 489.98. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 7. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC5 out of range. Value: -3.4295. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
5. OPS PC10 out of range. Value: 2.8007. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
6. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
8. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

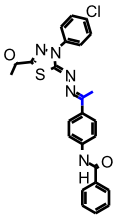
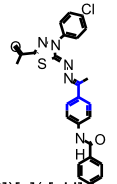
Feature Contribution

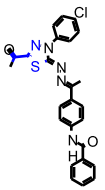
Top features for positive contribution

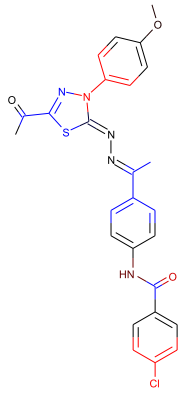
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348



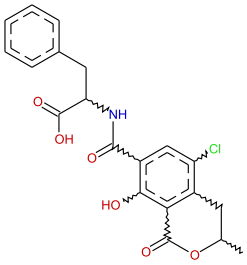
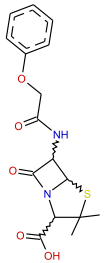
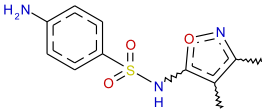
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 3.39e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 3.91e-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	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.978	1.206	1.303
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

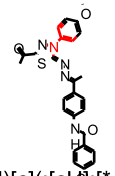
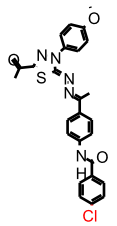
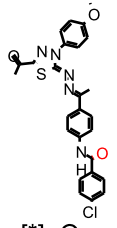
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 520. 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: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC5 out of range. Value: -4.6392. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
5. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
7. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

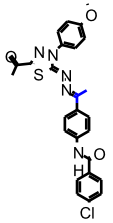
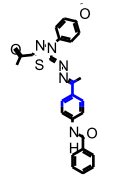
Feature Contribution

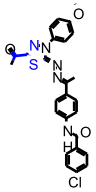
Top features for positive contribution

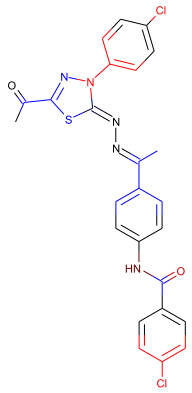
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348



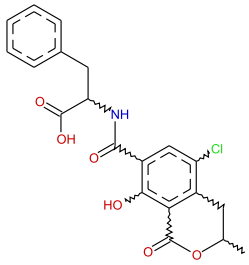
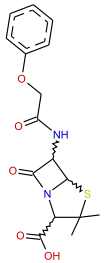
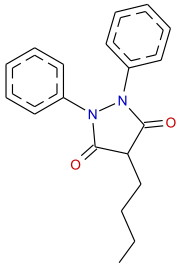
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 6.42e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 9.8
Mahalanobis Distance p-value: 1.08e-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	OCHRATOXIN	PENICILLIN VK	PHENYLBUTAZONE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	3.48909
Predicted Endpoint (-log C)	5.12358	3.9702	3.17333
Distance	0.966	1.241	1.250
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-367

Model Applicability

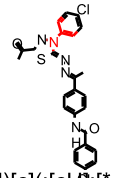
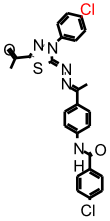
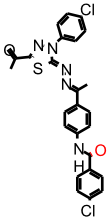
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 524.42. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 7. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC5 out of range. Value: -3.4476. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
5. OPS PC10 out of range. Value: 2.6731. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
6. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
8. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

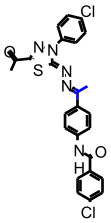
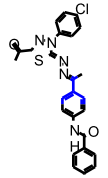
Feature Contribution

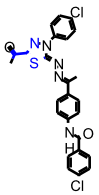
Top features for positive contribution

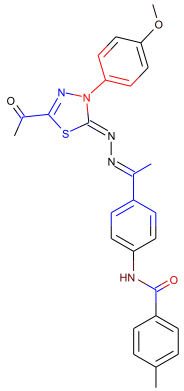
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348



