Design, Synthesis, and Evaluation of Novel Thiadiazole Derivatives as Potent VEGFR-2 Inhibitors: A Comprehensive *In vitro* and *In silico* Study

Ibrahim H. Eissa[a*], Walid E. Elgammal[b], Hazem A. Mahdy[a], Susi Zara [c], Simone Carradori [c], Dalal Z. Husein[d], Maymounah N. Alharthi[e], Ibrahim M. Ibrahim[f], Eslam B. Elkaeed[g*], Hazem Elkady[a], Ahmed M. Metwaly[h*]

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

[b]Department of Chemistry, Faculty of Science, Al-Azhar University, Nasr City, Cairo, Egypt

[c]Department of Pharmacy, "G. d'Annunzio" University of Chieti-Pescara, Chieti, 66100, Italy

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

[e] Department of Chemistry, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

[f] Biophysics Department, Faculty of Science, Cairo University, Giza 12613, Egypt

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

[h] Pharmacognosy and Medicinal Plants Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, 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

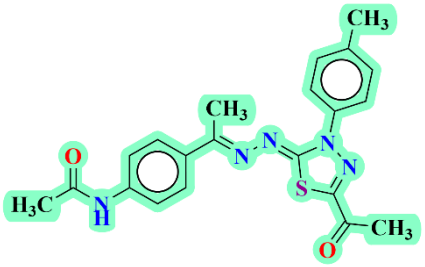
Lab code	Paper code
WA-6	**16**
PH-5	**20a**
PH-6	**20b**
PH-10	**21a**
WC-10	**21b**

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.

N-(4-((E)-1-(((Z)-5-Acetyl-3-(p-tolyl)-1,3,4-thiadiazol-2(3H)-ylidene)hydrazono)ethyl) phenyl) acetamide 16

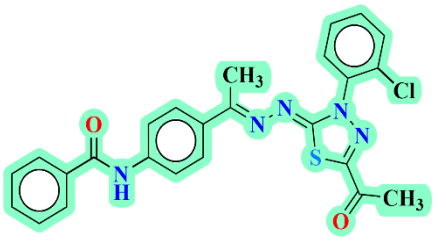


Yellow powder (yield, 74%); m.p. = 201- 202 °C. FT-IR (v max, cm^{-1}): 3307 (NH), 3080, 3032, 3009 (C-H aromatic), 2922, 2863 (C-H aliphatic), 1687, 1666 (C=O); ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.15 (s, 1H, NH), 7.94 – 7.92 (m, 2H, Ar-H), 7.85 – 7.83 (m, 2H, Ar-H), 7.67 (d, *J* = 8.7 Hz, 2H, Ar-H), 7.40 (d, *J* = 8.4 Hz, 2H, Ar-H), 2.60 (s, 3H, CH_3), 2.40 (s, 3H, CH_3), 2.39 (s, 3H, CH_3), 2.09 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 190.2, 169.0, 164.5, 160.3, 150.8, 141.4, 137.2, 136.9, 132.2, 130.0, 127.6, 122.5, 119.0, 25.5, 24.6, 21.1, 15.6; Anal. Calcd. for $C_{21}H_{21}N_5O_2S$ (407.49): C, 61.90; H, 5.19; N, 17.19. Found: C, 61.79; H, 5.46; N, 17.42%.

3.1.1. ***General procedure for the synthesis of derivatives 20a,b and 21a,b.***

A solution of intermediates **19a,b** (1 molar) in ethanol (20 mL) was combined with triethylamine (1 molar) and the corresponding hydrazonoyl chlorides 5a,b and 8 (1 molar). The mixtures were heated for a period of 6 hours. The resulting solids were collected and subjected to purification with hexane, resulting in the desired products **20a,b** and **21a,b**, respectively.

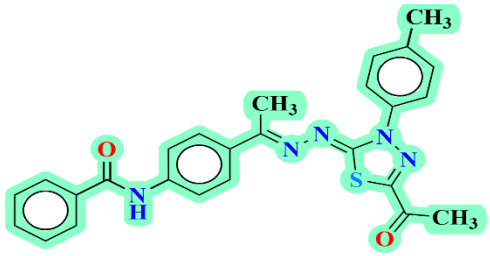
N-(4-((E)-1-(((Z)-5-Acetyl-3-(2-chlorophenyl)-1,3,4-thiadiazol-2(3H)-ylidene)hydrazono) ethyl) phenyl)benzamide 20a



Pale yellow powder (yield, 70%); m.p. = 190-192 °C. FT-IR (v max, cm^{-1}): 3353 (NH), 3062, 3001 (C-H aromatic), 2919 (C-H aliphatic), 1675, 1655 (C=O);^{1}H NMR (400 MHz, DMSO-d_6) δ 10.45 (s, 1H, NH), 8.01 – 7.98 (m, 4H, Ar-H), 7.89 (m, 3H, Ar-H), 7.67-7.61 (m, 3H, Ar-H), 7.59-7.54 (m, 3H, Ar-H), 2.54 (s, 3H, CH_3), 2.23 (s, 3H, CH_3); ^{13}C NMR (126 MHz, DMSO-d_6) δ 190.1, 166.2, 164.2, 160.9, 151.5, 141.4, 138.3, 135.4, 132.9, 132.1, 131.6, 129.6, 128.9, 128.2, 127.5, 123.9,

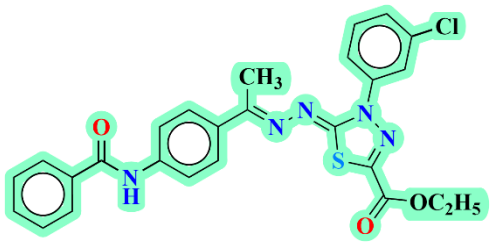
120.5, 25.5, 15.8; Anal. Calcd. for $C_{25}H_{20}ClN_5O_2S$ (489.98): C, 61.28; H, 4.11; N, 14.29. Found: C, 61.39; H, 4.24; N, 14.53%.

N-(4-((E)-1-(((Z)-5-Acetyl-3-(p-tolyl)-1,3,4-thiadiazol-2(3H)-ylidene)hydrazono) ethyl) phenyl) benzamide 20b



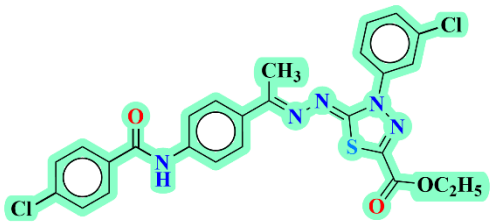
Yellow powder (yield, 78%); m.p. = 185- 187 ^oC. FT-IR (v max, cm^{-1}): 3351, 3110 (NH), 3057 (C-H aromatic), 2986, 2916 (C-H aliphatic), 1746, 1706, 1655 (C=O); ^1H NMR (400 MHz, DMSO-d_6) δ 10.46 (s, 1H, NH), 8.00 – 7.91 (m, 8H, Ar-H), 7.65 (d, J = 6.9 Hz, 1H, Ar-H), 7.57 (t, J = 7.4 Hz, 2H, Ar-H), 7.41 (d, J = 8.1 Hz, 2H, Ar-H), 2.61 (s, 3H, CH_3), 2.43 (s, 3H, CH_3), 2.40 (s, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ 190.2, 166.2, 164.7, 160.3, 150.9, 141.3, 137.3, 136.9, 135.2, 132.8, 132.2, 130.0, 128.9, 128.2, 127.4, 122.6, 120.3, 25.5, 21.1, 15.7; Anal. Calcd. for $C_{26}H_{23}N_5O_2S$ (469.56): C, 66.51; H, 4.94; N, 14.91. Found: C, 66.32; H, 5.06; N, 15.07%.

Ethyl (Z)-5-(((E)-1-(4-Benzamidophenyl)ethylidene)hydrazono)-4-(3-chlorophenyl)-4,5-dihydro-1,3,4-thiadiazole-2-carboxylate 21a



Yellowish white powder (yield, 73%); m.p. = 230-232 ^oC. FT-IR (v max, cm^{-1}): 3285 (NH), 3080, 3027 (C-H aromatic), 2981, 2955, 2918 (C-H aliphatic), 1715, 1655 (C=O); ^1H NMR (400 MHz, DMSO-d_6) δ 10.47 (s, 1H, NH), 8.22 (t, J = 2.1 Hz, 1H, Ar-H), 8.03-7.98 (m, 3H, Ar-H), 7.91-7.88 (m, 4H, Ar-H), 7.65 – 7.55 (m, 4H, Ar-H), 7.49 – 7.45 (m, 1H, Ar-H), 4.43 (q, J = 7.1 Hz, 2H, CH_2), 2.46 (s, 3H, CH_3), 1.36 (t, J = 7.1 Hz, 3H, CH_3); ^{13}C NMR (101 MHz, DMSO-d_6) δ ^{13}C NMR (101 MHz, DMSO-d_6) δ 166.2, 164.0, 160.9, 158.4, 143.9, 141.4, 140.4, 135.2, 133.8, 132.6, 132.2, 131.4, 128.9, 128.2, 127.5, 127.1, 121.6, 120.4, 120.3, 63.4, 15.8, 14.5; Anal. Calcd. for $C_{26}H_{22}ClN_5O_3S$ (520.00): C, 60.05; H, 4.26; N, 13.47. Found: C, 60.29; H, 4.37; N, 13.60%.

Ethyl (Z)-5-(((E)-1-(4-(4-Chlorobenzamido)phenyl)ethylidene)hydrazono)-4-(3-chlorophenyl)-4,5-dihydro-1,3,4-thiadiazole-2-carboxylate 21b



Yellow powder (yield, 75%); m.p. = 203- 205 ^{o}C. FT-IR (v max, cm^{-1}): 3310 (NH), 3077, 3000 (C-H aromatic), 2979, 2959, 2917 (C-H aliphatic), 1740, 1651 (C=O); ^{1}H NMR (400 MHz, DMSO-d_6) δ ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.51 (s, 1H, NH), 8.04 – 8.01 (m, 2H, Ar-H), 7.92-7.86 (m, 6H, Ar-H), 7.66 – 7.62 (m, 2H, Ar-H), 7.38 (d, *J* = 8.2 Hz, 2H, Ar-H), 4.42 (q, *J* = 7.1 Hz, 2H), 2.42 (s, 3H, CH_3), 2.39 (s, 3H, CH_3), 1.35 (t, *J* = 7.1 Hz, 3H, 2.42 (s, 3H, CH_3)); ^{13}C NMR (101 MHz, DMSO-d_6) δ 164.5, 160.1, 158.6, 142.9, 141.0, 137.2, 136.8, 133.9, 133.0, 130.2, 130.0, 129.0, 127.4, 122.6, 120.4, 63.2, 21.1, 15.7, 14.5; Anal. Calcd. for $C_{26}H_{21}Cl_2N_5O_3S$ (554.45): C, 56.32; H, 3.82; N, 12.63. Found: C, 56.57; H, 3.98; N, 12.67%.

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.

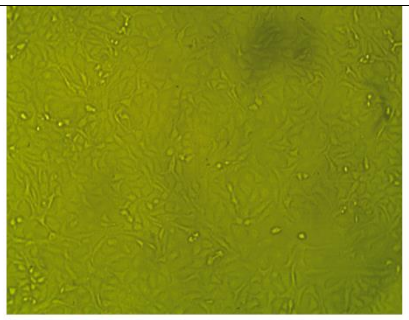
S.2.1.2. Morphological assay

❖ 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.

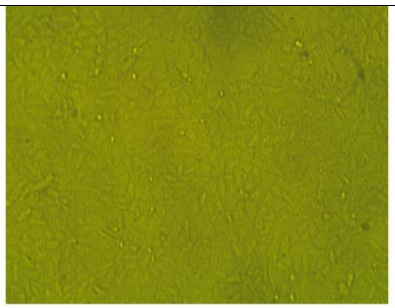
❖ Necrotic cells: nuclear swelling, chromatin flocculation, loss of nuclear basophilia

❖ 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

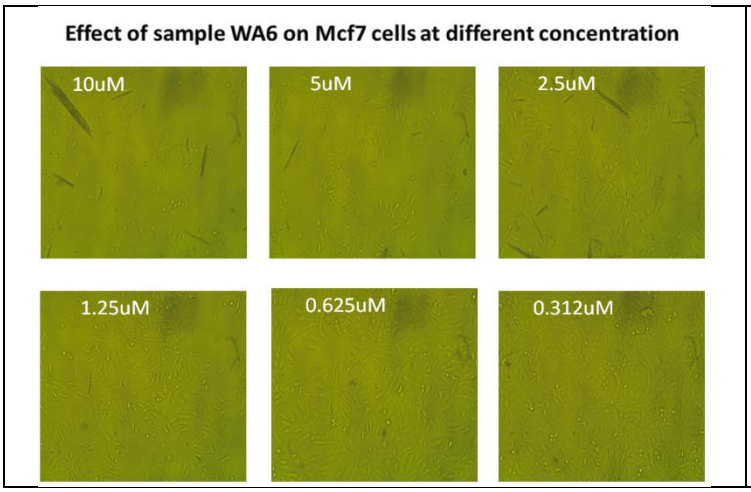
Organism : *Homo sapiens*, human
Tissue : liver
Cell Type : epithelial
Culture Properties : adherent
Disease : hepatocellular carcinoma
ATCC : HB-8065

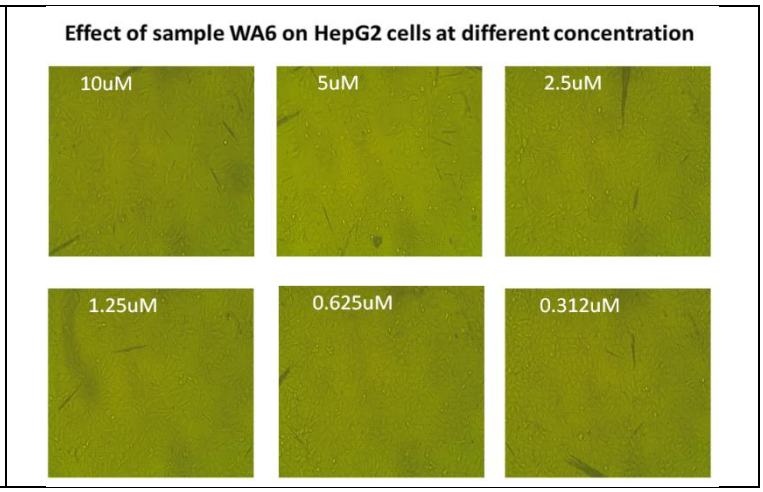
***** Cytotoxicity of WA-6 (compound 16) against MCF-7 and HepG2**

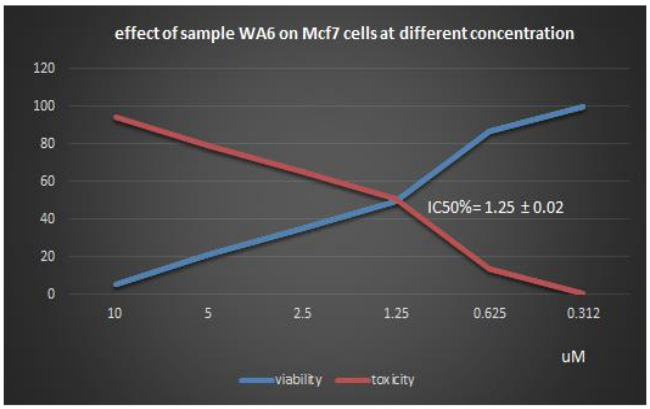
ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
Mcf7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
WA6	10	0.046	0.032	0.047	0.041667	0.004842	5.39025442	94.60974558	1.25 ± 0.02
	5	0.173	0.156	0.162	0.163667	0.004978	21.17291936	78.82708064	
	2.5	0.263	0.269	0.271	0.267667	0.002404	34.62699439	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	

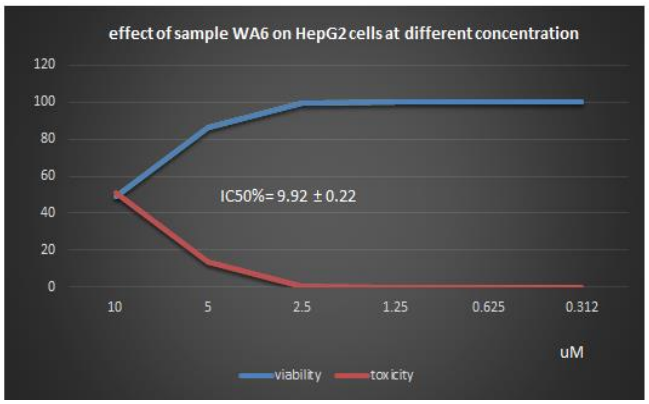
	0.312	0.771	0.764	0.778	0.771	0.004041	99.74126779	0.258732212	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
WA6	10	0.348	0.371	0.356	0.358333	0.006741	48.81925522	51.18074478	9.92 ± 0.22
	5	0.643	0.618	0.632	0.631	0.007234	85.96730245	14.03269755	
	2.5	0.735	0.722	0.732	0.729667	0.00393	99.40962761	0.590372389	
	1.25	0.731	0.729	0.738	0.732667	0.002728	99.81834696	0.181653043	
	0.625	0.733	0.725	0.74	0.732667	0.004333	99.81834696	0.181653043	
	0.312	0.743	0.724	0.734	0.733667	0.005487	99.95458674	0.045413261	







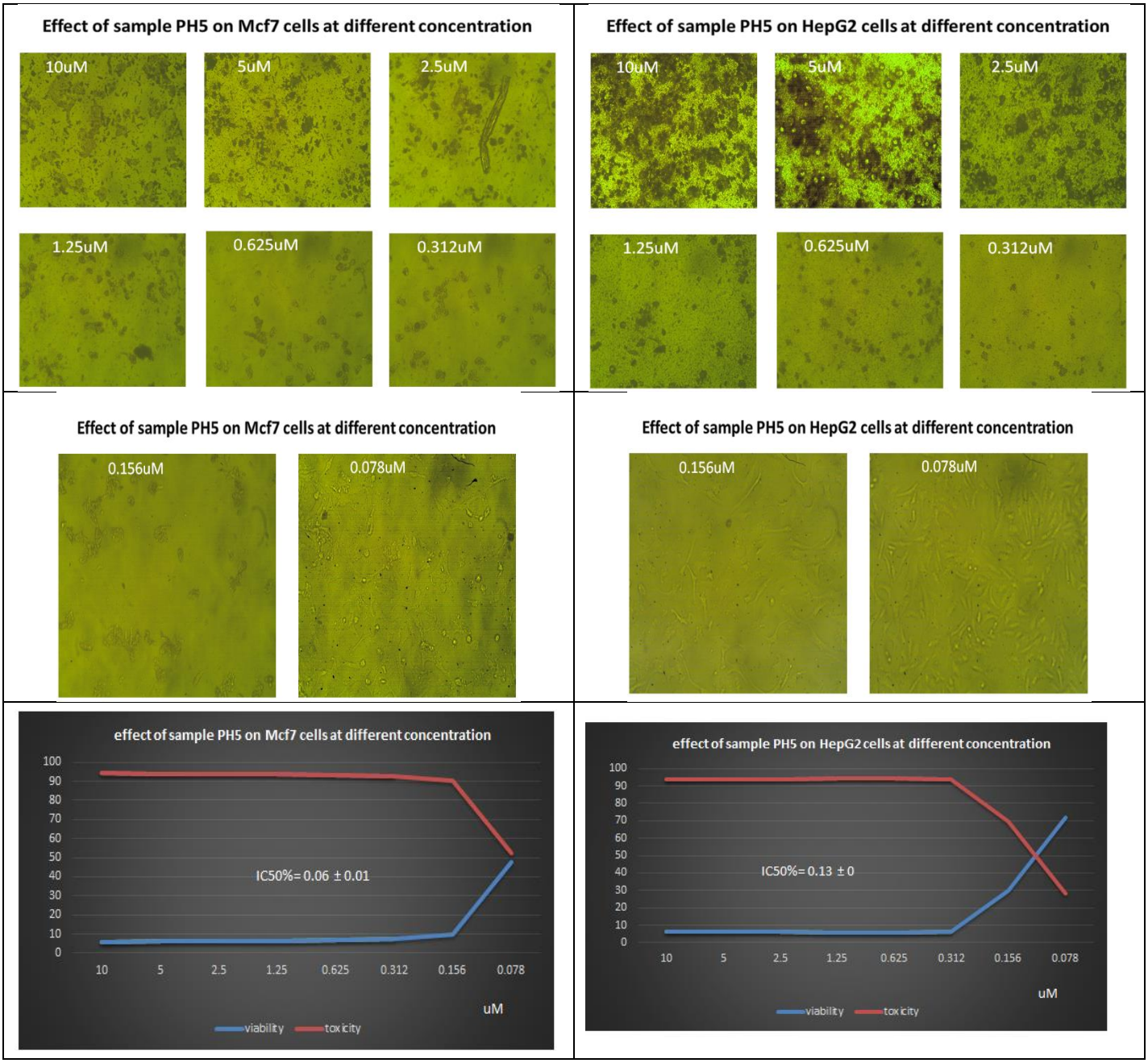


***** Cytotoxicity of PH-5 (compound 20a) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
Mcf7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
PH5	10	0.042	0.048	0.044	0.044667	0.001764	5.778352738	94.22164726	0.06 ± 0.01
	5	0.047	0.05	0.046	0.047667	0.001202	6.166451056	93.83354894	
	2.5	0.044	0.043	0.057	0.048	0.004509	6.209573092	93.79042691	
	1.25	0.051	0.047	0.049	0.049	0.001155	6.338939198	93.6610608	
	0.625	0.06	0.045	0.048	0.051	0.004583	6.59767141	93.40232859	
	0.312	0.055	0.061	0.058	0.058	0.001732	7.503234153	92.49676585	
	0.156	0.073	0.069	0.081	0.074333	0.003528	9.616213885	90.38378611	
	0.078	0.385	0.377	0.34	0.367333	0.01386	47.52048297	52.47951703	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
PH5	10	0.044	0.045	0.046	0.045	0.000577	6.130790191	93.86920981	0.13 ± 0
	5	0.044	0.044	0.046	0.044667	0.000667	6.08537693	93.91462307	
	2.5	0.043	0.045	0.046	0.044667	0.000882	6.08537693	93.91462307	
	1.25	0.042	0.038	0.047	0.042333	0.002603	5.767484105	94.23251589	
	0.625	0.042	0.044	0.045	0.043667	0.000882	5.949137148	94.05086285	