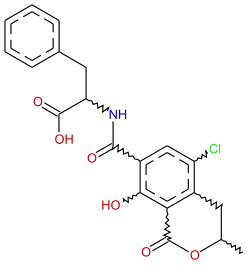
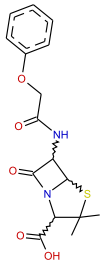
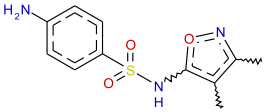
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.000324
Unit: g/kg_body_weight
Mahalanobis Distance: 12.4
Mahalanobis Distance p-value: 2.5e-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	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	1.005	1.169	1.237
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

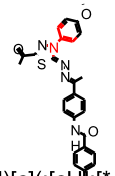
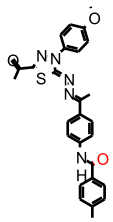
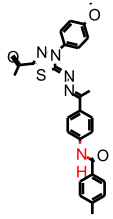
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 499.58. 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: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC6 out of range. Value: -3.2186. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
5. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
7. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

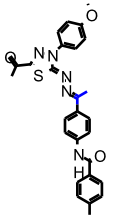
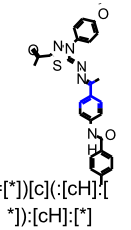
Feature Contribution

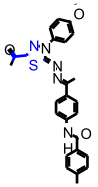
Top features for positive contribution

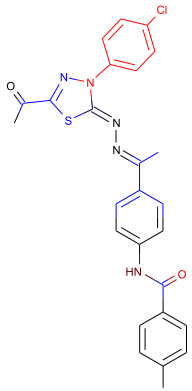
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	1	 [*]=O	0.511
FCFP_2	3	 [*]N[*]	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348



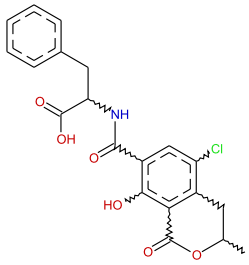
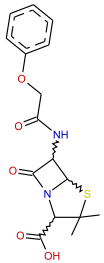
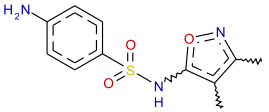
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 7.05e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 9.76
Mahalanobis Distance p-value: 1.21e-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	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.949	1.217	1.228
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

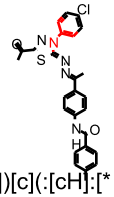
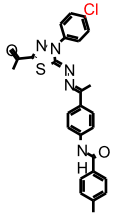
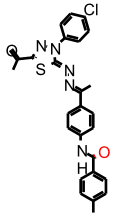
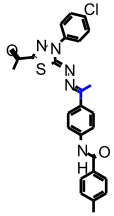
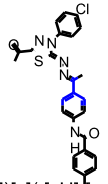
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

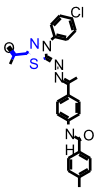
1. Molecular_Weight out of range. Value: 504. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 7. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC5 out of range. Value: -3.4178. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
5. OPS PC10 out of range. Value: 2.7196. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
6. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
8. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

Feature Contribution

Top features for positive contribution

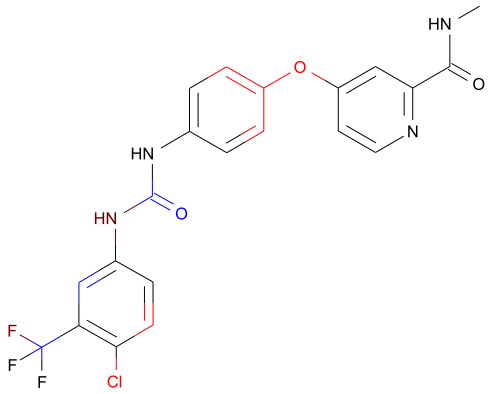
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.000918

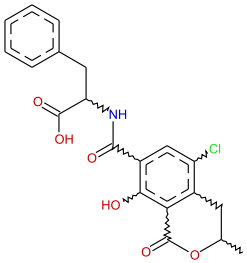
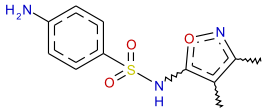
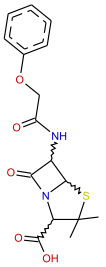
Unit: g/kg_body_weight

Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 4.69e-009

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

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

Model Applicability

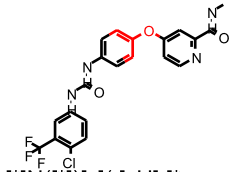
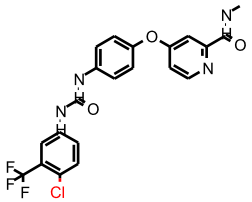
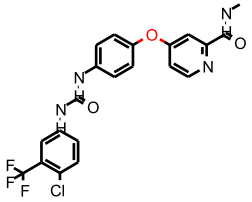
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 464.82. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
3. OPS PC5 out of range. Value: -3.5737. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
4. OPS PC7 out of range. Value: -3.8342. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
5. Unknown FCFP_2 feature: 1499521844: [*]NC(=O)N[*]
6. Unknown FCFP_2 feature: -1029533685: [*]:[c](:[*])C(F)(F)F
7. Unknown FCFP_2 feature: 136686699: [*]NC

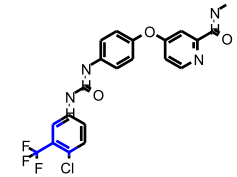
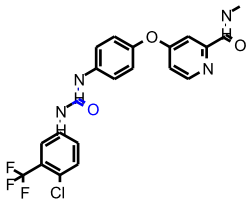
Feature Contribution

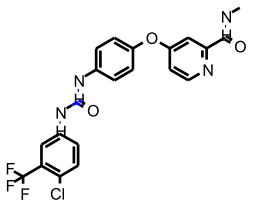
Top features for positive contribution

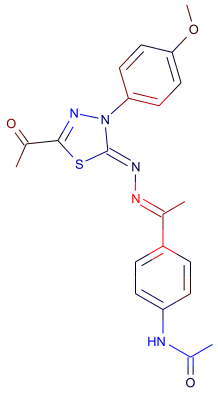
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307

FCFP_2	0	 [*]C(=[*])[*]	-0.29



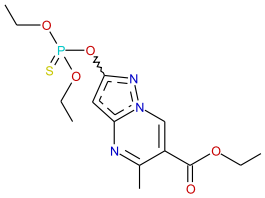
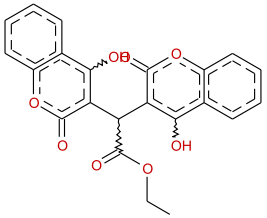
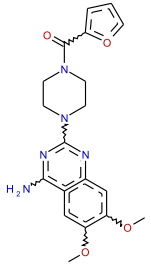
$C_{21}H_{21}N_5O_3S$
Molecular Weight: 423.48813
ALogP: 3.064
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.308
Unit: g/kg_body_weight
Mahalanobis Distance: 23.6
Mahalanobis Distance p-value: 1.39e-023

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	PYRAZOPHOS	ETHYL-bis-COUMACETATE	PRASOZIN .HCl (HCl STRIPPED)
Structure			
Actual Endpoint (-log C)	3.234	2.687	2.294
Predicted Endpoint (-log C)	3.35181	2.7054	3.00765
Distance	0.644	0.647	0.654
Reference	KHZDAN 23;538;80	FEPRA7 10;303;51	NIIRDN 6;688;82

Model Applicability

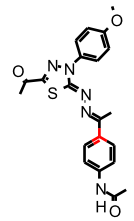
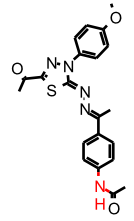
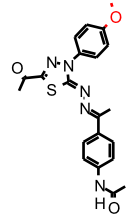
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]

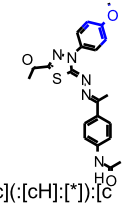
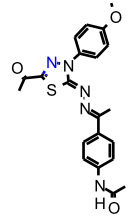
Feature Contribution

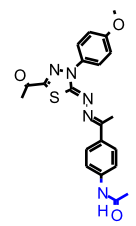
Top features for positive contribution

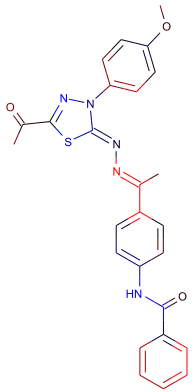
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	136627117	 [*]OC	0.17