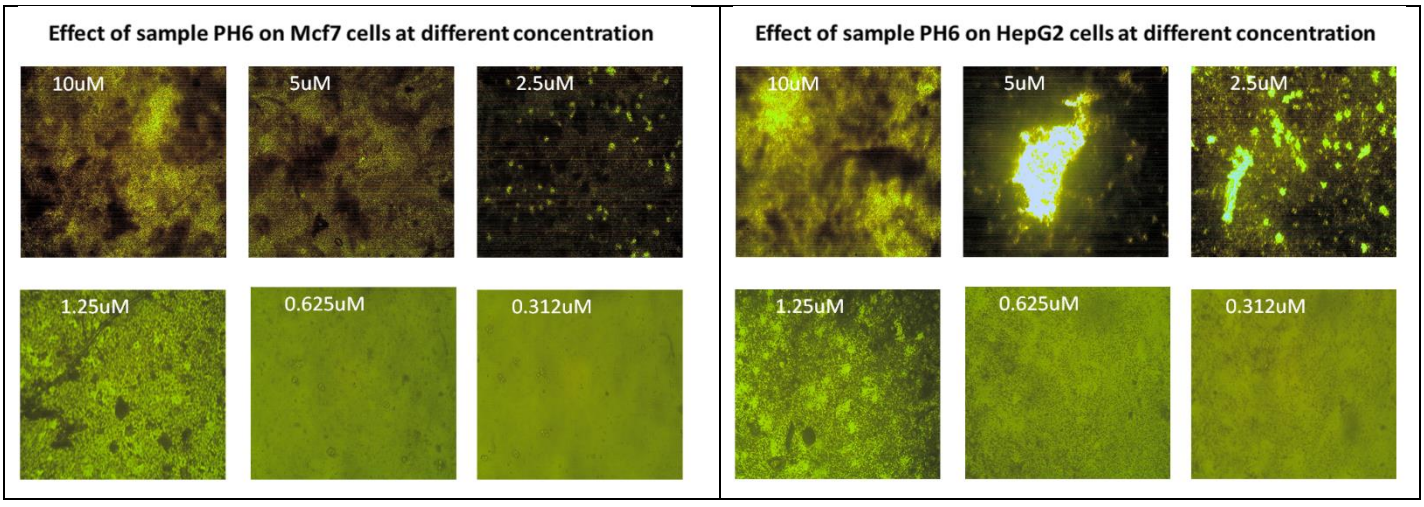
	0.312	0.045	0.047	0.043	0.045	0.001155	6.130790191	93.86920981	
	0.156	0.236	0.22	0.208	0.221333	0.00811	30.15440509	69.84559491	
	0.078	0.538	0.512	0.529	0.526333	0.007623	71.7075386	28.2924614	

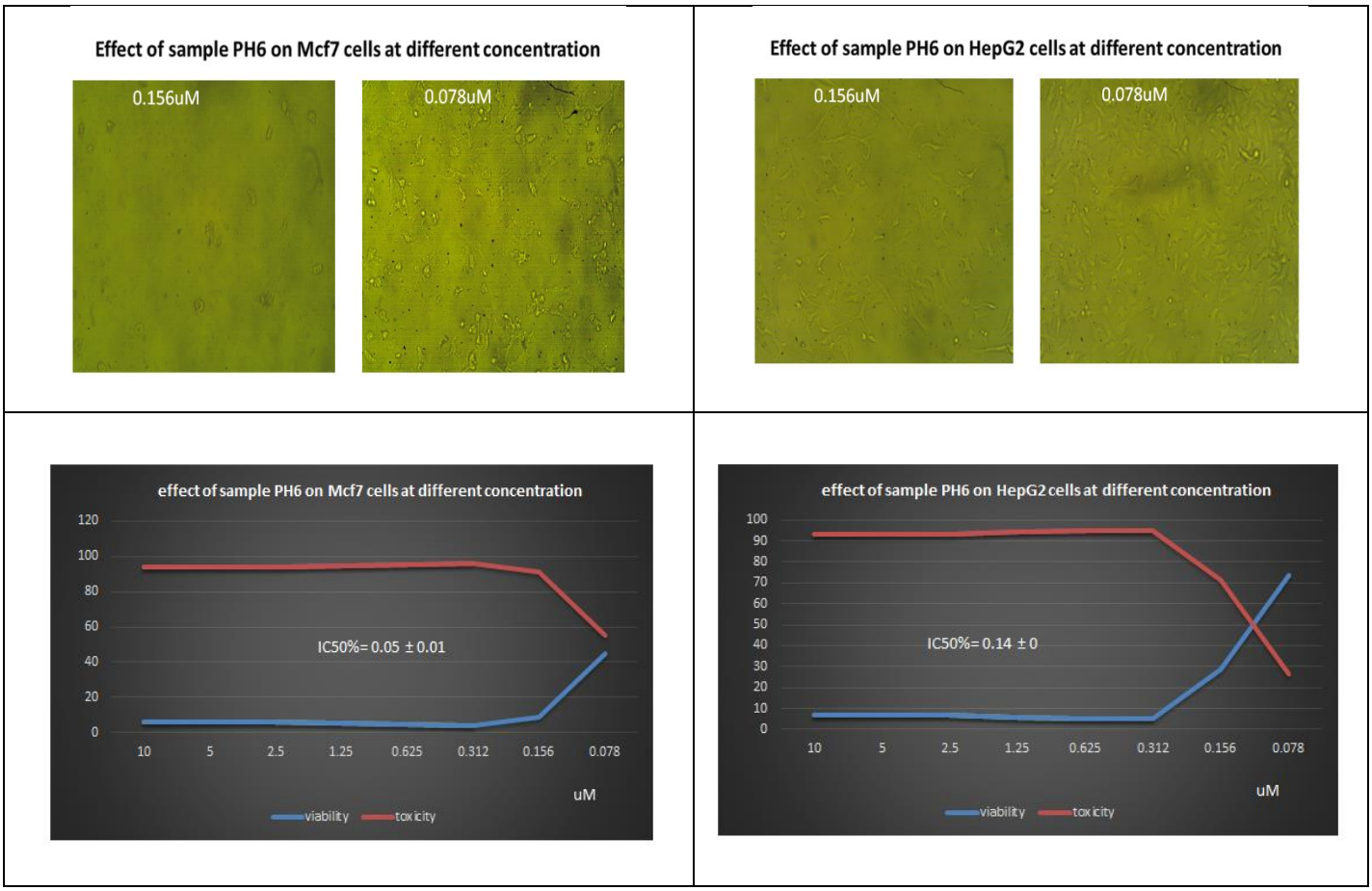


***** Cytotoxicity of PH-6 (compound 20b) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
Mcf7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
PH6	10	0.05	0.05	0.05	0.05	4.91E-18	6.468305304	93.5316947	0.05 ± 0.01
	5	0.05	0.05	0.05	0.05	4.91E-18	6.468305304	93.5316947	
	2.5	0.05	0.05	0.05	0.05	4.91E-18	6.468305304	93.5316947	
	1.25	0.036	0.043	0.044	0.041	0.002517	5.304010349	94.69598965	
	0.625	0.032	0.039	0.04	0.037	0.002517	4.786545925	95.21345408	
	0.312	0.032	0.033	0.036	0.033667	0.001202	4.355325571	95.64467443	
	0.156	0.083	0.061	0.068	0.070667	0.006489	9.141871496	90.8581285	
	0.078	0.35	0.321	0.37	0.347	0.014224	44.89003881	55.10996119	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
PH6	10	0.05	0.05	0.05	0.05	4.91E-18	6.811989101	93.1880109	0.14 ± 0
	5	0.05	0.05	0.05	0.05	4.91E-18	6.811989101	93.1880109	
	2.5	0.05	0.05	0.05	0.05	4.91E-18	6.811989101	93.1880109	
	1.25	0.038	0.047	0.045	0.043333	0.002728	5.903723887	94.09627611	
	0.625	0.037	0.042	0.04	0.039667	0.001453	5.40417802	94.59582198	
	0.312	0.038	0.035	0.039	0.037333	0.001202	5.086285195	94.9137148	
	0.156	0.195	0.208	0.224	0.209	0.008386	28.47411444	71.52588556	
	0.078	0.549	0.559	0.513	0.540333	0.013968	73.61489555	26.38510445	

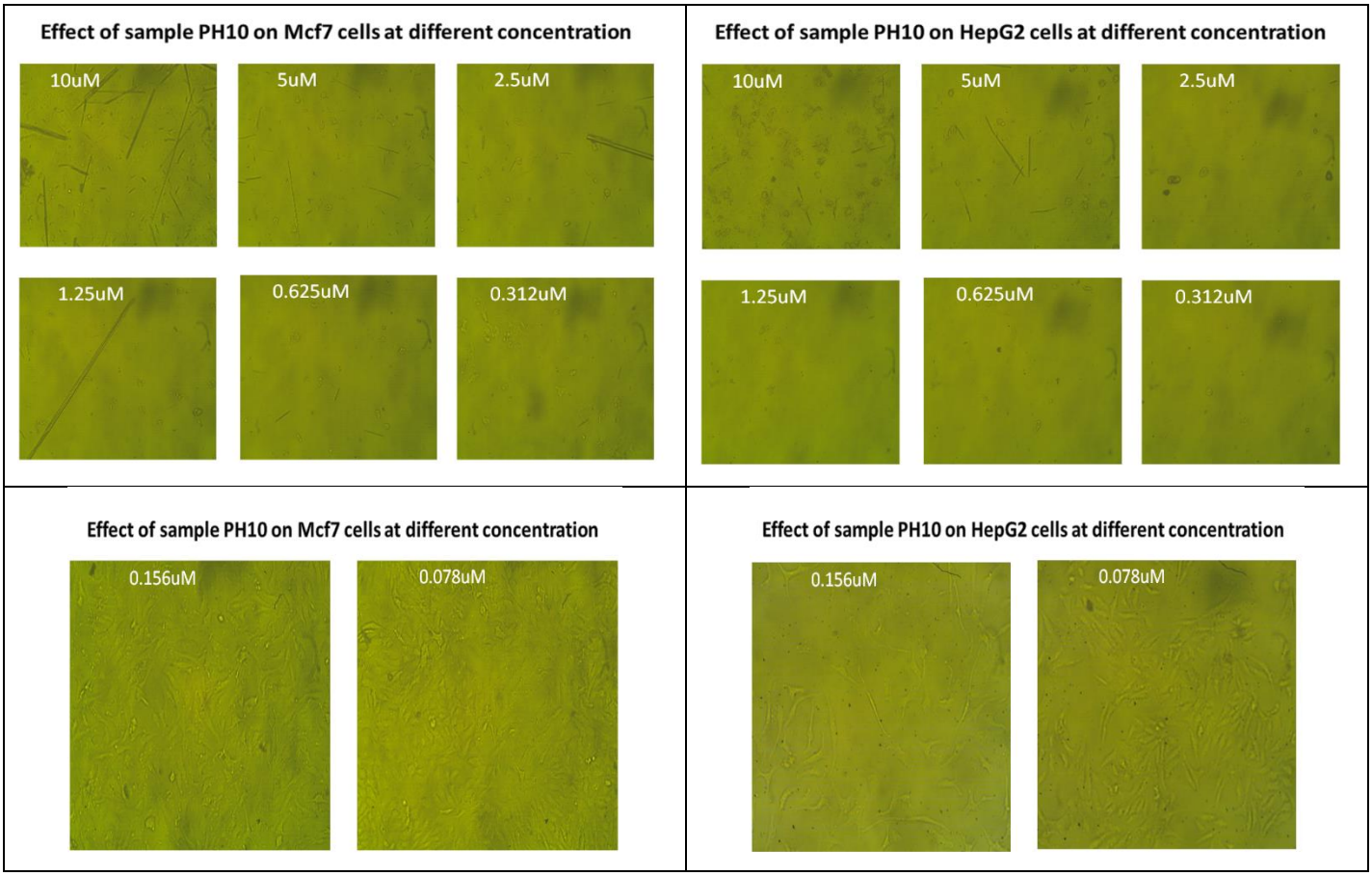


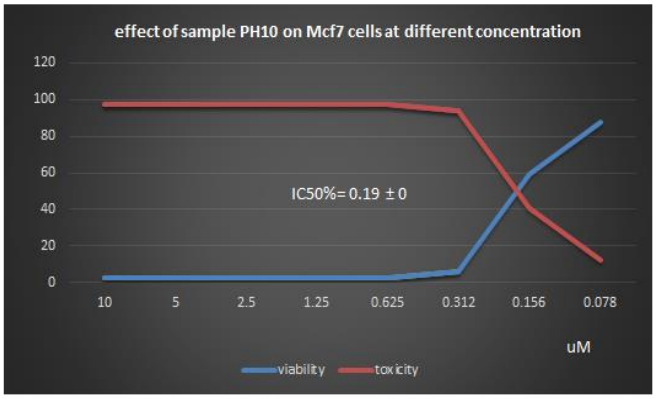


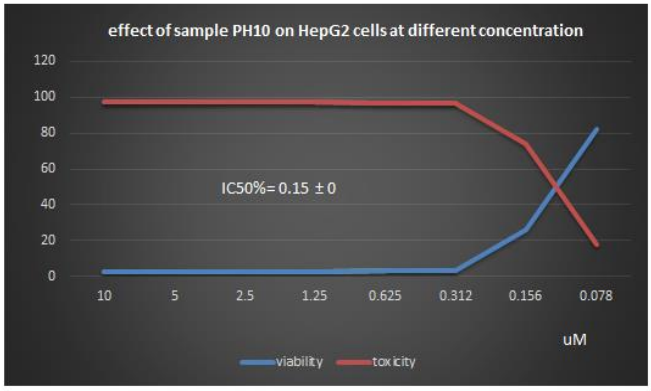
***** Cytotoxicity of PH-10 (compound 21a) against MCF-7 and HepG2**

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
Mcf7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
PH10	10	0.019	0.018	0.019	0.018667	0.000333	2.41483398	97.58516602	0.19 ± 0
	5	0.019	0.02	0.018	0.019	0.000577	2.457956016	97.54204398	
	2.5	0.019	0.022	0.021	0.020667	0.000882	2.673566192	97.32643381	
	1.25	0.02	0.022	0.018	0.02	0.001155	2.587322122	97.41267788	
	0.625	0.02	0.018	0.023	0.020333	0.001453	2.630444157	97.36955584	
	0.312	0.046	0.052	0.049	0.049	0.001732	6.338939198	93.6610608	
	0.156	0.462	0.435	0.477	0.458	0.012288	59.24967658	40.75032342	
	0.078	0.67	0.699	0.664	0.677667	0.010806	87.66709789	12.33290211	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
PH10	10	0.024	0.022	0.02	0.022	0.001155	2.997275204	97.0027248	0.15 ± 0
	5	0.025	0.021	0.02	0.022	0.001528	2.997275204	97.0027248	
	2.5	0.02	0.019	0.023	0.020667	0.001202	2.815622162	97.18437784	
	1.25	0.02	0.021	0.019	0.02	0.000577	2.72479564	97.27520436	
	0.625	0.019	0.027	0.024	0.023333	0.002333	3.178928247	96.82107175	
	0.312	0.026	0.03	0.022	0.026	0.002309	3.542234332	96.45776567	
	0.156	0.2	0.175	0.195	0.19	0.007638	25.88555858	74.11444142	
	0.078	0.618	0.588	0.598	0.601333	0.008819	81.92552225	18.07447775	





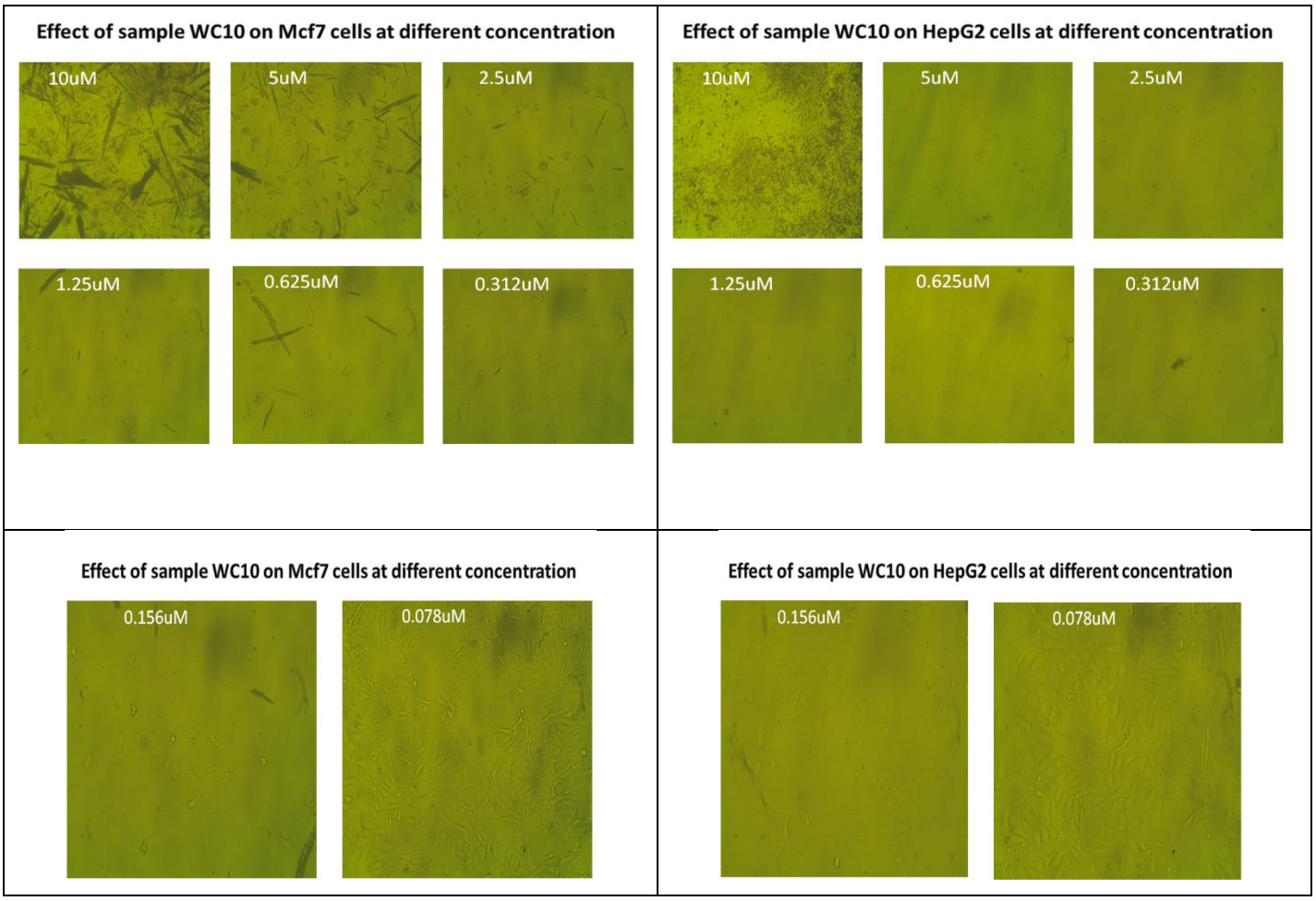


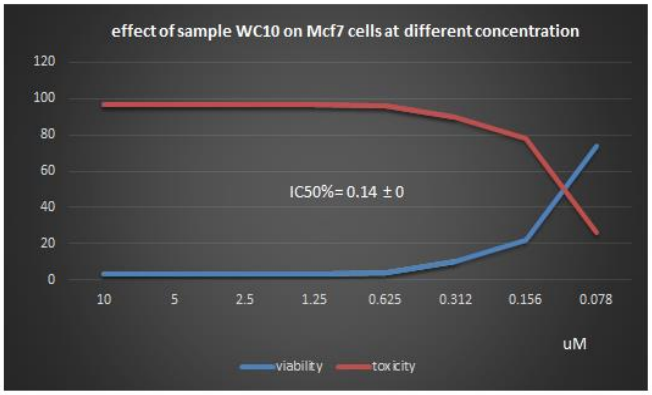
***** Cytotoxicity of WC-10 (compound 21b) against MCF-7 and HepG2**

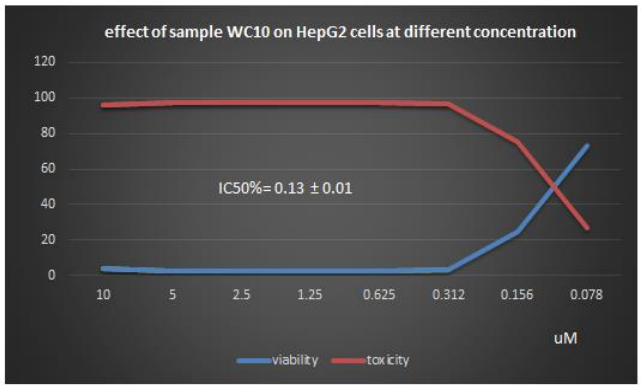
ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
Mcf7	--------	0.772	0.768	0.779	0.773	0.003215	100	0	uM
WC10	10	0.033	0.026	0.027	0.028667	0.002186	3.708495041	96.29150496	0.14 ± 0
	5	0.028	0.029	0.028	0.028333	0.000333	3.665373006	96.33462699	
	2.5	0.024	0.03	0.028	0.027333	0.001764	3.5360069	96.4639931	
	1.25	0.024	0.025	0.029	0.026	0.001528	3.363518758	96.63648124	
	0.625	0.026	0.034	0.031	0.030333	0.002333	3.924105218	96.07589478	
	0.312	0.073	0.076	0.09	0.079667	0.005239	10.30616645	89.69383355	
	0.156	0.152	0.188	0.167	0.169	0.01044	21.86287193	78.13712807	
	0.078	0.563	0.559	0.582	0.568	0.007095	73.47994825	26.52005175	

ID	uM	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
HepG2	--------	0.742	0.728	0.732	0.734	0.004163	100	0	uM
WC10	10	0.03	0.03	0.029	0.029667	0.000333	4.0417802	95.9582198	0.13 ± 0.01
	5	0.019	0.018	0.02	0.019	0.000577	2.588555858	97.41144414	
	2.5	0.023	0.019	0.021	0.021	0.001155	2.861035422	97.13896458	
	1.25	0.02	0.019	0.018	0.019	0.000577	2.588555858	97.41144414	

	0.625	0.02	0.02	0.023	0.021	0.001	2.861035422	97.13896458	
	0.312	0.019	0.026	0.024	0.023	0.002082	3.133514986	96.86648501	
	0.156	0.182	0.166	0.194	0.180667	0.00811	24.61398728	75.38601272	
	0.078	0.583	0.555	0.472	0.536667	0.033328	73.11534968	26.88465032	







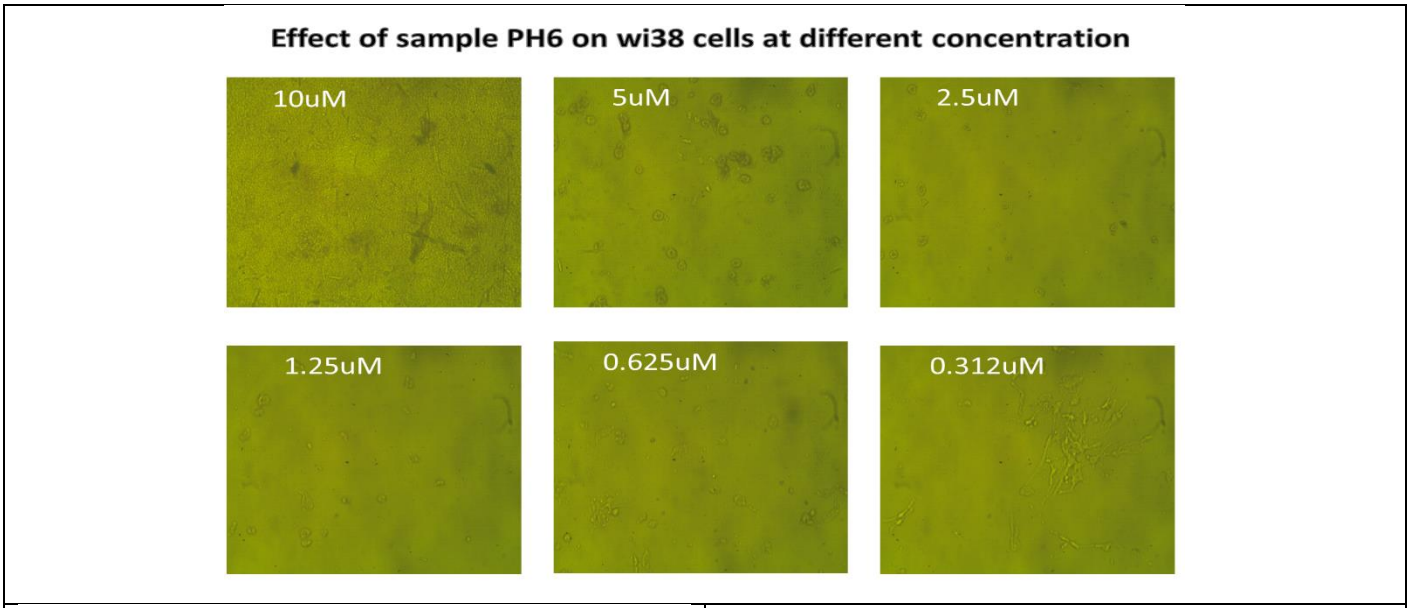
S.2.2. Safety assay

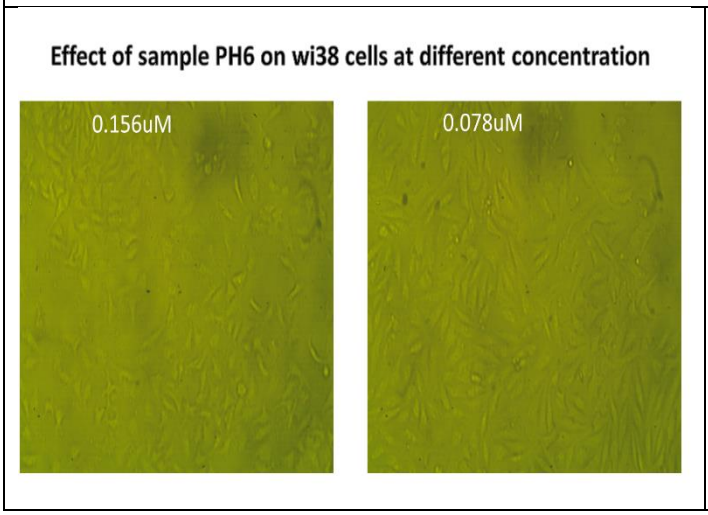
The safety profile of compound **20b** was checked on against normal human lung cells (WI-38) to determine the treatments concentrations that do not depict toxic effects against the tested cells. A portion of 100.0 µl of 6×10^4 cell/ml cells was seeded into each well of a 96-well plate and then the plates 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.