Top Features for negative contribution

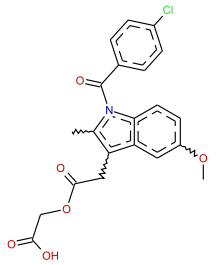
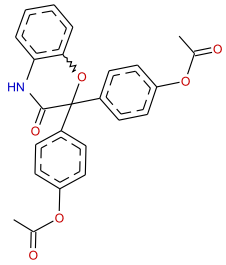
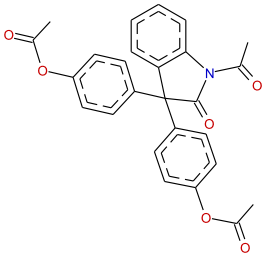
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	-0.257
ECFP_6	655739385	 [*]N=[*]	-0.239

FCFP_6	566058135	 [*]NC(=O)C	-0.216



$C_{26}H_{23}N_5O_3S$
Molecular Weight: 485.55751
ALogP: 4.728
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Structural Similar Compounds

Name	ACEMETACIN	bis-OXATIN ACETATE	1-ACETYL-3;3-bis-[4-(ACETYLOXY)PHENYL]-1;3-DIHYDRO-2H-INDOL-2-ONE
Structure			
Actual Endpoint (-log C)	4.235	1.717	2.948
Predicted Endpoint (-log C)	3.39415	2.40947	2.6866
Distance	0.651	0.663	0.698
Reference	ARZNAD 30;1398;80	NIIRDN 6;609;82	JAPMA8 42;468;53

Model Prediction

Prediction: 0.211
Unit: g/kg_body_weight
Mahalanobis Distance: 23.5
Mahalanobis Distance p-value: 3.37e-023

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Model Applicability

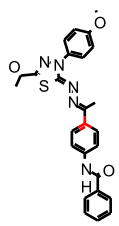
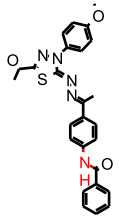
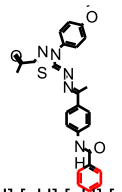
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]

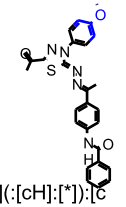
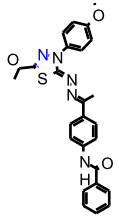
Feature Contribution

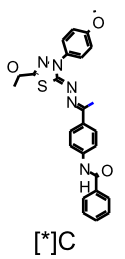
Top features for positive contribution

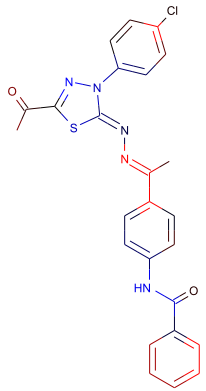
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	-0.257
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201



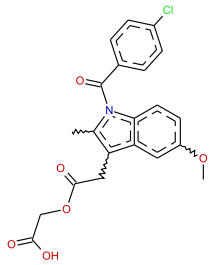
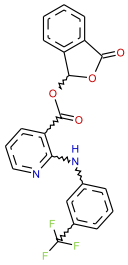
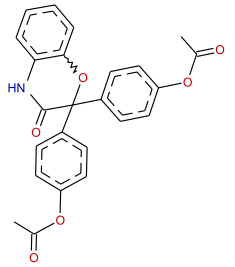
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.343
Unit: g/kg_body_weight
Mahalanobis Distance: 23.7
Mahalanobis Distance p-value: 3.46e-024

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACEMETACIN	TALNIFLUMATE	bis-OXATIN ACETATE
Structure			
Actual Endpoint (-log C)	4.235	1.538	1.717
Predicted Endpoint (-log C)	3.39415	2.82541	2.40947
Distance	0.630	0.639	0.662
Reference	ARZNAD 30;1398;80	FRPSAX 36;372;81	NIIRDN 6;609;82

Model Applicability

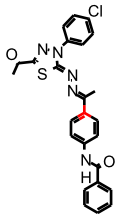
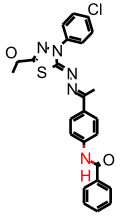
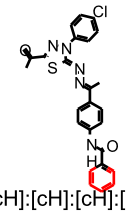
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
11. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

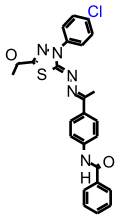
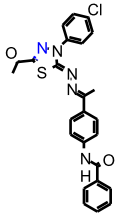
Feature Contribution

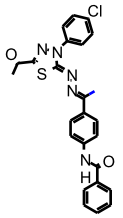
Top features for positive contribution

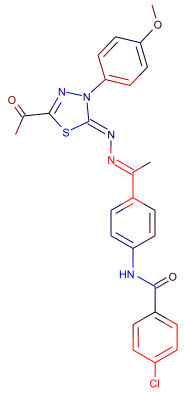
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-817402818	 [*]Cl	-0.263
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201



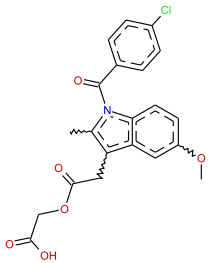
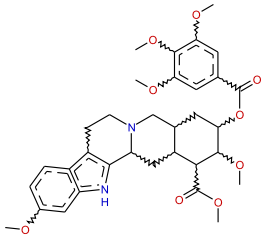
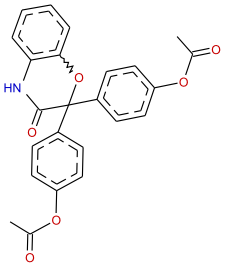
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.392
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.148
Unit: g/kg_body_weight
Mahalanobis Distance: 23.8
Mahalanobis Distance p-value: 1.55e-024

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACEMETACIN	RESERPINE	bis-OXATIN ACETATE
Structure			
Actual Endpoint (-log C)	4.235	3.161	1.717
Predicted Endpoint (-log C)	3.39415	2.72801	2.40947
Distance	0.678	0.704	0.736
Reference	ARZNAD 30;1398;80	PSSCBG 11;555;80	NIIRDN 6;609;82

Model Applicability

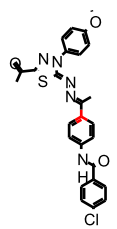
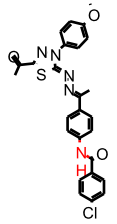
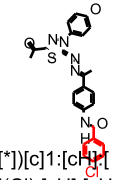
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
11. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

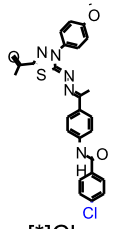
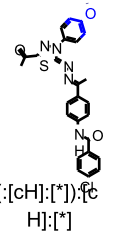
Feature Contribution

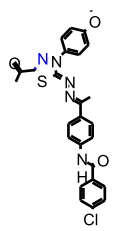
Top features for positive contribution

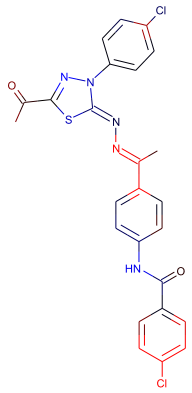
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	-149636017	 [*]C(=[*])[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.193

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-817402818	 [*]Cl	-0.263
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	-0.257

ECFP_6	655739385	 [*]N=[*]	-0.239



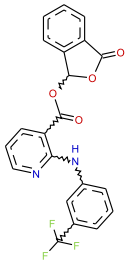
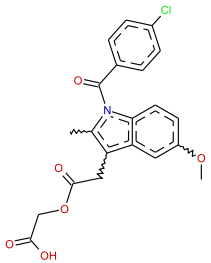
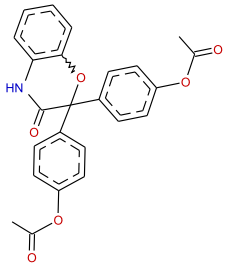
$C_{25}H_{19}Cl_2N_5O_2S$
Molecular Weight: 524.42166
ALogP: 6.073
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.292
Unit: g/kg_body_weight
Mahalanobis Distance: 23.8
Mahalanobis Distance p-value: 7.26e-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 non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	TALNIFLUMATE	ACEMETACIN	bis-OXATIN ACETATE
Structure			
Actual Endpoint (-log C)	1.538	4.235	1.717
Predicted Endpoint (-log C)	2.82541	3.39415	2.40947
Distance	0.699	0.704	0.748
Reference	FRPSAX 36;372;81	ARZNAD 30;1398;80	NIIRDN 6;609;82