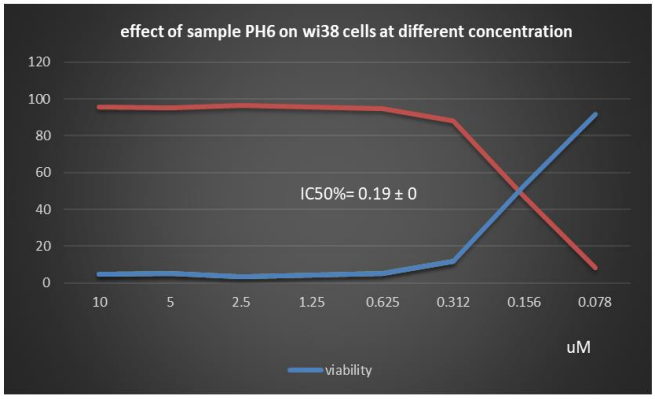
***** Cytotoxicity of PH-6 (compound 20b) against normal human lung cells (WI-38)**

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
PH6	10	0.031	0.033	0.026	0.03	0.002082	4.47761194	95.52238806	0.19 ± 0
	5	0.033	0.036	0.03	0.033	0.001732	4.925373134	95.07462687	
	2.5	0.021	0.022	0.027	0.023333	0.001856	3.482587065	96.51741294	
	1.25	0.027	0.03	0.031	0.029333	0.001202	4.378109453	95.62189055	

	0.625	0.04	0.027	0.036	0.034333	0.003844	5.124378109	94.87562189	
	0.312	0.077	0.082	0.079	0.079333	0.001453	11.84079602	88.15920398	
	0.156	0.352	0.341	0.368	0.353667	0.007839	52.78606965	47.21393035	
	0.078	0.611	0.638	0.597	0.615333	0.012032	91.84079602	8.15920398	







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

Inhibitory activity of compound **20b** 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
PH6		100	2	95.7	30	0	30	4.33	0	4.33	3.333	5.196
		10	1	87.7	30	0	30	12.29	0	12.29	3.333	14.748
		1	0	75.3	30	0	30	24.66	0	24.66	3.333	29.592
		0.1	-1	60.2	30	0	30	39.81	0	39.81	3.333	47.772
		0.01	-2	41.2	30	0	30	58.75	0	58.75	3.333	70.5
EC				0	30	0	30	100	0	100	3.333	120
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. Kinase profiling test

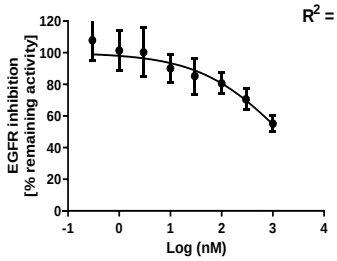
- The in vitro inhibitory activity of compound **20b** against CDK8, PIK3α, BRAF & EGFR kinases enzyme activities was accomplished using Assay Kits (BPS Bioscience, USA) at different eight dilutions of 1000, 300, 100, 30, 10, 3, 1, and 0.3 nM prepared in DMSO.
- The activities of Kinases were observed by measuring chemiluminescence using BioTek™ Synergy2 Microplate Reader (BioTek, USA).
- Different standards were used.
- All samples and controls were verified in triplicates to calculate the concentration that caused 50% inhibition of the kinase activity.

Best-fit values	
LogIC50	3.156
HillSlope	-0.5354
IC50	1431

20b

IC_{50} = 1431

R^2 = 0.7368

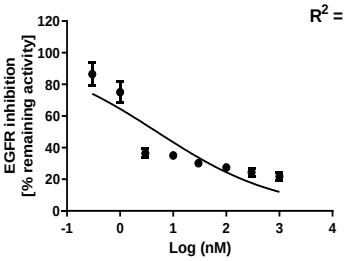


Best-fit values	
LogIC50	0.6930
HillSlope	-0.3747
IC50	4.932

Erlotinib

IC_{50} = 4.932

R^2 = 0.7872

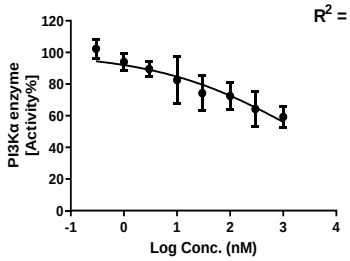


Best-fit values	
LogIC50	3.331
HillSlope	-0.3190
IC50	2143

20b

IC_{50} = 2143

R^2 = 0.7225

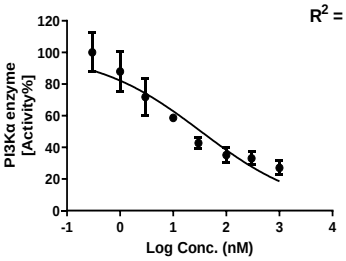


Best-fit values	
LogIC50	1.526
HillSlope	-0.4357
IC50	33.55

PIK-90

IC_{50} = 33.55

R^2 = 0.8703

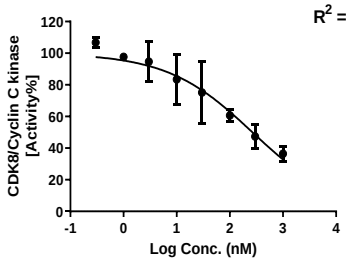


Best-fit values	
LogIC50	2.426
HillSlope	-0.5415
IC50	266.5

20b

IC_{50} = 266.5

R^2 = 0.8587

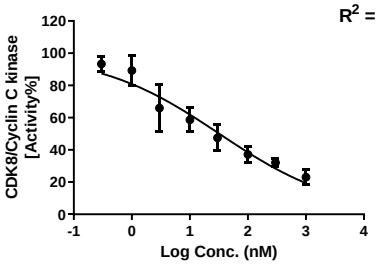


Best-fit values	
LogIC50	1.508
HillSlope	-0.4150
IC50	32.24

Cortistatin A

IC_{50} = 32.24

R^2 = 0.8942

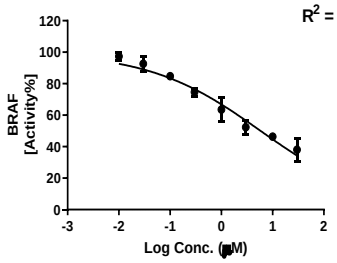


Best-fit values	
LogIC50	0.7665
HillSlope	-0.3977
IC50	5.841

20b

IC_{50} = 5.841

R^2 = 0.9417

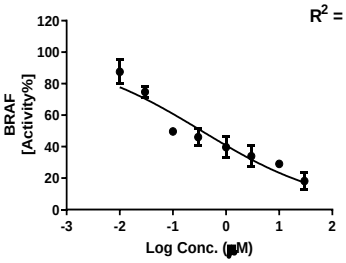


Best-fit values	
LogIC50	-0.4588
HillSlope	-0.3531
IC50	0.3477

Vemurafenib

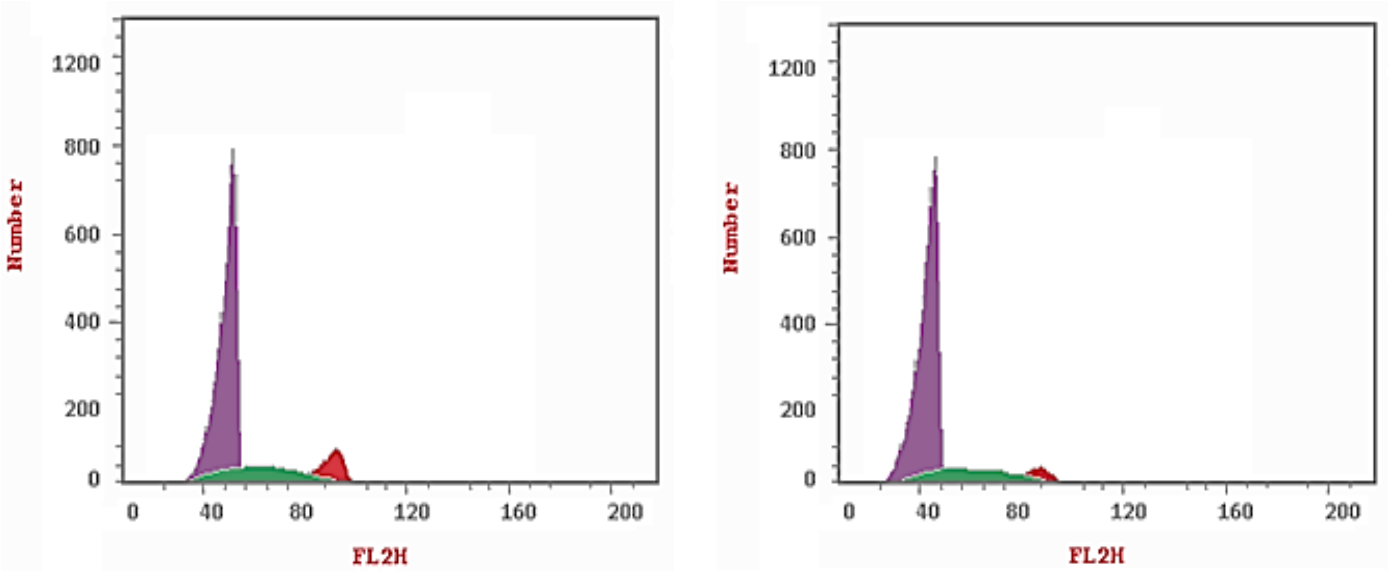
IC_{50} = 0.3477

R^2 = 0.8854



S.2.5. 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 **20b** at a concentration of 0.06 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.

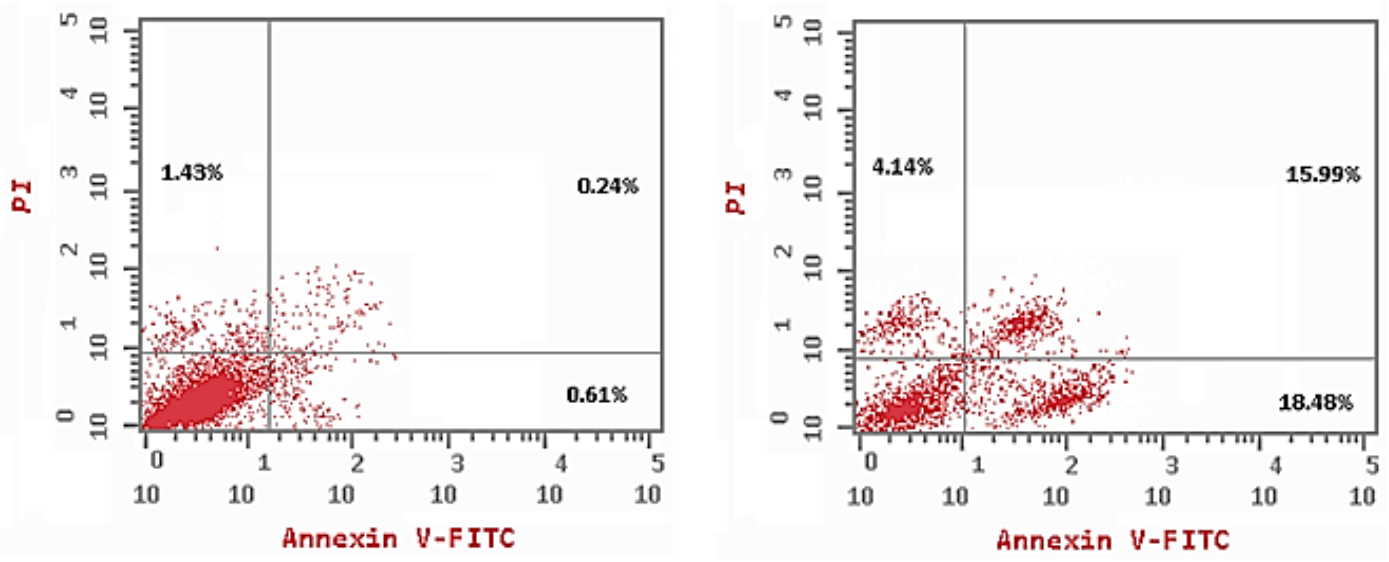


(**A**) Control (MCF-7 cells) (**B**) Compound **20b**

Fig. S1. The distribution of MCF-7 cells across the cell cycle phases was assessed using flow cytometry. Panel (A) represents the control condition, while panel (B) illustrates the influence of compound **20b** on the cell cycle distribution.

S.2.6. Apoptosis analysis

The annexin V externalization assay for apoptosis was performed using flow cytometry after treatment MCF-7 cells with compound **20b** (at concentrations of 0.06 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 **20b**, 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+).



(**A**) Control (MCF-7 cells) (**B**) Compound **20b**

Fig. S2. Compound **20b** triggered apoptosis in MCF-7 cells. Panel (A) represents the control condition, while panel (B) illustrates the impact of compound **20b** on inducing apoptosis.

S.2.7. *In vitro* assay for caspase-8, caspase-9 using RT-PCR Tanique

The molecular anticancer mode of action of compound **20b** was investigated by screening their ability to affect the level of caspase-8 and caspase-9 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 0.06 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	
Casp8	: F 5'- AGAAGAGGGTCATCCTGGGAGA-3',
Casp8	: R 5'- TCAGGACTTCCTTCAAGGCTGC-3'.
Casp9	: F 5'- GTTTGAGGACCTTCGACCAGCT -3',
Casp9	: R 5'- CAACGTACCAGGAGCCACTCTT -3'.
GAPDH	: F 5'- GTCTCCTCTGACTTCAACAGCG-3'
GAPDH	: R 5'- ACCACCCTGTTGCTGTAGCCAA-3'

Sample	Gene Expression		
	Control cells	Test cells	FLD

MCF7

Ser	code	Conc	GAPDH	Casp8	ΔCTC	GAPDH	Casp8	ΔCTE	ΔΔ CT	2^ ΔΔCT
			HC	TC	TC-HC	HE	TE	TE-HE	ΔCTE-ΔCTC	E=1.872
1	PH6		21.66	33.61	11.95	21.85	31.84	9.99	-1.96	3.4176
2	Control		21.66	33.61	11.95	21.66	33.61	11.95	0	1

MCF7

Ser	code	Conc	GAPDH	Casp9	ΔCTC	GAPDH	Casp9	ΔCTE	ΔΔ CT	2^ ΔΔCT
			HC	TC	TC-HC	HE	TE	TE-HE	ΔCTE-ΔCTC	E=1.872
1	PH6		21.66	33.95	12.29	21.85	31.44	9.59	-2.7	5.4353
2	Control		21.66	33.95	12.29	21.66	33.95	12.29	0	1

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_**22** 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 **20b** 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 **22** 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:

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

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

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

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

$E_{SA} = \gamma.SASA$ 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.

The ProLIF technique was employed to identify amino acids engaged in ligand interactions and their occurrence frequency. For sorafenib, 13 amino acids were detected with hydrophobic binding frequencies surpassing 80%. Specific amino acids crucial for compound interactions were pinpointed, including Leu838 (91.2%), Lys866 (87.9%), Ile886 (92.9%), Val846 (98.6%), Leu887 (95.1%), Val897 (88.9%), Val914 (99.7%), Ile890 (95.2%), Leu1017 (95.4%), Leu1033 (91%), His1024 (86.6%), Cys1043 (99.8%), and Asp1044 (99.3%). These high percentages underscore their pivotal role in interactions and highlight their significance in binding within the complex (**Fig. S3**). Moreover, Asp1044 exhibited an 88.2% hydrogen bond rate. On the other hand, compound 20b interacted with 12 amino acids predominantly through hydrophobic interactions, each with a frequency of at least 80% (**Fig. S4**). These amino acids include Leu838 (95.7%), Val846 (99.9%), Ala864 (100%), Val865 (99.3%), Lys866 (97.9%), Leu887 (96%), Val897

(100%), Val914 (99.9%), Leu1033 (99.4%), Cys1043 (99.9%), Asp1044 (99.9%), and Phe1045 (89.2%). Notably, Asp1044 also showed a higher rate of hydrogen bond formation (97.1%) compared to sorafenib. Nine of these 12 amino acids were common with sorafenib, indicating a similar mode of interaction. Furthermore, six of these amino acids exhibited stronger contributions compared to sorafenib, as depicted in **Fig. S3**.

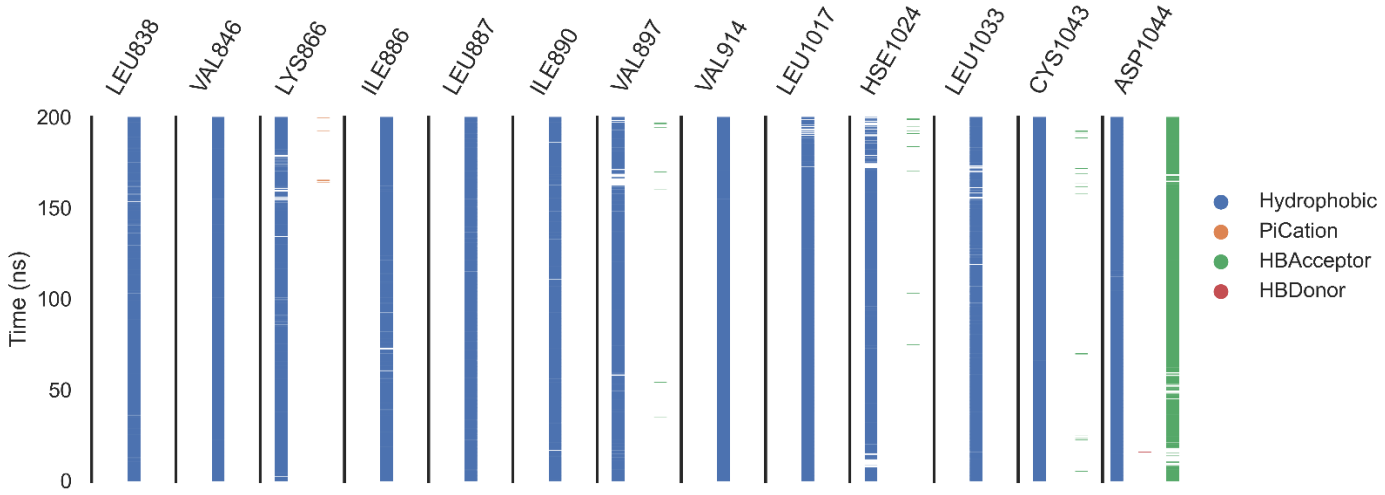


Fig. S3. a comprehensive depiction of the amino acids, the interactions' nature, and their frequency throughout the entire simulation period of VEGFR-2_sorafenib complex, as analyzed through the ProLIF Python library.

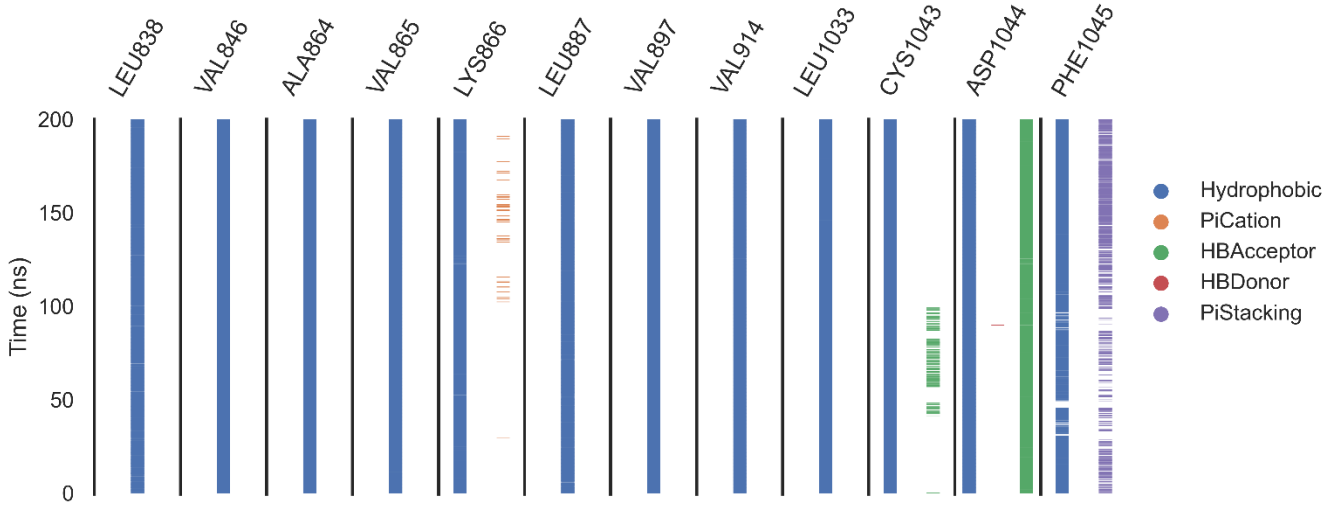


Fig. S4. a comprehensive depiction of the amino acids, the interactions' nature, and their frequency throughout the entire simulation period of VEGFR-2_**20b** complex, as analyzed through the ProLIF Python library.