Model Applicability

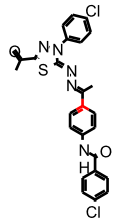
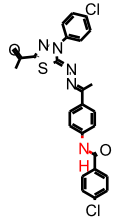
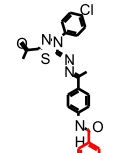
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
11. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

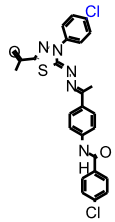
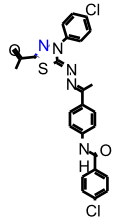
Feature Contribution

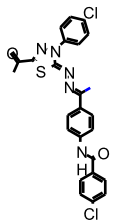
Top features for positive contribution

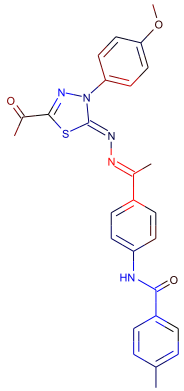
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	-149636017	 [*]C(=[*])[c]1:[cH]:[cH]:[c]:[cH]:[c](Cl):[cH]:[cH]:1	0.193

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-817402818	 [*]Cl	-0.263
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201



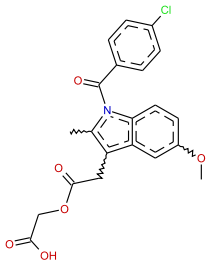
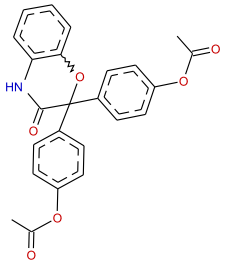
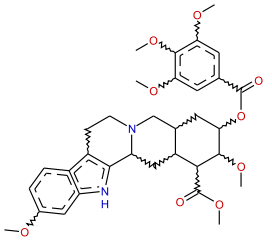
$C_{27}H_{25}N_5O_3S$
Molecular Weight: 499.58409
ALogP: 5.214
Rotatable Bonds: 7
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.358
Unit: g/kg_body_weight
Mahalanobis Distance: 23.6
Mahalanobis Distance p-value: 9.56e-024

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	ACEMETACIN	bis-OXATIN ACETATE	RESERPINE
Structure			
Actual Endpoint (-log C)	4.235	1.717	3.161
Predicted Endpoint (-log C)	3.39415	2.40947	2.72801
Distance	0.678	0.703	0.714
Reference	ARZNAD 30;1398;80	NIIRDN 6;609;82	PSSCBG 11;555;80

Model Applicability

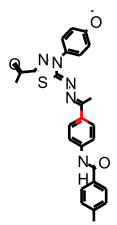
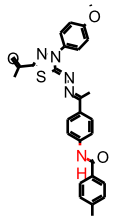
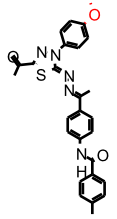
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]

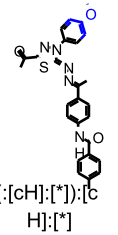
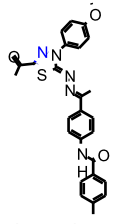
Feature Contribution

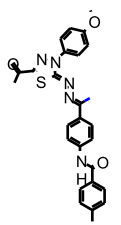
Top features for positive contribution

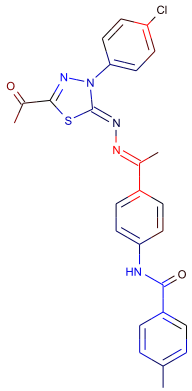
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	136627117	 [*]OC	0.17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[c H]:[*]	-0.257
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201



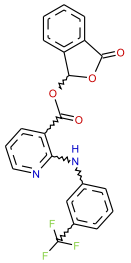
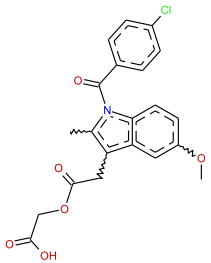
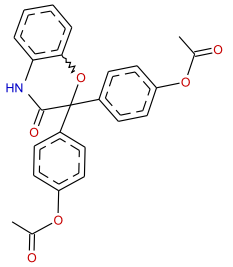
$C_{26}H_{22}ClN_5O_2S$
Molecular Weight: 504.00318
ALogP: 5.895
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.581
Unit: g/kg_body_weight
Mahalanobis Distance: 23.8
Mahalanobis Distance p-value: 1.2e-024