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.

Principal Component Analysis (PCA) was employed to detect coordinated movements. The determination of the reduced subspace size was guided by several criteria detailed in the methods section, including the examination of the scree plot, distribution of eigenvectors, and cumulative variance captured by additional eigenvectors. The scree plot clearly shows a change in slope starting from the 2nd principal component (PC). The first eigenvector alone explained nearly 81% of the total variance, while the combined contribution of the first three eigenvectors accounted for approximately 90% of the overall variance (**Fig. S5**). It's noteworthy that the distribution of the first three PCs deviated from a Gaussian pattern (**Fig. S6**). Therefore, it was decided to utilize the top three eigenvectors as representatives of the fundamental subspace.

To assess the level of randomness in the behavior of the initial ten eigenvectors, the cosine content was computed for both systems (**Fig. S7**). The analysis revealed that, with the exception of the second principal component in the VEGFR-2_**20b** system and the sixth PC in VEGFR-2_sorafenib, the cosine content of the first 10 eigenvectors remained below 0.2 in both systems. Moreover, the Root Mean Square Inner Product (RMSIP) indicated a limited overlap between the two subspaces, particularly for the first three eigenvectors, with a value of 18%. Additionally, the RMSIP analysis demonstrated only a 27% similarity between the C matrices of the two systems, suggesting distinctive sampling characteristics for each system.

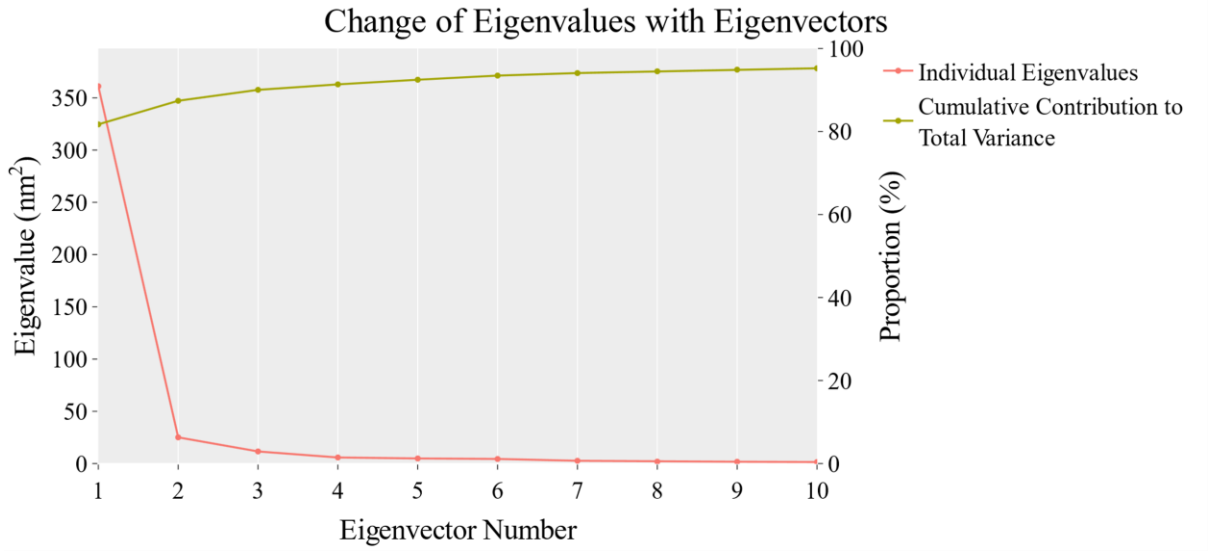


Fig. S5. the variation in eigenvalues as the number of eigenvectors increases (blue line). Furthermore, it illustrates the cumulative retained variance in the eigenvectors (red line).

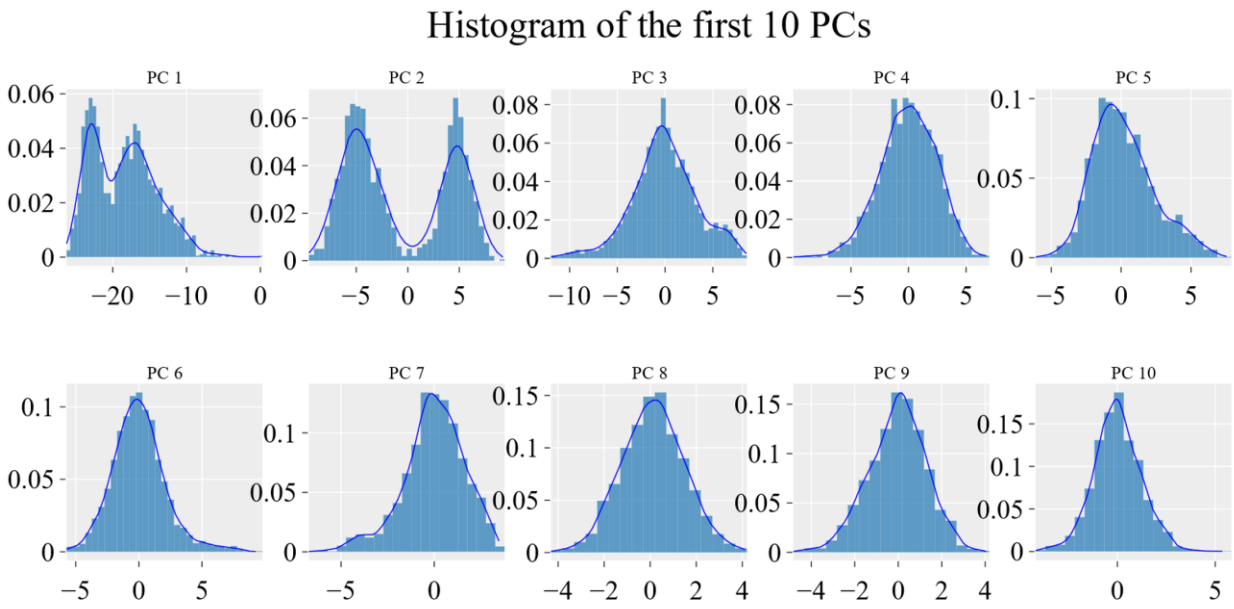


Fig. S6. The arrangement of the initial ten eigenvectors

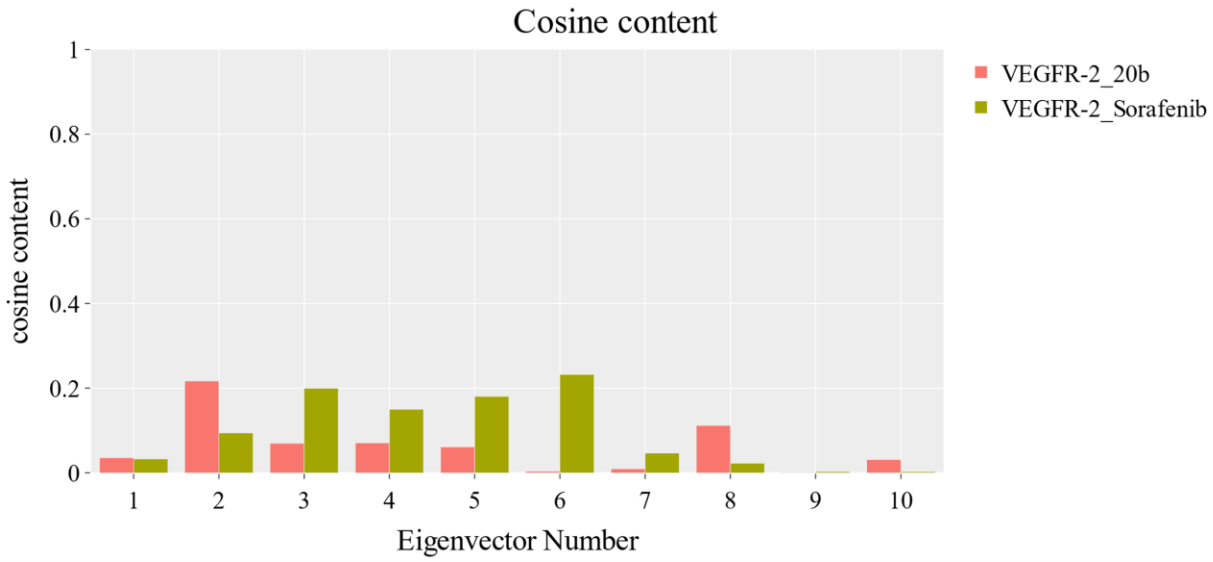


Fig. S7. Cosine content's Values of the first ten eigenvectors.

The results obtained from projecting each trajectory onto the first three eigenvectors of the new C matrix are presented in **Fig. S8-S10**. In these graphs, the larger dot represents the average structure of the respective trajectory. **Fig. S8** illustrates the projection onto the first two eigenvectors, revealing distinct average structures for each trajectory. Moreover, the frames demonstrate separate sampling, with only a small amount of overlap observed at the initial stages of the simulation (pale red and white dots). Furthermore, the sampling patterns for the reference trajectory exhibit clustered frames indicating similarity within the sampled frames. Conversely, VEGFR-2_20b shows two clusters indicating two sampled regions (orange dots and dark red dots). Similarly, **Fig. S9** displays a similar pattern to **Fig. S8**, with different average structures and distinct clustered sampling. Finally, **Fig. S10** illustrates the projection onto the second and third eigenvectors, indicating that the two trajectories differ from each other and exhibit only slight overlap in the early frames. Notably, in the VEGFR-2_**20b** complex frames, two clusters are observed (similar to PC1/PC2 projection), with one cluster (orange dots) being in proximity to the region of VEGFR-2_sorafenib, while the other cluster is located further away. This suggests the sampling of a new region after spending some time near the initial structure. In the visualization, small white-to-black dots represent frames from the VEGFR-2 in the VEGFR-2_sorafenib simulation, while small white-to-red dots represent frames from the VEGFR-2 in the VEGFR-2_**20b** simulation.

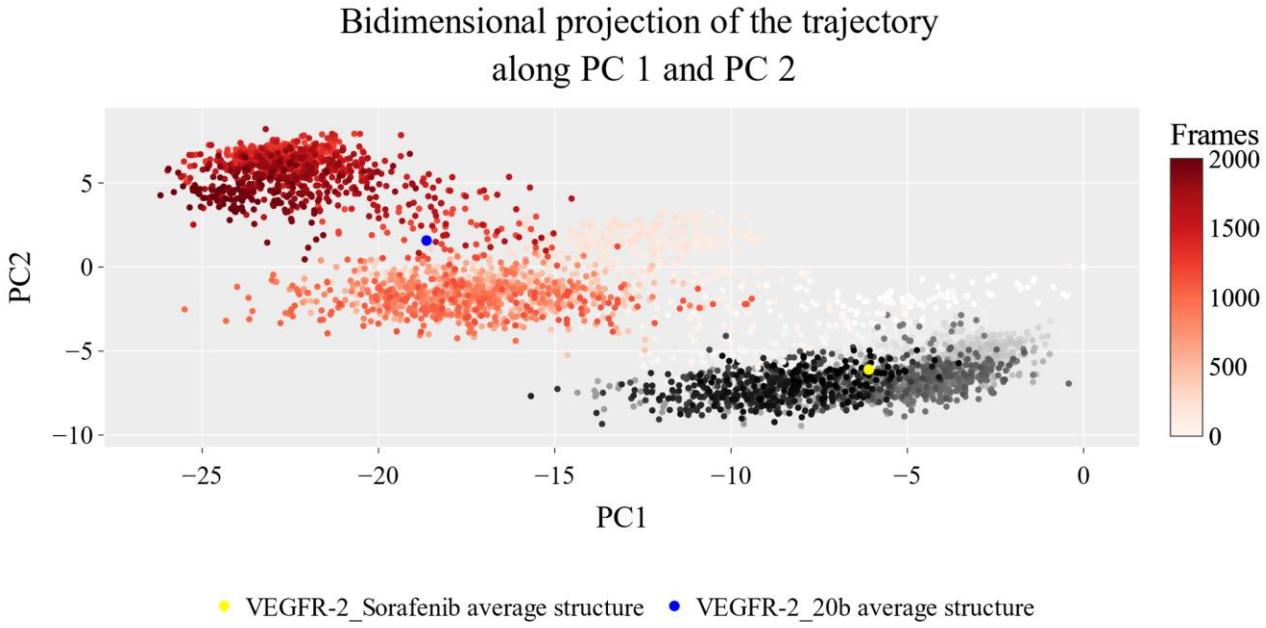


Fig. S8. Each trajectory is projected onto the first two eigenvectors.

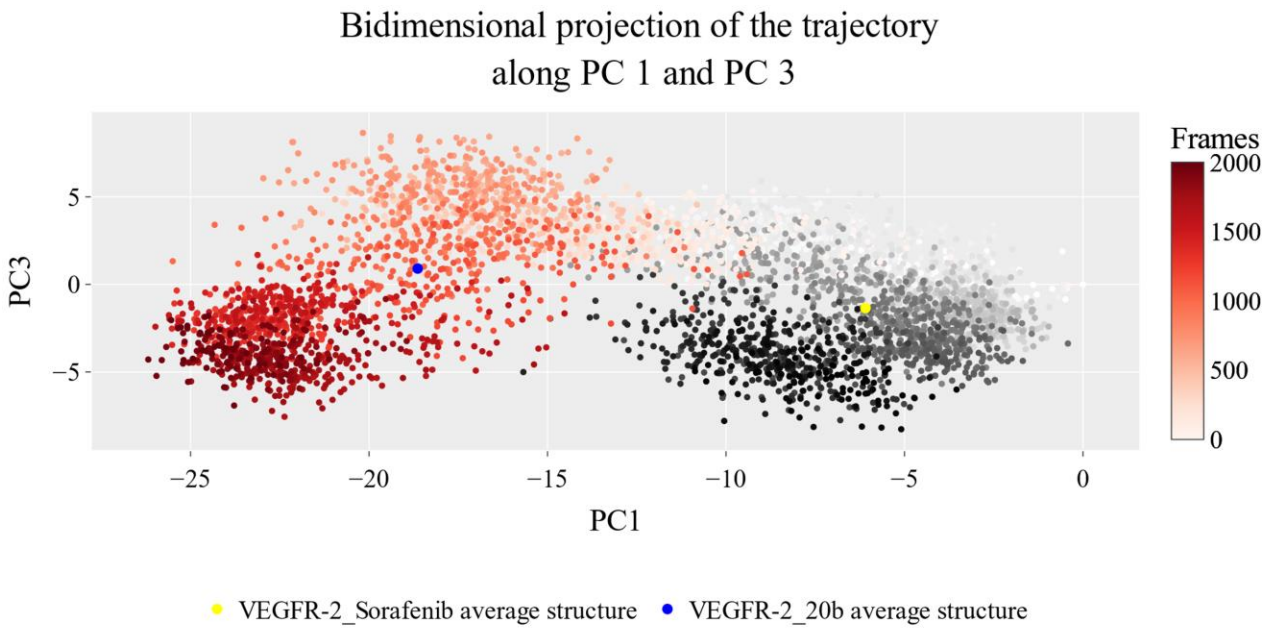


Fig. S9. Each trajectory is projected onto the first and third eigenvectors.

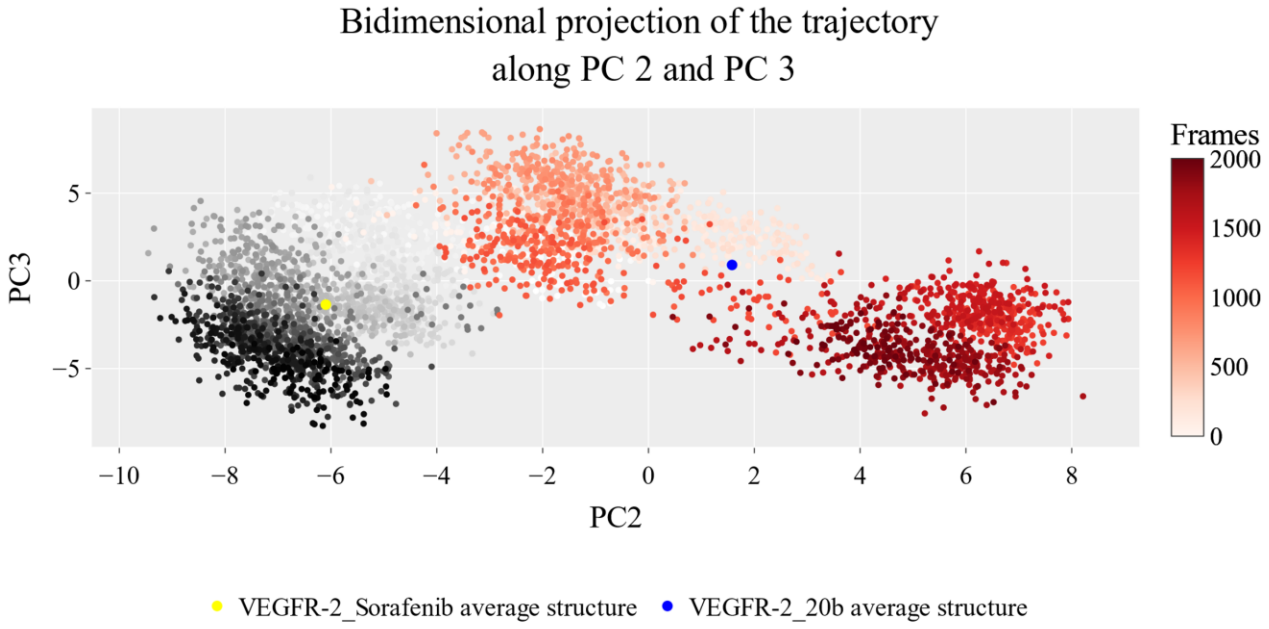


Fig. S10. Each trajectory is projected onto the second and third eigenvectors.

S.3.4. Density Function Theory (DFT) calculations

The calculations have been done using DFT/B3LYP/6-31+G (d, p) theory level utilizing Gaussian 09 software with the aid of various software including GaussSum, Multiwfn, AIMALL and Gauss View 5. Mulliken charge analysis as well as TDOS; Total density of states, MEP; molecular electrostatic potential, QTAIM; quantum theory of atoms in molecules, analyses have performed. FMO; frontiers molecular orbital analysis and reactivity descriptors including global softness(σ), global hardness(η), chemical potential(μ), electron affinity (EA), electronegativity (χ), electrophilicity index(ω), ionization potential (IP), and energy change (ΔE) have been determined according to the following equations:

$IP = -E_{HOMO}$

$EA = -E_{LUMO}$

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

$\eta = (IP-EA)$

$\chi = -\mu$

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

$\sigma = 1/\eta$

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

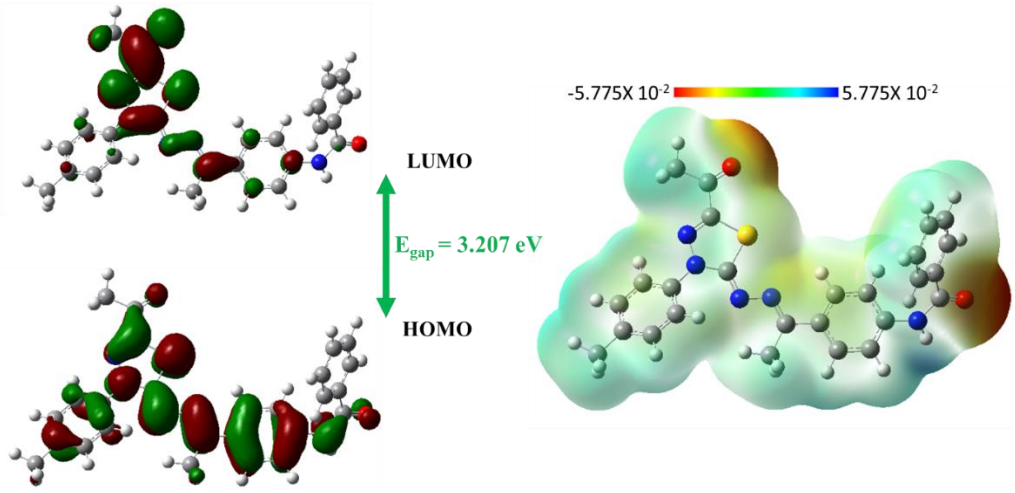
$\Delta E = -\omega$

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

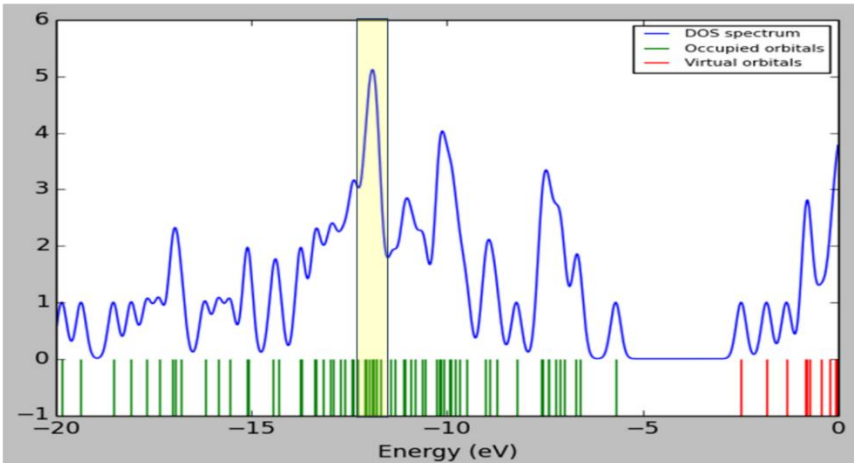
The HOMO and LUMO (frontier molecular orbitals) are electronic characteristics that are linked to chemical reactivity. Their energies have been determined to be -5.684 and -2.476 eV, respectively. This results in values of the ionization potential (IP), electron affinity (EA), chemical potential (μ), hardness (η), and electrophilicity (ω). The wave function distribution of the HOMO and LUMO is found over the conjugated π-system, suggesting a π-π intramolecular charge transfer due to the lowest energy gap (E_{gap}) of 3.207 eV, **Fig. S11.A**. Such a result indicates easier excitations of electrons from HOMO to LUMO which imply potential chemical reactivity [49]. The electron density map over electrostatic potential in **Fig. S11.A** shows that the structure is composed of several charged (negative and positive) areas. The oxygen atoms in the carbonyl groups belong to the negatively charged areas. In contrast, the carbonyl carbon and hydrogen atoms are the most positively charged atomic sites.