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	TALNIFLUMATE	ACEMETACIN	bis-OXATIN ACETATE
Structure			
Actual Endpoint (-log C)	1.538	4.235	1.717
Predicted Endpoint (-log C)	2.82541	3.39415	2.40947
Distance	0.671	0.672	0.717
Reference	FRPSAX 36;372;81	ARZNAD 30;1398;80	NIIRDN 6;609;82

Model Applicability

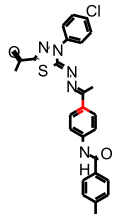
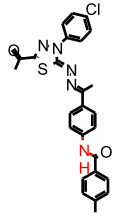
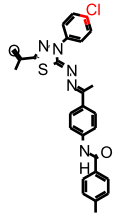
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
11. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

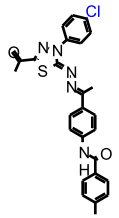
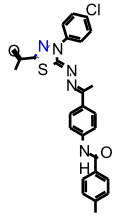
Feature Contribution

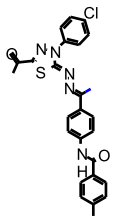
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
ECFP_6	99947387	 [*]:[c](:[*])Cl	0.181

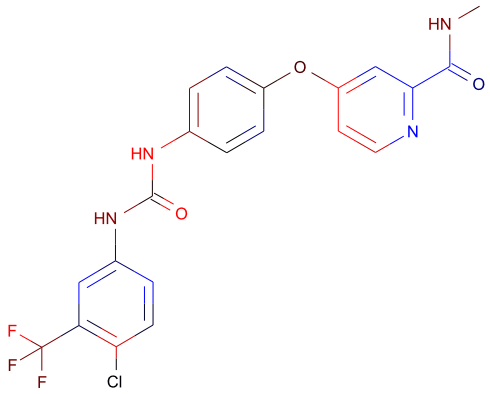
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-817402818	 [*]Cl	-0.263
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201

Sorafenib

TOPKAT_Rat_Oral_LD50



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

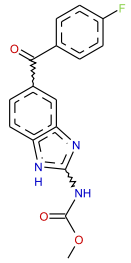
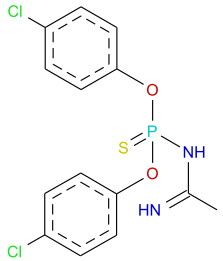
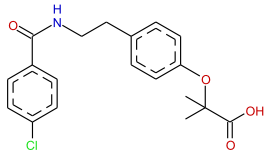
ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Structural Similar Compounds

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

Model Prediction

Prediction: 0.823

Unit: g/kg_body_weight

Mahalanobis Distance: 21

Mahalanobis Distance p-value: 1.93e-012

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Model Applicability

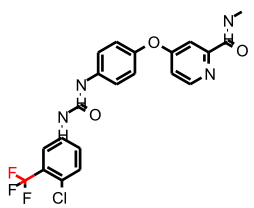
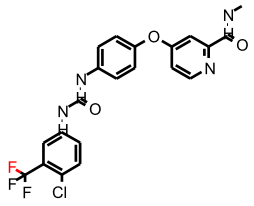
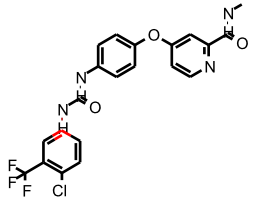
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

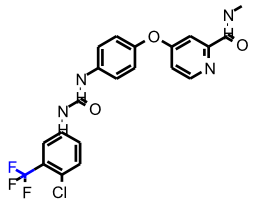
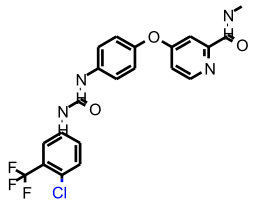
Feature Contribution

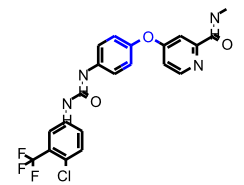
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	71953198	 [*]C([*])([*])F	0.392
ECFP_6	-1046436026	 [*]F	0.349
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	226796801	 [*]C([*])([*])F	-0.32
ECFP_6	-817402818	 [*]Cl	-0.263

ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[c H]:[*]	-0.257