To gain deep insights into the electronic structure and then the chemical behavior of **20b**, the total density of states (TDOs) has been analyzed and demonstrated in **Fig. S11.B**. The highest peak of electronic density (in yellow color) defines the orbitals that are involved in bonding with the target. The narrow HOMO-LUMO gap in energy suggests easier excitation of electrons and possibly increased reactivity [50]. The topology analysis of the electron density (ρ) in the quantum theory of atoms in molecule (QTAIM) has been performed based on the obtained ρ from the DFT method. QTAIM gains a comprehensive insight into the electronic structure, bonding nature and reactivity of a molecule. In QTAIM, bond critical points (BCP) and bond paths are identified based on the gradient of ρ. In addition, the Laplacian of electron density ($\nabla^2\rho$), total electron energy density (H(r)), kinetic electron energy density (G(r)) and potential electron energy density (V(r)) at each BCP are determined. **Fig. S11.C** demonstrated that three bonds are generated within **20b**.

A



B



C

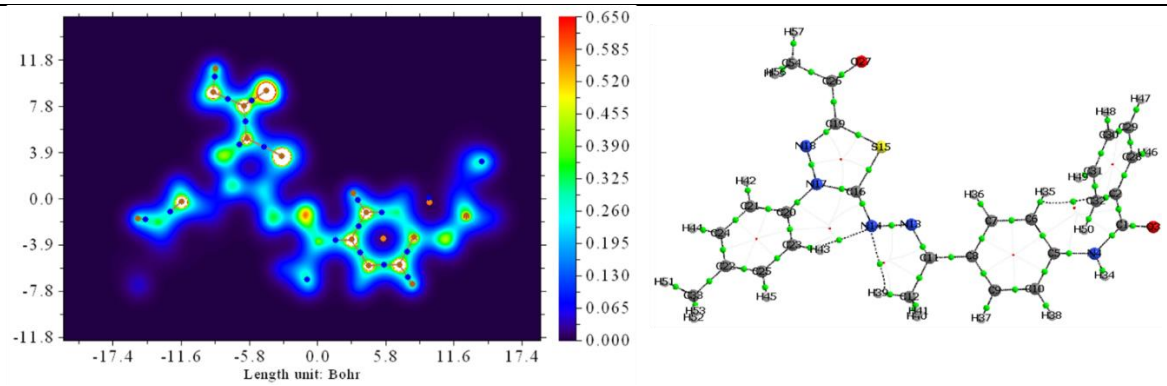


Fig. S11. The FMO and energy gap, the ESP (A), the TDOS spectrum (B), and the QTAIM maps (C) for compound **20b**. These results were obtained using B3LYP/6-31+G(d,p).

The results in **Table S1** show that all bonds within the molecule are covalent while the newly generated bonds are electrostatic in nature which provide the molecule with stability.

Table S1. The QTAIM parameters (a.u.) at bond critical points (BCPs) of **20b**.

BCP #	Atoms	(ρ)	(∇2ρ)	K(r)	G(r)	V(r)	H(r)
1	C1 - C2	0.26569	-0.66472	0.227058	0.060877	-0.28794	-0.22706
2	C1 - O3	0.401135	-0.07824	0.692595	0.673035	-1.36563	-0.6926
3	C1 - N4	0.307836	-0.98721	0.45509	0.208287	-0.66338	-0.45509
4	C6 - C7	0.315624	-0.86494	0.318946	0.102711	-0.42166	-0.31895
5	C7 - C8	0.304828	-0.81305	0.296259	0.092995	-0.38925	-0.29626
6	C5 - C6	0.307943	-0.83457	0.304689	0.096045	-0.40073	-0.30469
7	N4 - H34	0.341404	-1.81804	0.506558	0.052049	-0.55861	-0.50656
8	N4 - C5	0.285186	-0.86617	0.395392	0.178851	-0.57424	-0.39539
9	C5 - C10	0.311734	-0.85085	0.311043	0.09833	-0.40937	-0.31104
10	C8 - C9	0.306803	-0.81898	0.300277	0.095532	-0.39581	-0.30028
11	C9 - C10	0.312384	-0.85008	0.312289	0.099768	-0.41206	-0.31229
12	C8 - C11	0.270685	-0.67357	0.233728	0.065336	-0.29906	-0.23373
13	C11 - N13	0.370388	-0.80275	0.638647	0.43796	-1.07661	-0.63865
14	C11 - C12	0.254597	-0.60216	0.211093	0.060553	-0.27165	-0.21109
15	**C12 - H39**	**0.283535**	**-1.01715**	**0.296139**	**0.041853**	**-0.33799**	**-0.29614**
16	N13 - N14	0.342813	-0.63817	0.331307	0.171764	-0.50307	-0.33131
17	S15 - C16	0.195021	-0.35677	0.148677	0.059485	-0.20816	-0.14868
18	N14 - H39	0.016938	0.074658	-0.00337	0.015297	-0.01193	0.003368

19	N14 - C16	0.386609	-1.19354	0.664567	0.366183	-1.03075	-0.66457
20	N17 - C20	0.271242	-0.72997	0.383585	0.201093	-0.58468	-0.38359
21	C16 - N17	0.301339	-0.93736	0.429118	0.194778	-0.6239	-0.42912
22	N18 - C19	0.3669	-0.65461	0.629766	0.466114	-1.09588	-0.62977
23	N17 - N18	0.364394	-0.69898	0.36572	0.190976	-0.5567	-0.36572
24	S15 - C19	0.199683	-0.37223	0.155012	0.061955	-0.21697	-0.15501
25	C20 - C21	0.311154	-0.84774	0.311198	0.099263	-0.41046	-0.3112
26	C21 - C24	0.312567	-0.85216	0.31302	0.09998	-0.413	-0.31302
27	C20 - C23	0.311644	-0.85191	0.312724	0.099745	-0.41247	-0.31272
28	C22 - C24	0.309513	-0.83387	0.305656	0.097188	-0.40284	-0.30566
29	N14 - H43	0.012326	0.04491	-0.0019	0.009333	-0.00744	0.001895
30	C22 - C25	0.310062	-0.83617	0.306702	0.09766	-0.40436	-0.3067
31	C24 - H44	0.283933	-1.01628	0.295652	0.041583	-0.33723	-0.29565
32	C23 - C25	0.311698	-0.84757	0.311184	0.099292	-0.41048	-0.31119
33	C19 - C26	0.272612	-0.69319	0.238605	0.065306	-0.30391	-0.23861
34	C26 - C54	0.257637	-0.62639	0.21709	0.060491	-0.27758	-0.21709
35	C26 - O27	0.401319	0.128282	0.68371	0.715781	-1.39949	-0.68371
36	C2 - H35	0.007972	0.028906	-0.00161	0.00562	-0.00401	0.001607
37	C2 - C28	0.308749	-0.8308	0.30444	0.096741	-0.40118	-0.30444
38	C28 - C29	0.313249	-0.85851	0.313653	0.099027	-0.41268	-0.31365
39	C29 - C30	0.31134	-0.85021	0.309849	0.097297	-0.40715	-0.30985
40	C2 - C32	0.307424	-0.81996	0.302336	0.097347	-0.39968	-0.30234
41	C30 - C31	0.311637	-0.85137	0.310415	0.097572	-0.40799	-0.31041
42	C31 - C32	0.312385	-0.85297	0.312165	0.098923	-0.41109	-0.31217
43	C22 - C33	0.253314	-0.60043	0.20794	0.057832	-0.26577	-0.20794
44	C33 - H53	0.276822	-0.94922	0.282194	0.04489	-0.32709	-0.2822
45	C6 - H35	0.286709	-1.04296	0.301562	0.040822	-0.34238	-0.30156
46	C7 - H36	0.2893	-1.07597	0.307201	0.038209	-0.34541	-0.3072
47	C10 - H38	0.283245	-1.00992	0.294717	0.042236	-0.33695	-0.29472
48	C9 - H37	0.286509	-1.035	0.300682	0.041932	-0.34261	-0.30068
49	C12 - H40	0.274437	-0.93097	0.278644	0.045903	-0.32455	-0.27864
50	C12 - H41	0.27294	-0.91933	0.276075	0.046242	-0.32232	-0.27608
51	C21 - H42	0.287081	-1.05044	0.302678	0.040069	-0.34275	-0.30268
52	C23 - H43	0.289631	-1.07738	0.308385	0.039041	-0.34743	-0.30838
53	C25 - H45	0.283987	-1.01647	0.295735	0.041618	-0.33735	-0.29574
54	C28 - H46	0.287509	-1.05879	0.303235	0.038539	-0.34177	-0.30324
55	C29 - H47	0.285076	-1.02896	0.297915	0.040675	-0.33859	-0.29792
56	C30 - H48	0.285212	-1.03066	0.298092	0.040427	-0.33852	-0.29809
57	C31 - H49	0.284988	-1.02766	0.297775	0.040859	-0.33863	-0.29777
58	C32 - H50	0.286819	-1.04349	0.301365	0.040493	-0.34186	-0.30137
59	C33 - H51	0.276638	-0.94792	0.281896	0.044915	-0.32681	-0.2819
60	C33 - H52	0.273723	-0.92686	0.276852	0.045137	-0.32199	-0.27685
61	C54 - H55	0.273911	-0.93559	0.278519	0.044621	-0.32314	-0.27852

62	C54 - H57	0.280521	-0.9905	0.289696	0.04207	-0.33177	-0.2897
63	C54 - H56	0.276034	-0.9537	0.282217	0.043792	-0.32601	-0.28222

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

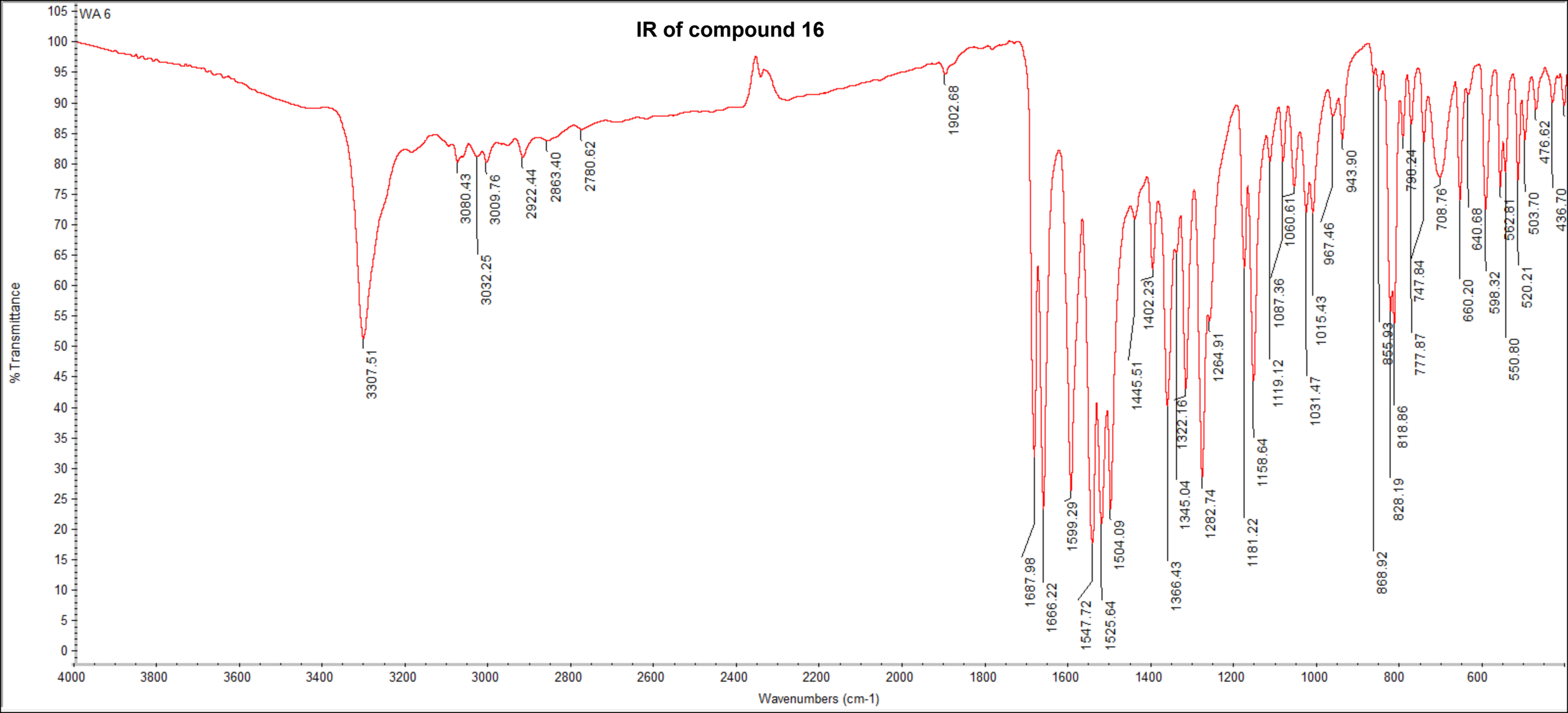
The toxicity parameters of the synthesized compounds were calculated using Discovery studio 4.0. Sorafenib was used as a reference molecule. At first, the CHARMM force field was applied then the compounds were prepared and minimized according to the preparation of small molecule protocol.

- **Preparation of the tested compounds:**

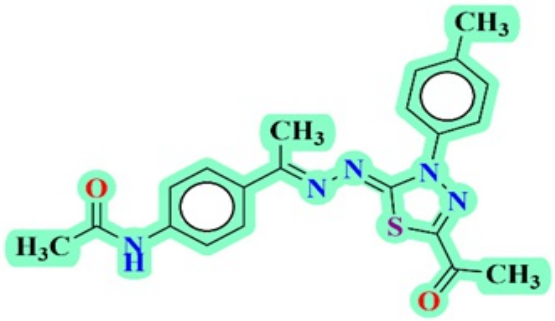
In this protocol, the general-purpose panel was utilized with the activation of the Prepare ligand option. The change ionization was switched on the true option using the Rule based as an ionization method. In Rule based task, we used the carboxylate as an acid ionization. Additionally, the primary, 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.

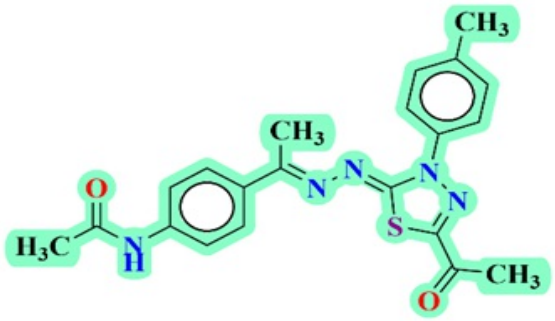

IR of compound 16
WA 6
% Transmittance
Wavenumbers (cm-1)
3307.51
3080.43
3032.25
3009.76
2922.44
2863.40
2780.62
1902.68
1687.98
1666.22
1599.29
1547.72
1525.64
1504.09
1445.51
1402.23
1366.43
1345.04
1322.16
1282.74
1264.91
1181.22
1158.64
1119.12
1087.36
1060.61
1031.47
1015.43
967.46
943.90
868.92
855.93
828.19
818.86
790.24
777.87
747.84
708.76
660.20
640.68
598.32
562.81
550.80
520.21
503.70
476.62
436.70

1H NMR of 16


CH3
CH3
O
H3C
N
H
N
N
N
N
S
CH3
O

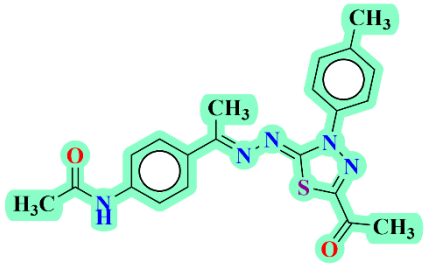
10.15
7.94
7.94
7.93
7.92
7.92
7.85
7.85
7.84
7.83
7.83
7.68
7.67
7.66
7.41
7.40
7.39
2.60
2.40
2.39
2.09
1.04
2.05
2.03
2.04
2.06
2.82
5.92
3.04
f1

13C NMR of 16

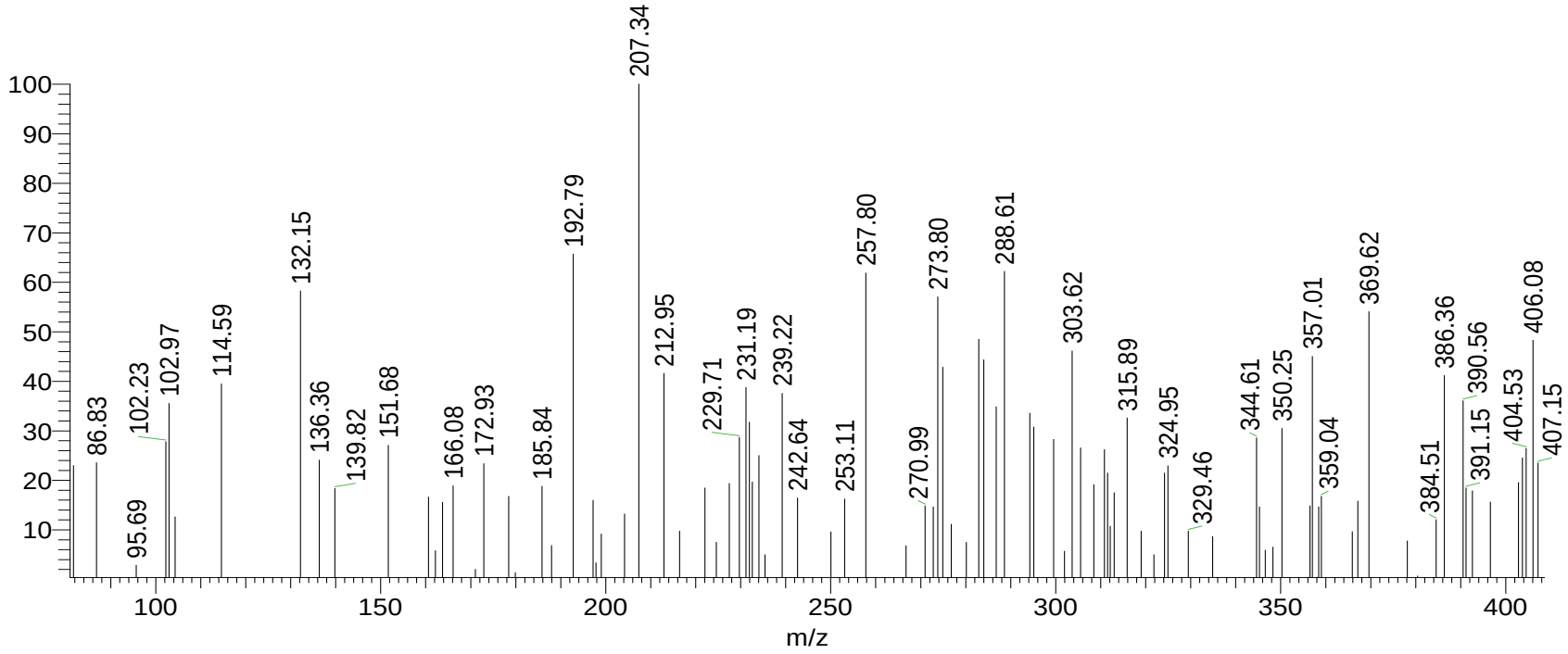

CH3
CH3
O
N
H
H3C
N
N
N
N
S
CH3
O

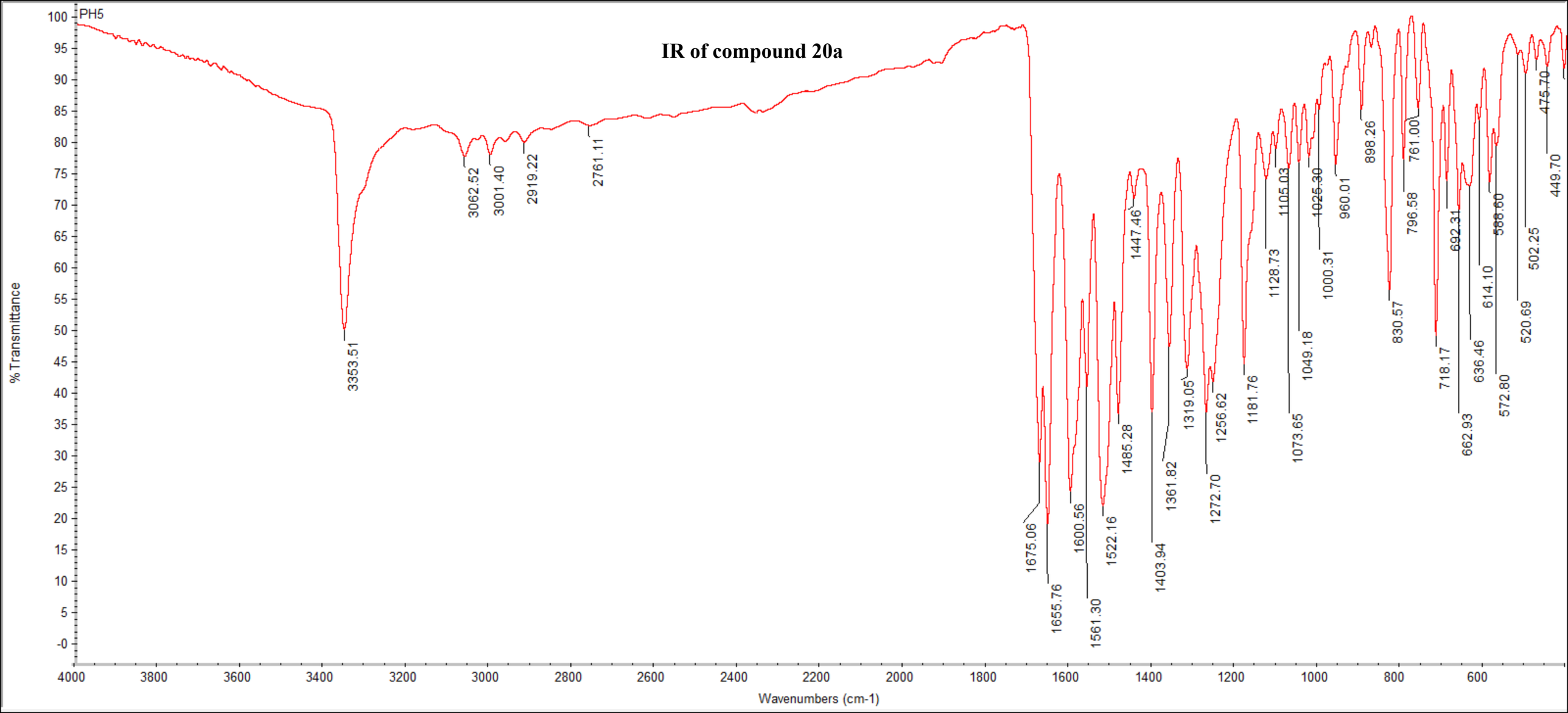
190.21
169.01
164.55
160.26
150.83
141.38
137.24
136.89
132.22
130.00
127.57
122.53
119.02
40.63 DMSO
40.42 DMSO
40.21 DMSO
40.00 DMSO
39.80 DMSO
39.59 DMSO
39.38 DMSO
25.50
24.57
21.10
15.63
f1

Mass spec. of compound 16

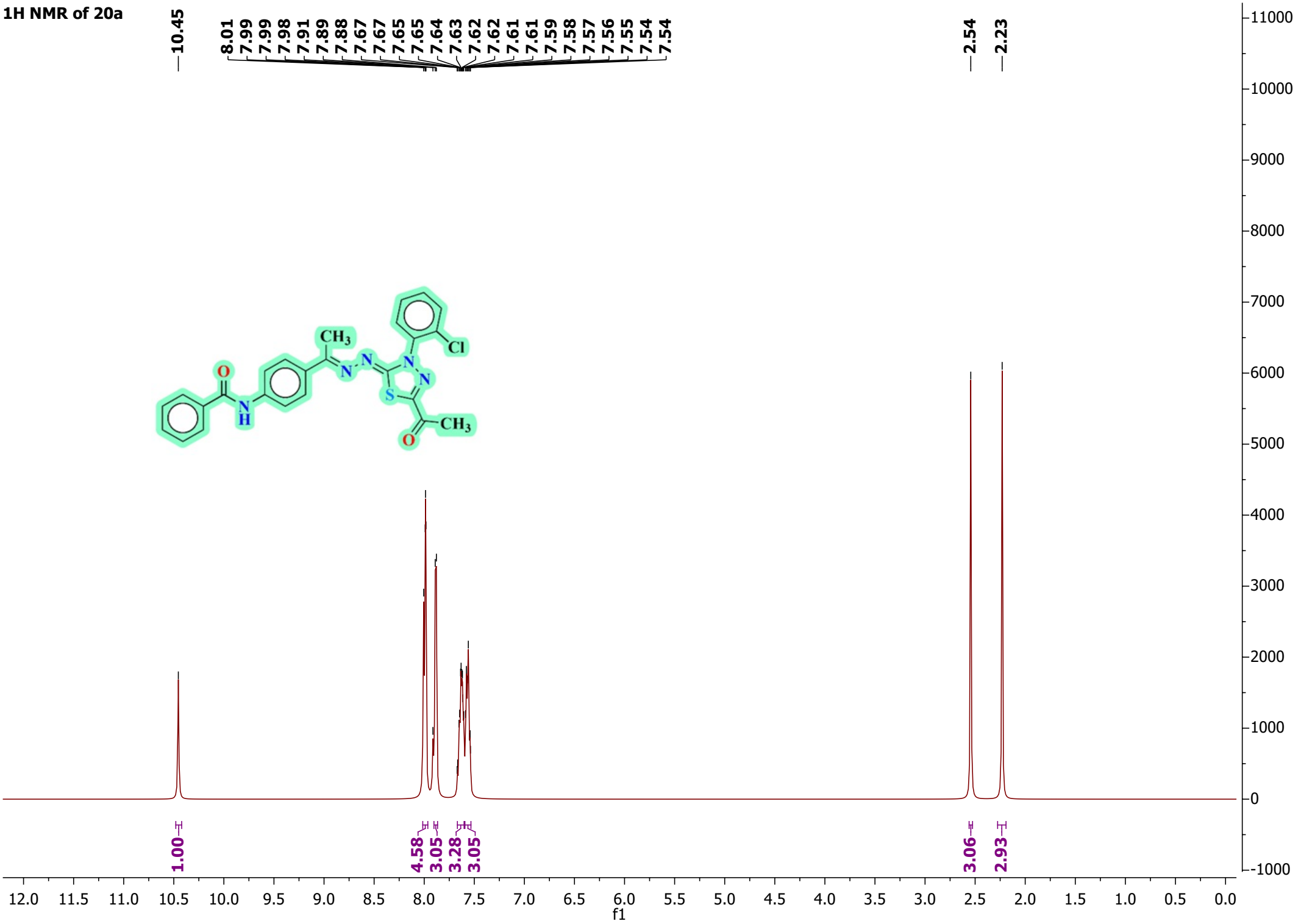


ibrahim-hassan-wa6 #190-195 RT: 3.20-3.28 AV: 6 SB: 26 1.21-1.34 , 0.87-1.14 NL: 8.99E1
T: {0,0} + c EI Full ms [40.00-1000.00]

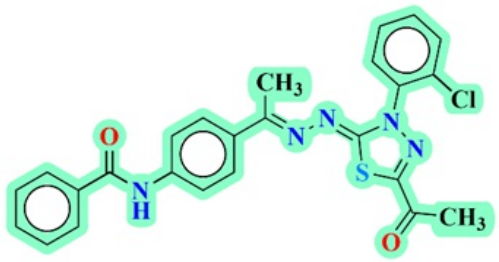



PH5
IR of compound 20a
% Transmittance
Wavenumbers (cm-1)
3353.51
3062.52
3001.40
2919.22
2761.11
1675.06
1655.76
1600.56
1561.30
1522.16
1485.28
1447.46
1403.94
1361.82
1319.05
1272.70
1256.62
1181.76
1128.73
1105.03
1073.65
1049.18
1025.30
1000.31
960.01
898.26
830.57
796.58
761.00
718.17
692.31
662.93
636.46
614.10
586.60
572.80
520.69
502.25
475.70
449.70

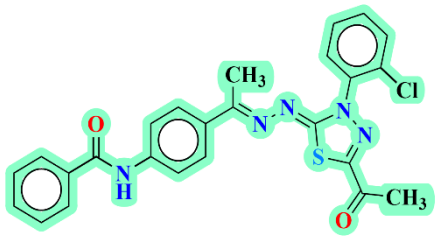
1H NMR of 20a


10.45
8.01
7.99
7.99
7.98
7.91
7.89
7.88
7.67
7.67
7.65
7.65
7.64
7.63
7.62
7.62
7.61
7.61
7.59
7.58
7.57
7.56
7.55
7.54
7.54
2.54
2.23
CH3
Cl
O
N
H
N
N
N
N
S
CH3
O
1.00
4.58
3.05
3.28
3.05
3.06
2.93
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
12.0
11.5
11.0
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1

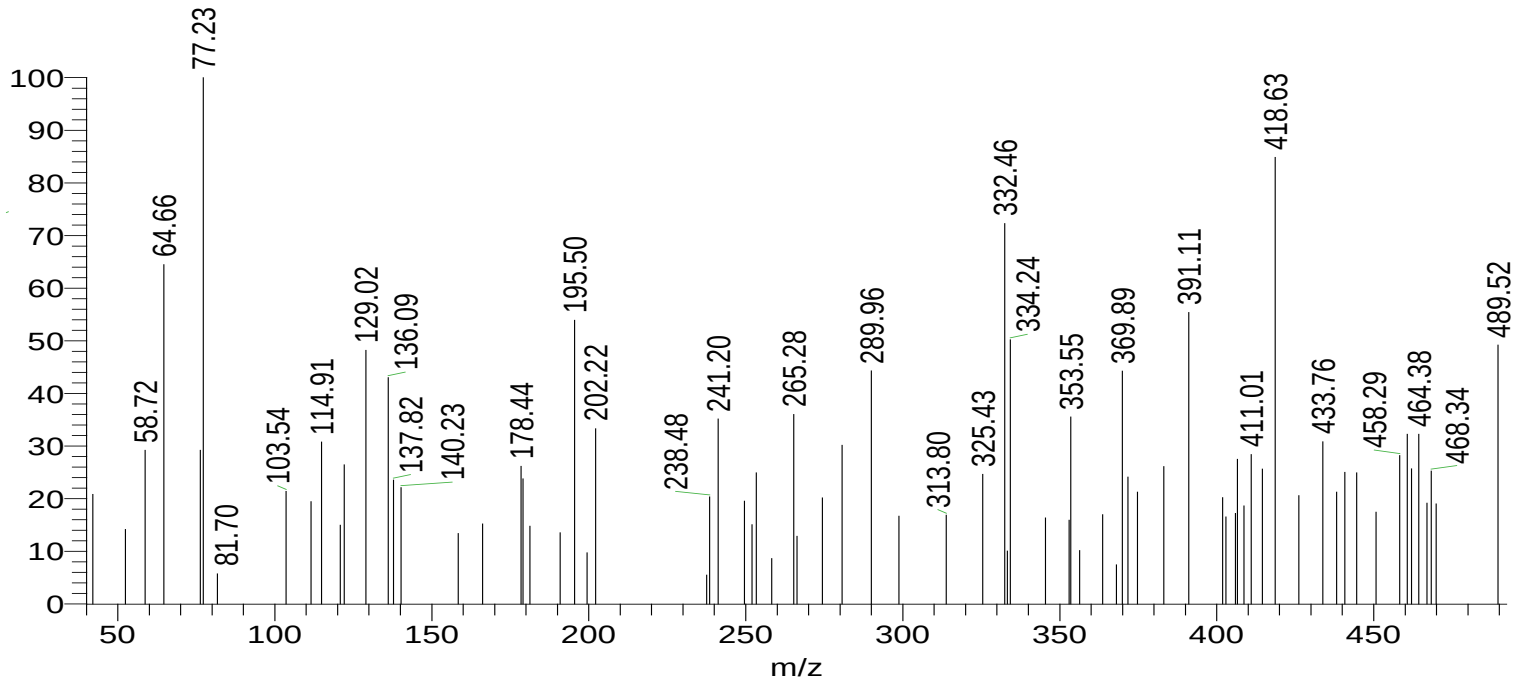
13C NMR of 20a



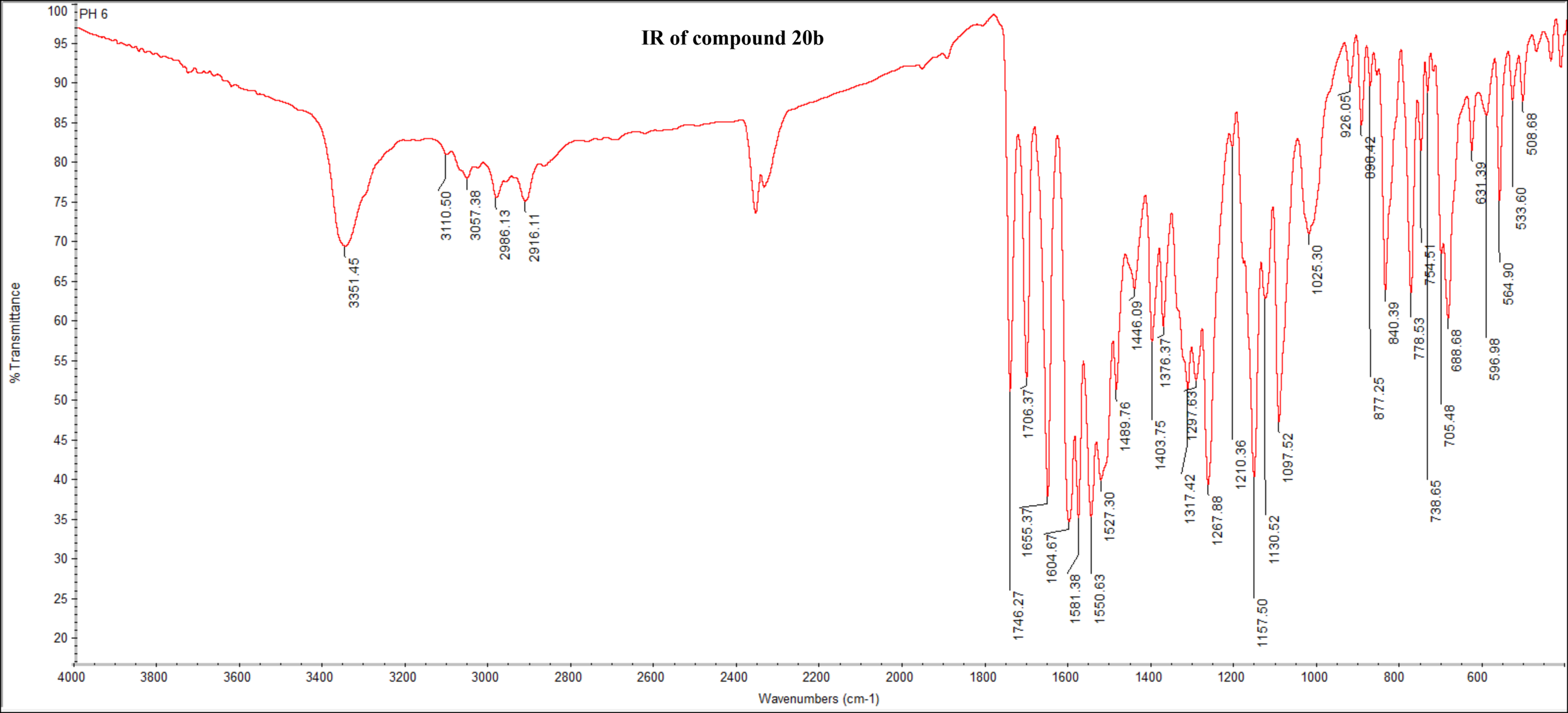
Mass spec. of compound 20a



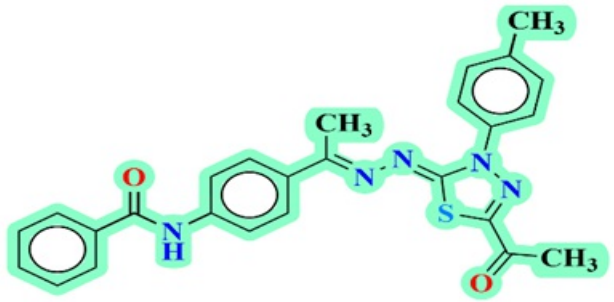
ibrahim-hassan-PH-5 #182 RT: 3.06 AV: 1 SB: 26 1.21-1.34 , 0.87-1.14 NL: 4.71E2
T: {0,0} + c EI Full ms [40.00-1000.00]



IR of compound 20b

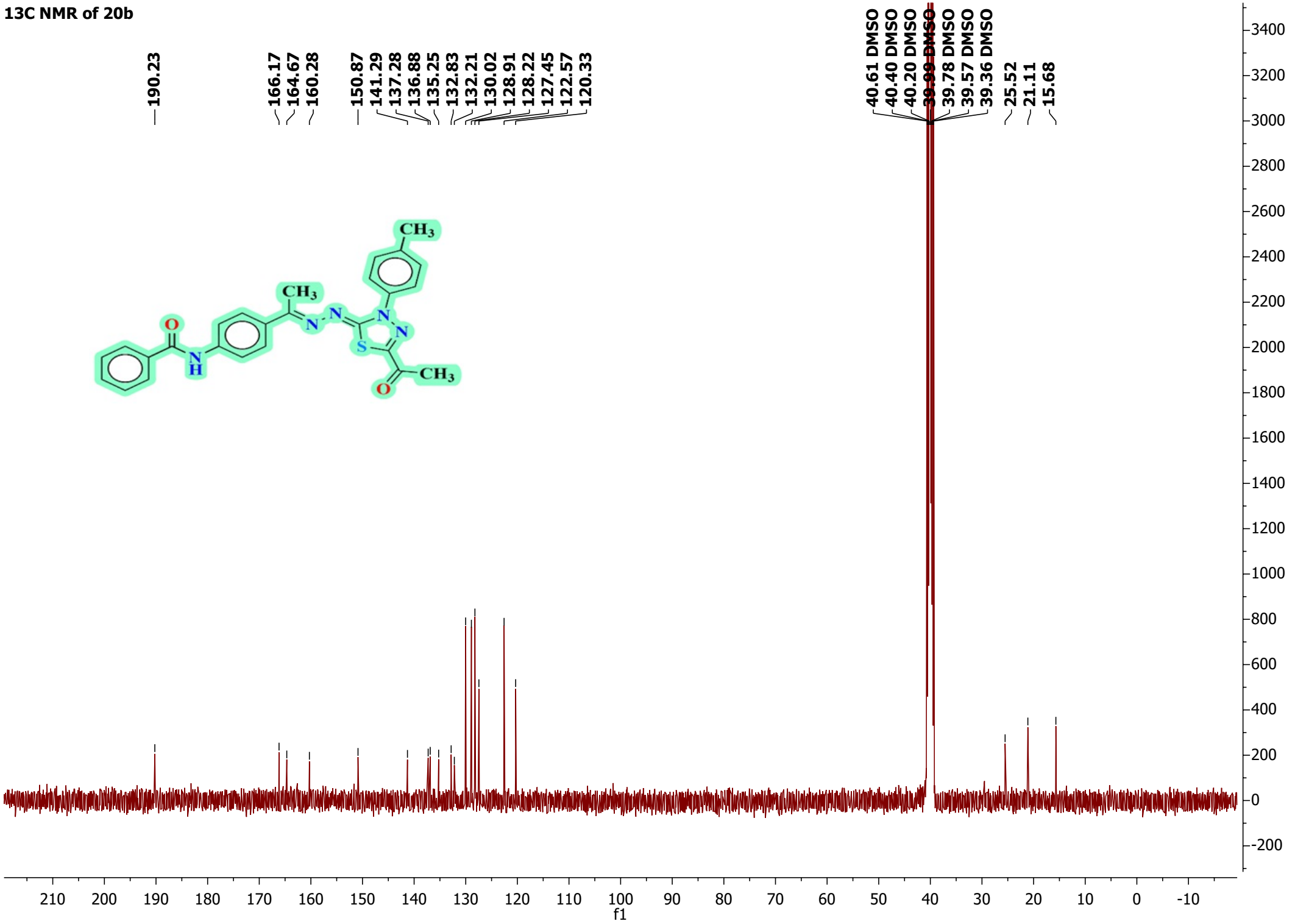


1H NMR of 20b

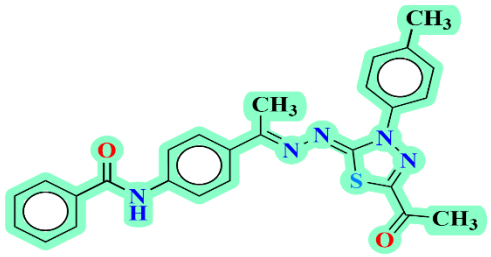

CH3
O
N
H
N
N
N
N
S
CH3
O

10.46
8.00
7.99
7.95
7.93
7.91
7.65
7.63
7.61
7.58
7.57
7.55
7.42
7.40
2.61
2.43
2.40
1.05
8.26
1.01
2.02
1.93
3.03
5.92
f1

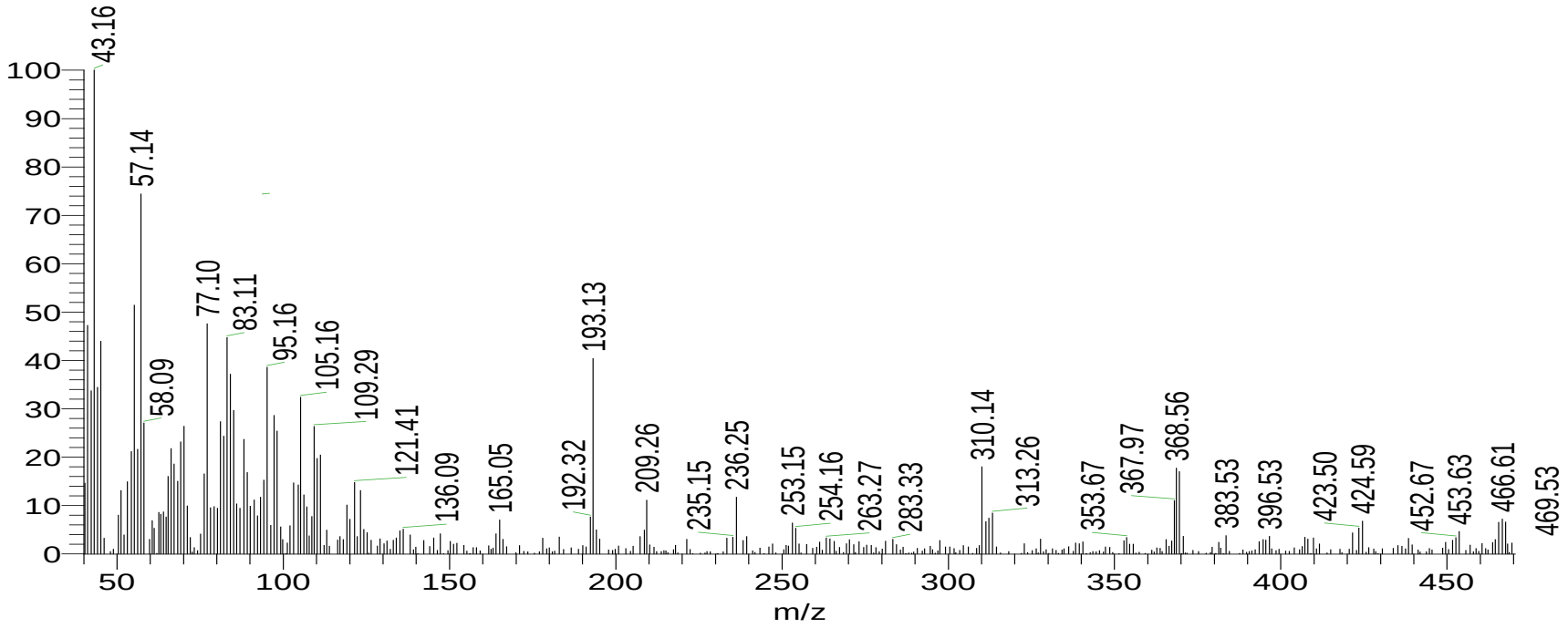
13C NMR of 20b



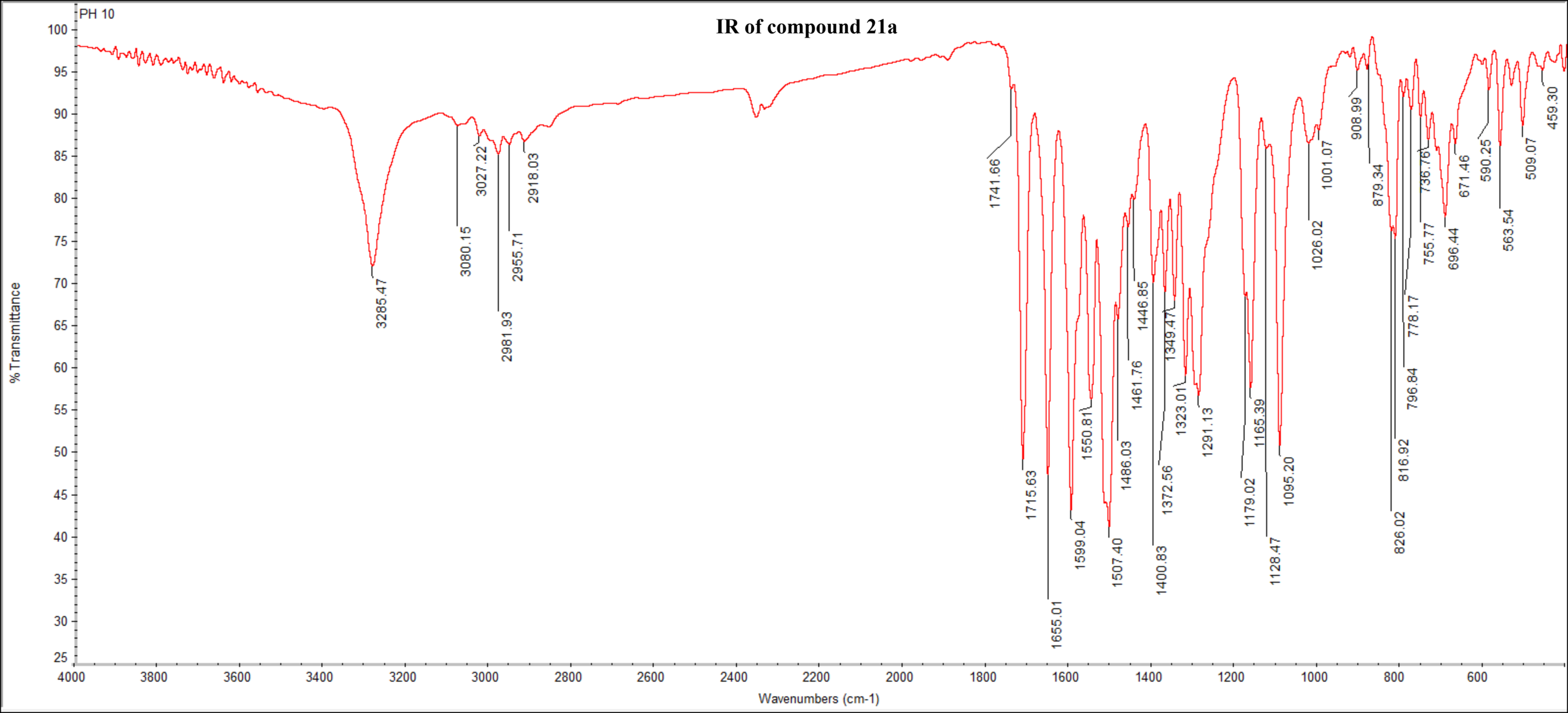
Mass spec. of compound 20b



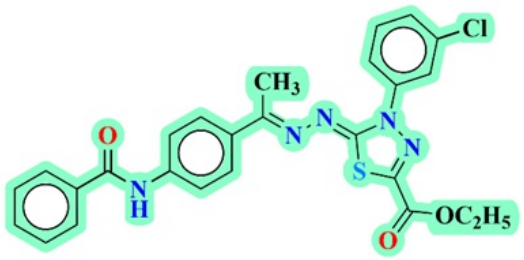
ibrahim-hassan-PH-7 #136-154 RT: 2.29-2.59 AV: 19 SB: 26 1.21-1.34 , 0.87-1.14 NL: 2.28E3
T: {0,0} + c EI Full ms [40.00-1000.00]



IR of compound 21a



1H NMR of 21a

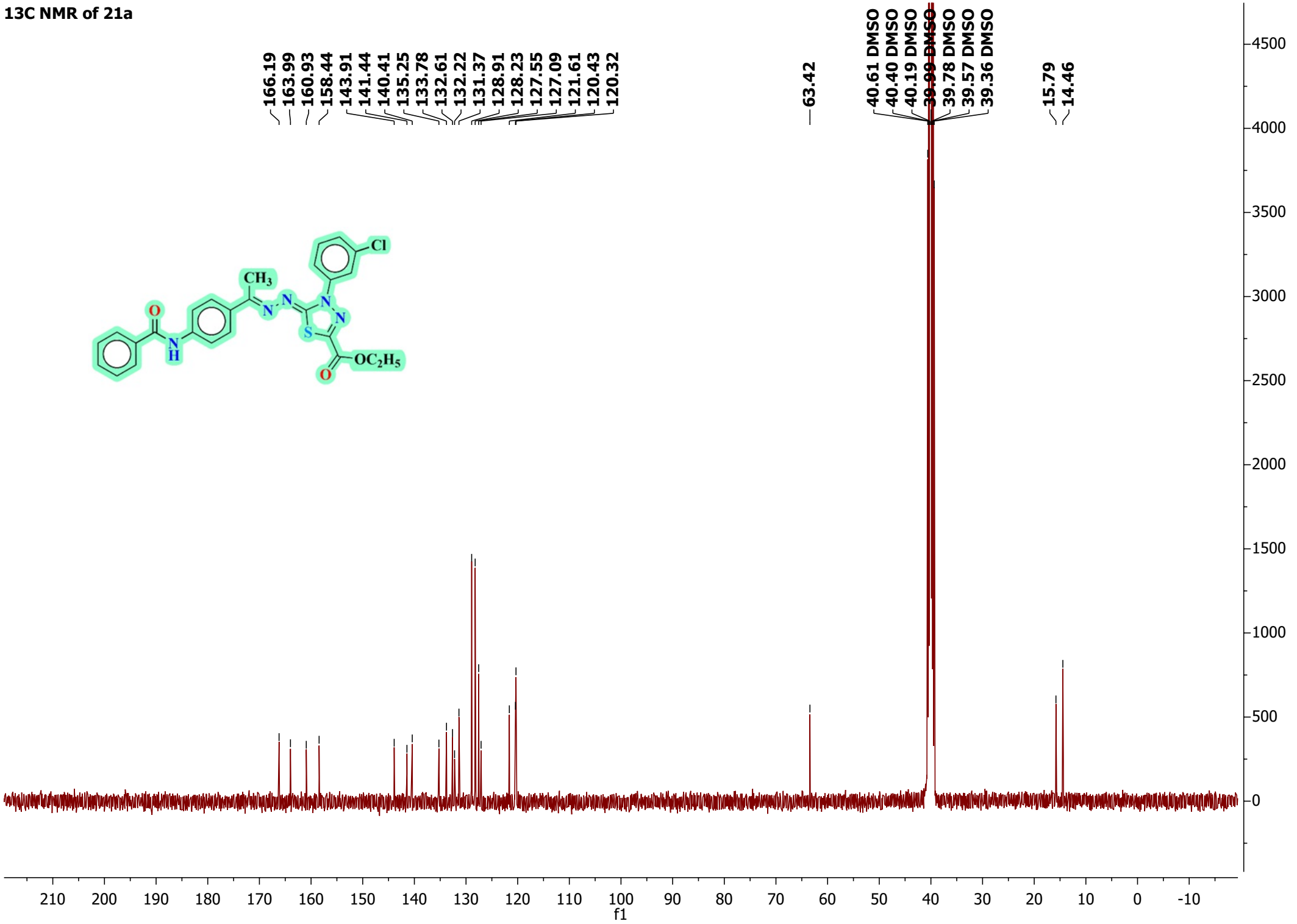


10.47 8.22 8.22 8.21 8.03 8.02 8.01 8.00 7.99 7.99 7.98 7.91 7.65 7.64 7.63 7.63 7.62 7.62 7.61 7.60 7.59 7.57 7.57 7.55 7.55 7.55 7.49 7.49 7.49 7.47 7.47 7.47 7.47 4.44 4.43 4.41 4.39 2.46 1.38 1.36 1.34

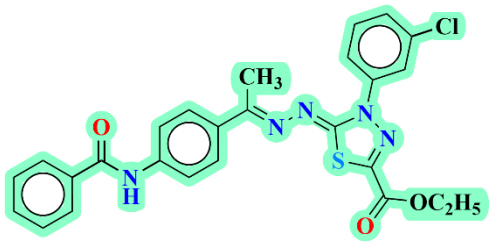
1.00 0.96 3.22 4.07 4.22 1.00 2.03 3.04 3.04

f1

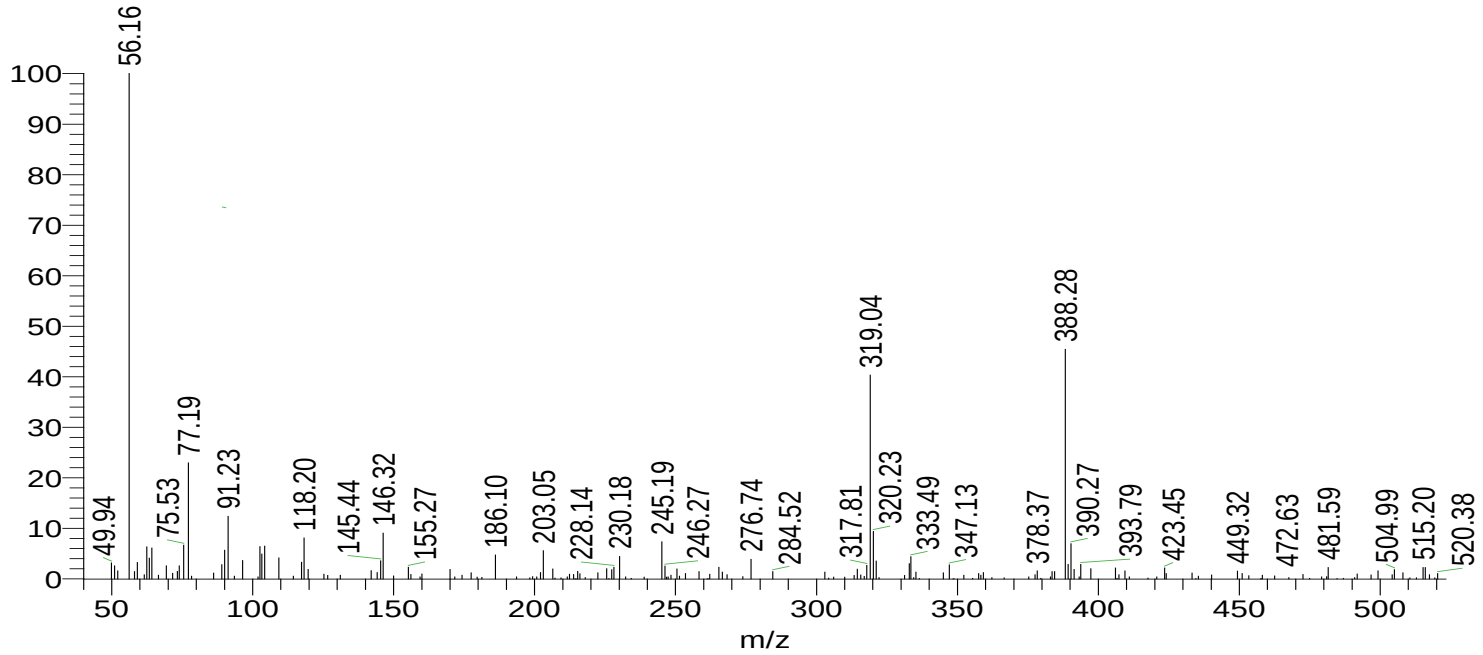
13C NMR of 21a

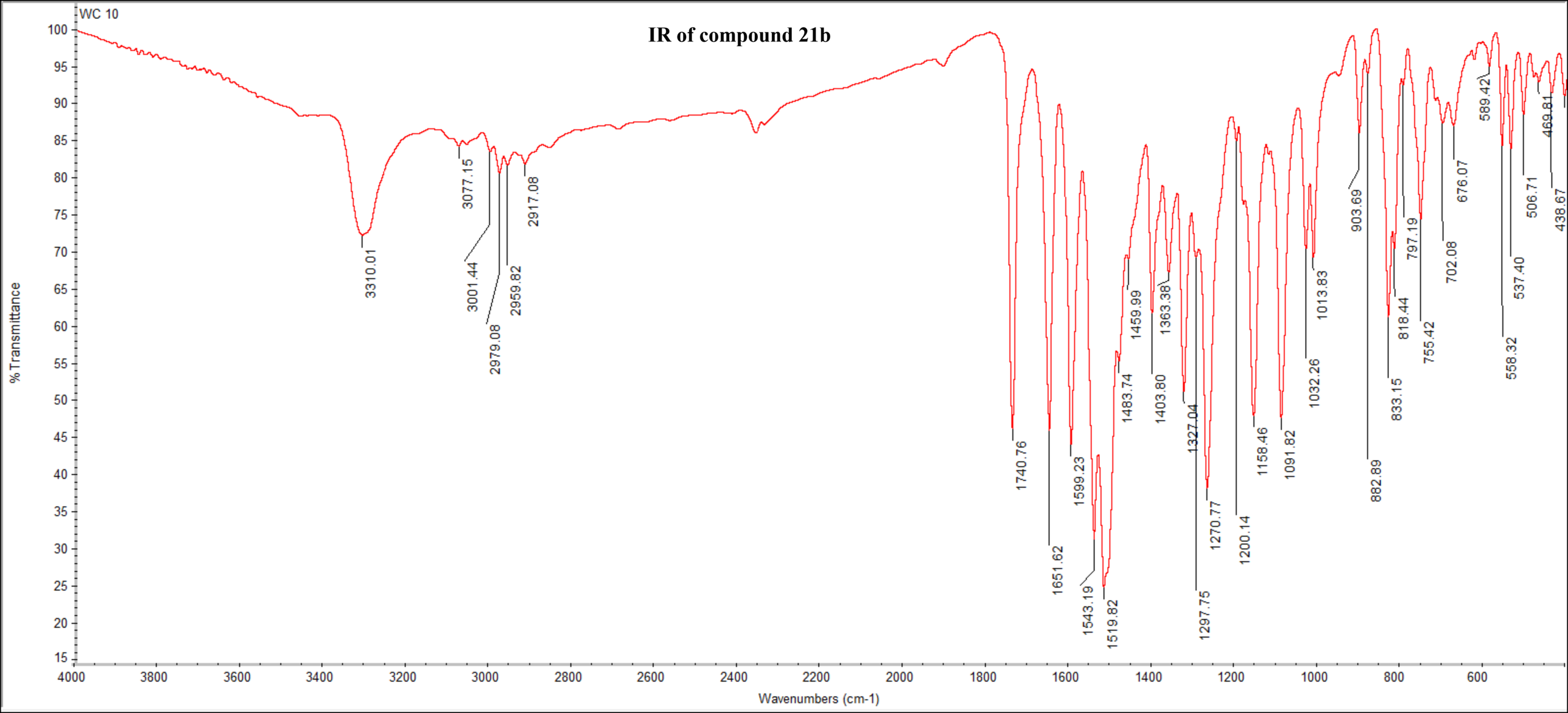


Mass spec. of compound 21a

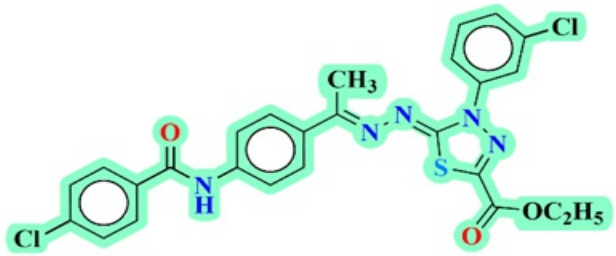


ibrahim-hassan-PH-8 #231 RT: 3.88 AV: 1 SB: 26 1.21-1.34 , 0.87-1.14 NL: 2.03E4
T: {0,0} + c EI Full ms [40.00-1000.00]



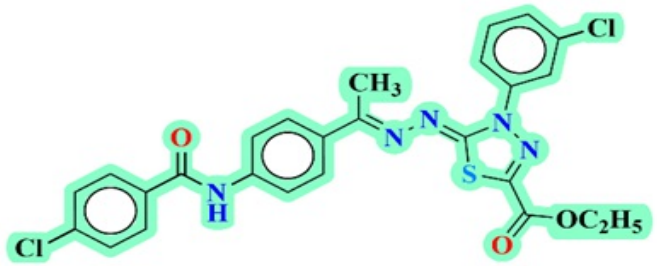

WC 10
IR of compound 21b
% Transmittance
Wavenumbers (cm-1)
3310.01
3077.15
3001.44
2979.08
2959.82
2917.08
1740.76
1651.62
1599.23
1543.19
1519.82
1483.74
1459.99
1403.80
1363.38
1327.04
1297.75
1270.77
1200.14
1158.46
1091.82
1032.26
1013.83
903.69
882.89
833.15
818.44
797.19
755.42
702.08
676.07
589.42
558.32
537.40
506.71
469.81
438.67

1H NMR of 21b


Cl
CH3
O
N
H
N
N
N
S
N
Cl
O
OC2H5

10.51
8.04
8.04
8.03
8.02
8.01
8.01
7.92
7.90
7.89
7.88
7.88
7.87
7.86
7.66
7.65
7.65
7.64
7.63
7.62
7.40
7.37
4.44
4.42
4.40
4.38
2.42
2.39
1.37
1.35
1.34
1.05
2.00
6.04
1.97
2.01
2.02
5.85
2.98
f1

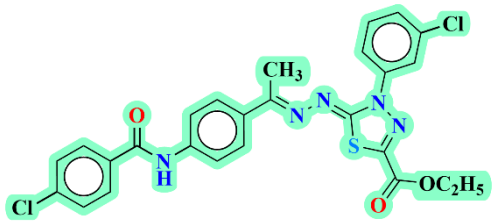
13C NMR of 21b



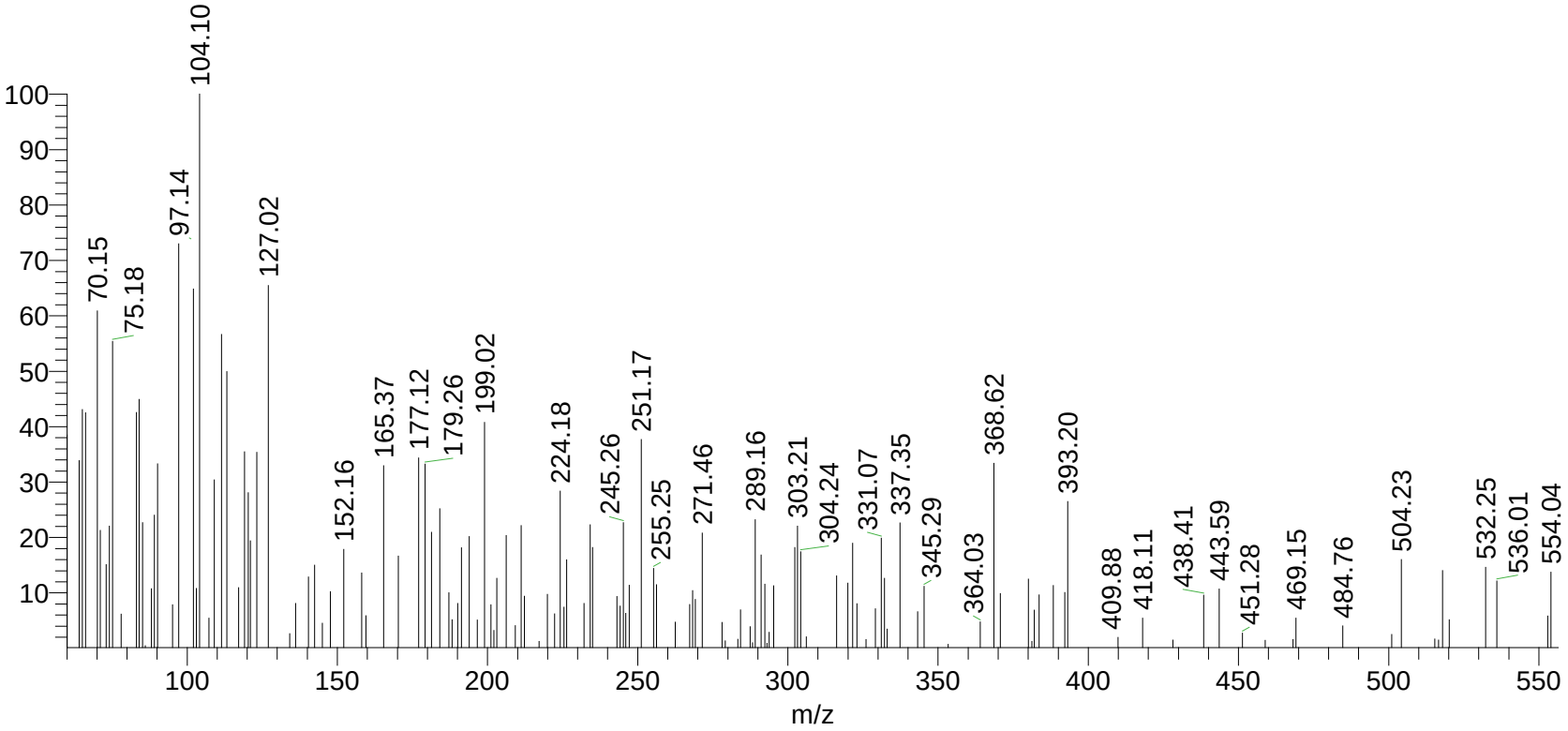
164.46
160.14
158.62
142.90
141.05
137.25
136.84
133.94
133.01
130.20
130.01
128.99
127.45
122.60
120.42
63.24
40.62 DMSO
40.41 DMSO
40.20 DMSO
39.99 DMSO
39.78 DMSO
39.58 DMSO
39.37 DMSO
21.10
15.68
14.48

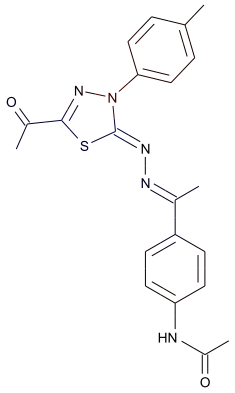
f1

Mass spec. of compound 21b



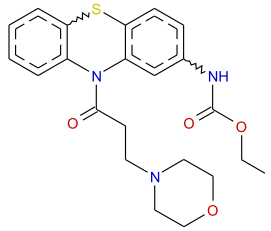
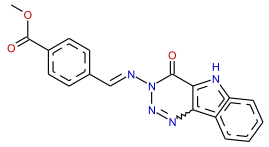
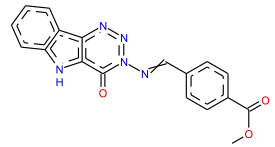
ibrahim-hassan-wc10 #25 RT: 0.44 AV: 1 SB: 22 0.22-0.37 , 0.28-0.47 NL: 3.53E3
T: {0,0} + c EI Full ms [40.00-1000.00]





$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Structural Similar Compounds

Name	MORICIZINE	139953-78-9	3-(4'-Acetylbenzylidenamino)-5H-1;2;3-triazin-[5;4b]indol-4-one
Structure			
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen
Distance	0.572	0.599	0.599
Reference	PDR 1994	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Mutagenesis 7(1):37-39; 1992

Model Prediction

Prediction: Non-Mutagen
Probability: 0.696
Enrichment: 1.25
Bayesian Score: -1.92
Mahalanobis Distance: 12
Mahalanobis Distance p-value: 0.000662

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

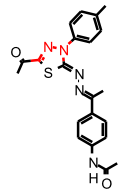
Model Applicability

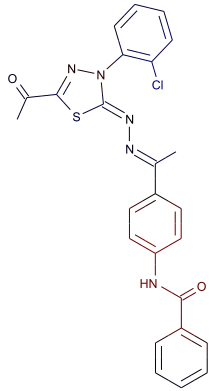
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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	-1325991669	 [*]N1[*][*]C(=N1)[*]	0.362	7 out of 8



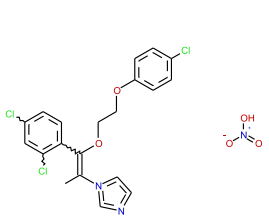
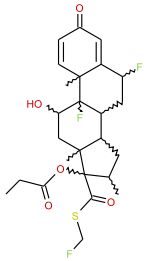
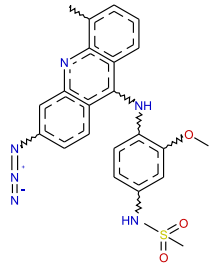
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.666
Enrichment: 1.19
Bayesian Score: -3.01
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	0.619	0.628	0.628
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	PDR 1994	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

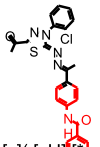
Model Applicability

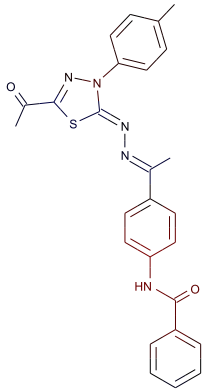
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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(=O)N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.447	22 out of 24



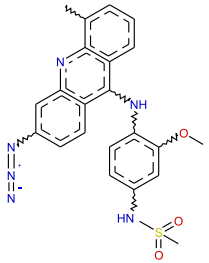
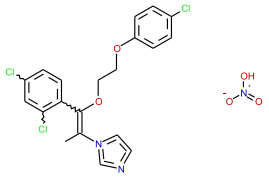
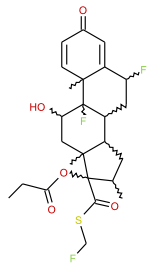
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Mutagen
Probability: 0.731
Enrichment: 1.31
Bayesian Score: -0.416
Mahalanobis Distance: 12.4
Mahalanobis Distance p-value: 0.000145

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	110004-69-8	83621-06-1	FLUTICASONE
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.610	0.619	0.620
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	PDR 1994

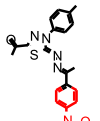
Model Applicability

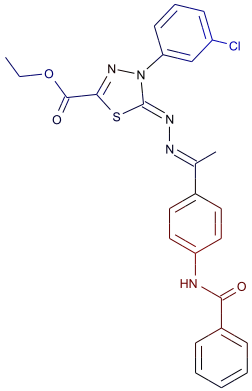
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

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



$C_{26}H_{22}ClN_5O_3S$

Molecular Weight: 520.00258

ALogP: 5.883

Rotatable Bonds: 8

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.54

Enrichment: 0.967

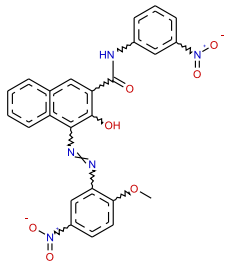
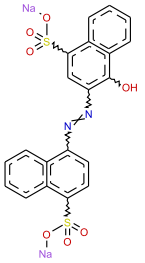
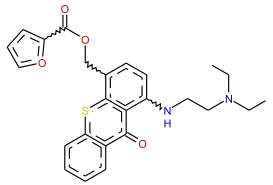
Bayesian Score: -6.58

Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.00137

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	6471-49-4	3567-69-9	HYCANTHONE FUROATE
Structure			
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Non-Mutagen	Mutagen
Distance	0.660	0.669	0.671
Reference	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	EMIC

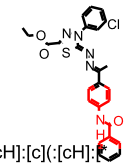
Model Applicability

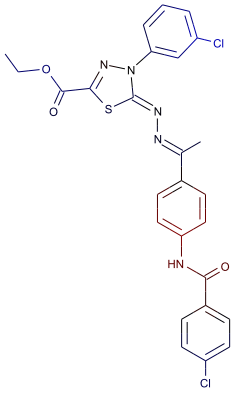
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

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



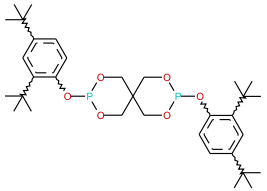
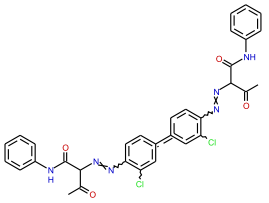
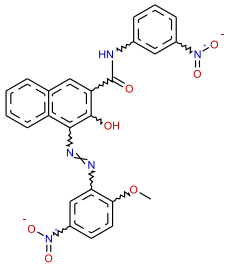
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.495
Enrichment: 0.887
Bayesian Score: -7.62
Mahalanobis Distance: 11.9
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	26741-53-7	6358-85-6	6471-49-4
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Distance	0.664	0.682	0.700
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

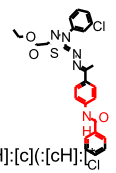
Model Applicability

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

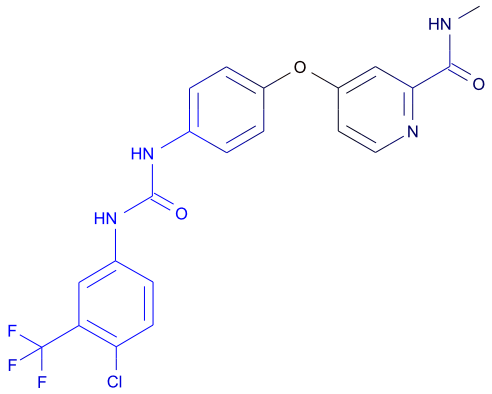
1. All properties and OPS components are within expected ranges.

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

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

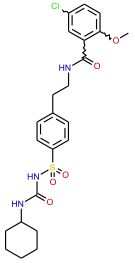
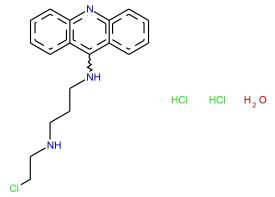
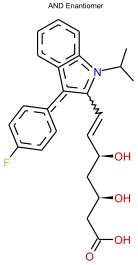
Rotatable Bonds: 6

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.

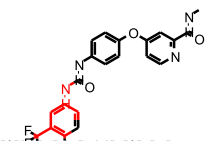
Model Applicability

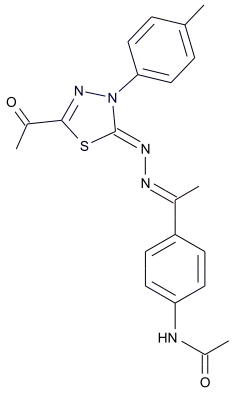
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



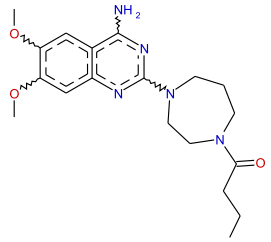
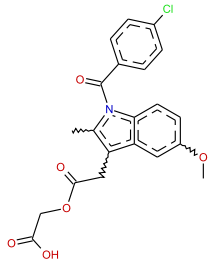
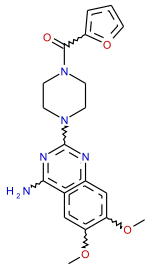
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.517
Enrichment: 0.984
Bayesian Score: -0.815
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.0016

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	0.601	0.603	0.615
Reference	Kiso to Rinsho 17:914-924; 1983	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 17:57-62; 1979

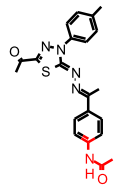
Model Applicability

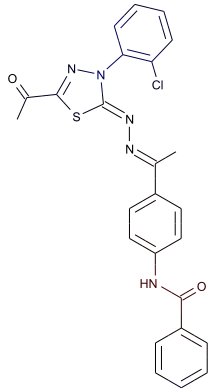
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: 4.7477. 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	1626825020	 [*]:[c](:[*])NC(=O)C	0.271	1 out of 1



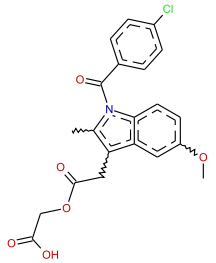
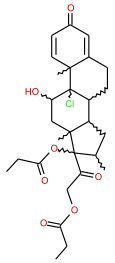
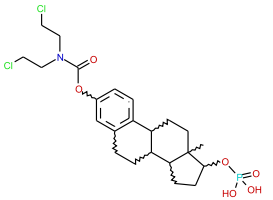
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.494
Enrichment: 0.938
Bayesian Score: -1.48
Mahalanobis Distance: 12.3
Mahalanobis Distance p-value: 7.15e-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	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	0.639	0.662	0.663
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 20(6):1219-1236; 1980

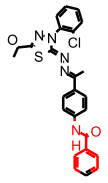
Model Applicability

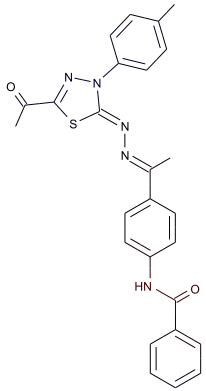
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC17 out of range. Value: 3.0225. Training min, max, SD, explained variance: -2.7025, 2.8536, 1.067, 0.0167.

Feature Contribution

Top features for positive contribution

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



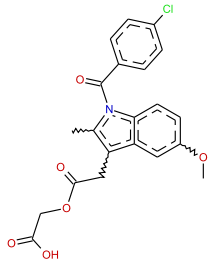
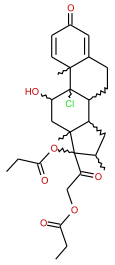
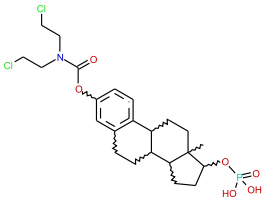
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.526
Enrichment: 0.999
Bayesian Score: -0.59
Mahalanobis Distance: 12.1
Mahalanobis Distance p-value: 1.77e-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	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	0.623	0.663	0.665
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 20(6):1219-1236; 1980

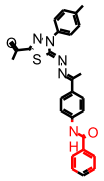
Model Applicability

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

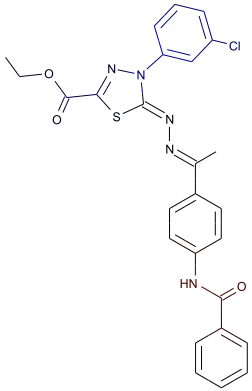
1. All properties and OPS components are within expected ranges.

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]:[c H]:[*]:[cH]:[cH]:1	0.441	3 out of 3

TOPKAT_Developmental_Toxicity_Potential



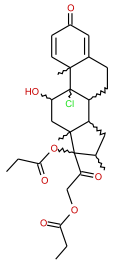
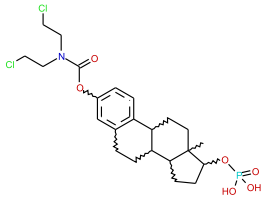
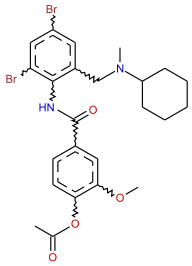
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.437
Enrichment: 0.832
Bayesian Score: -3.16
Mahalanobis Distance: 13.5
Mahalanobis Distance p-value: 3.92e-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	Beclomethasone Dipropionate	Estramustine Phosphate Disodium (Free acid form)	Brovanexine .HCl (Free base form)
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.659	0.660	0.681
Reference	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 20(6):1219-1236; 1980	Kiso to Rinsho 16(13):7179-7195; 1982

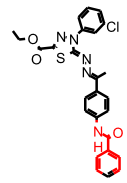
Model Applicability

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

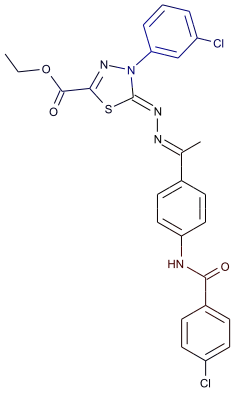
1. All properties and OPS components are within expected ranges.

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]:[*]:[cH]:[cH]:1	0.441	3 out of 3

TOPKAT_Developmental_Toxicity_Potential



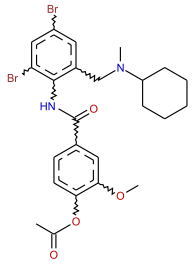
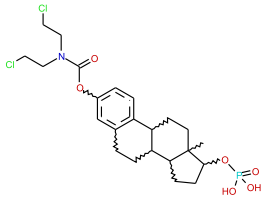
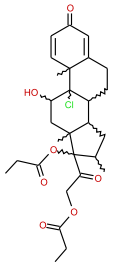
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.442
Enrichment: 0.84
Bayesian Score: -3.03
Mahalanobis Distance: 12.4
Mahalanobis Distance p-value: 4.85e-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	Brovanexine .HCl (Free base form)	Estramustine Phosphate Disodium (Free acid form)	Beclomethasone Dipropionate
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.677	0.688	0.716
Reference	Kiso to Rinsho 16(13):7179-7195; 1982	Oyo Yakuri 20(6):1219-1236; 1980	Oyo Yakuri 18(6):1021-1038; 1979

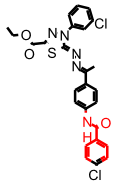
Model Applicability

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.

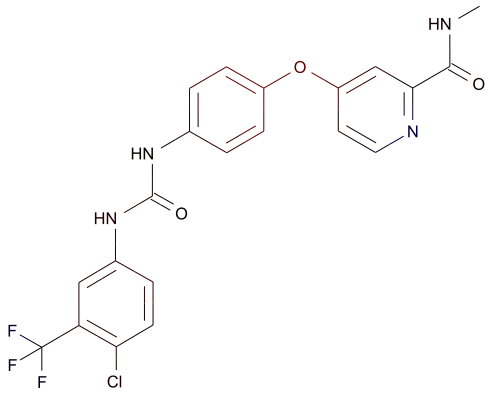
Feature Contribution

Top features for positive contribution

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

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

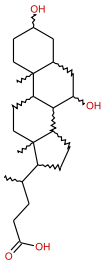
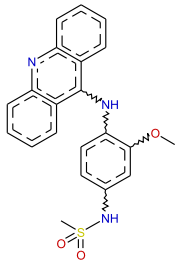
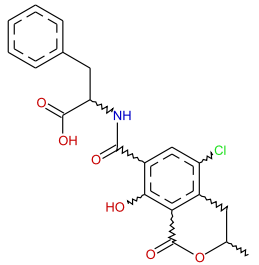
Bayesian Score: 1.15

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

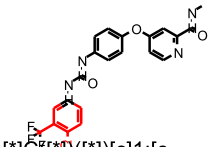
Model Applicability

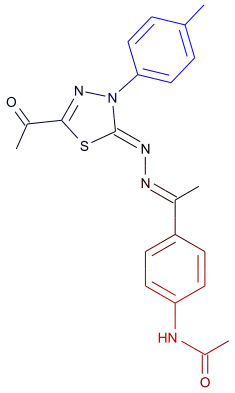
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



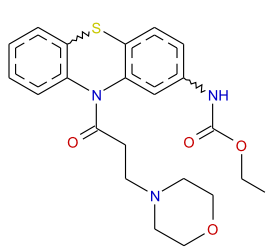
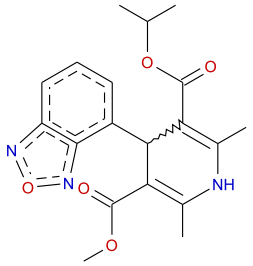
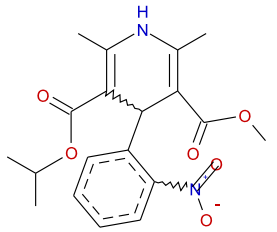
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.208
Enrichment: 0.649
Bayesian Score: -4.72
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.0666

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Isradipine	Nisoldipine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.575	0.618	0.633
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

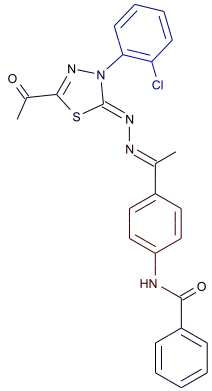
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



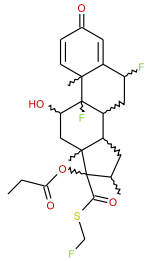
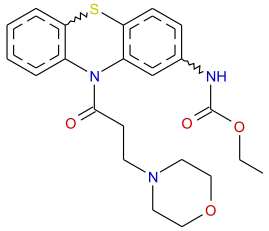
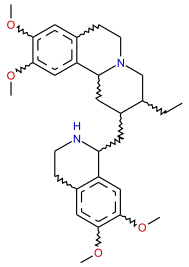
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.206
Enrichment: 0.642
Bayesian Score: -5.37
Mahalanobis Distance: 13.7
Mahalanobis Distance p-value: 4e-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	Fluticasone	Moricizine	Emetine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.632	0.686	0.692
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

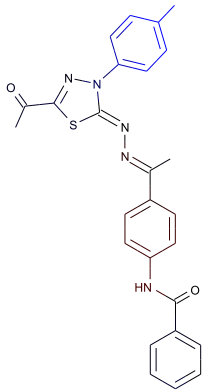
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC28 out of range. Value: 3.464. Training min, max, SD, explained variance: -2.8298, 3.1935, 1.043, 0.0111.
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



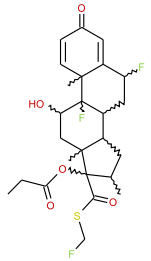
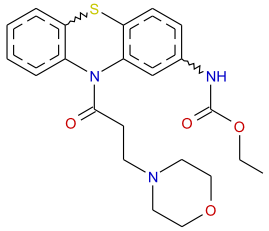
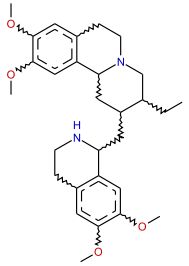
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.206
Enrichment: 0.644
Bayesian Score: -7.46
Mahalanobis Distance: 12.4
Mahalanobis Distance p-value: 0.00322

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Emetine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.628	0.670	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

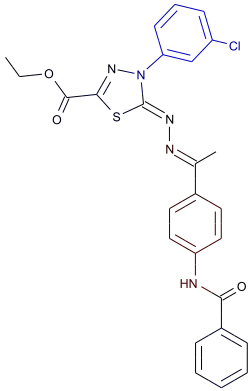
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



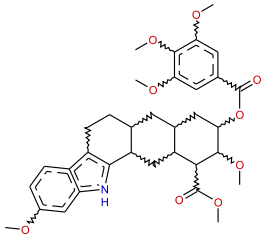
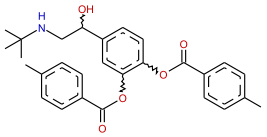
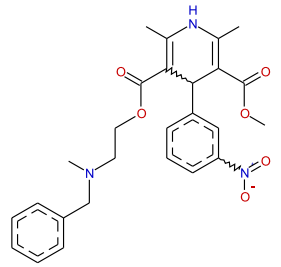
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.205
Enrichment: 0.64
Bayesian Score: -6.85
Mahalanobis Distance: 20
Mahalanobis Distance p-value: 4.36e-018

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

Structural Similar Compounds

Name	Reserpine	Bitolterol	Nicardipine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.693	0.733	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

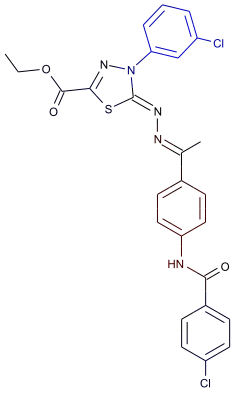
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC8 out of range. Value: -4.0148. Training min, max, SD, explained variance: -3.693, 7.8709, 1.782, 0.0323.
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



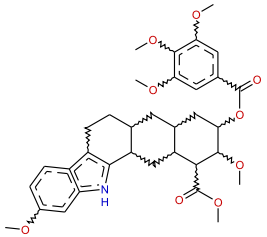
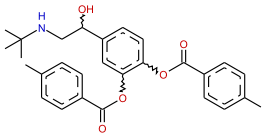
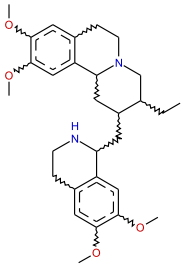
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.206
Enrichment: 0.643
Bayesian Score: -7.32
Mahalanobis Distance: 14.2
Mahalanobis Distance p-value: 5.77e-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	Reserpine	Bitolterol	Emetine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.674	0.766	0.785
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 PC8 out of range. Value: -4.4388. Training min, max, SD, explained variance: -3.693, 7.8709, 1.782, 0.0323.
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](:[*]):[*]

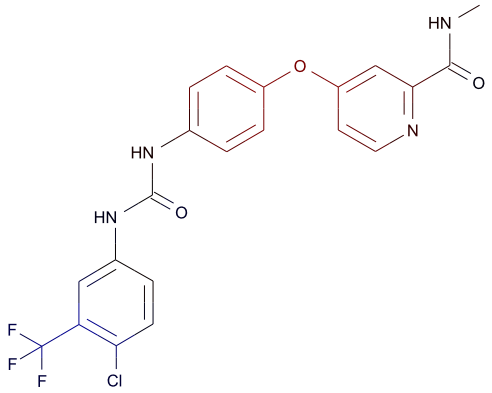
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

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

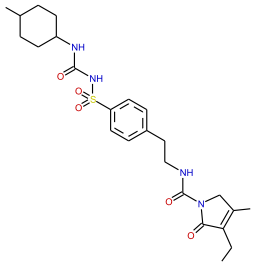
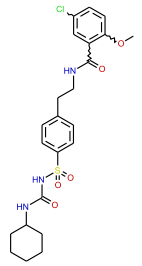
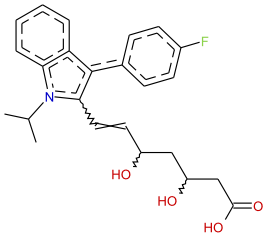
Bayesian Score: -0.321

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

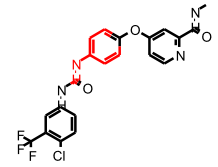
Model Applicability

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.

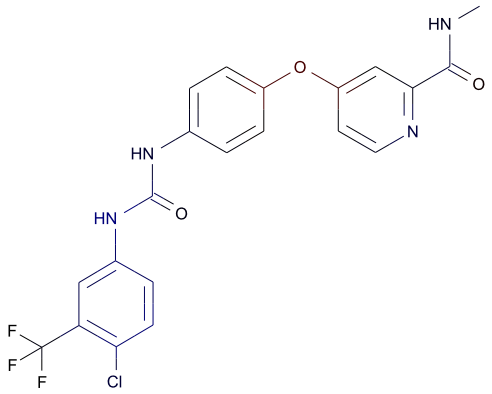
Feature Contribution

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

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

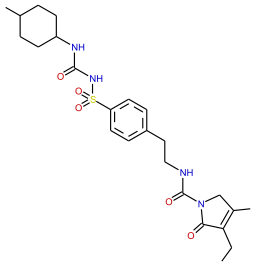
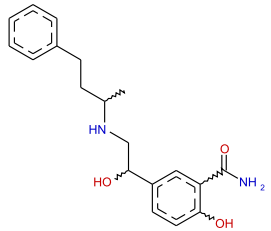
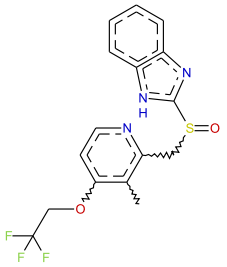
Bayesian Score: -3.89

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

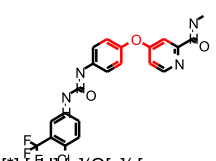
Model Applicability

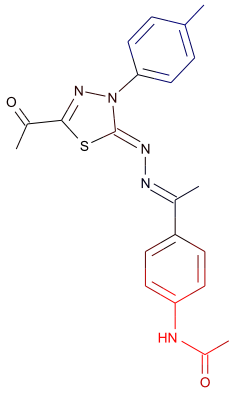
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



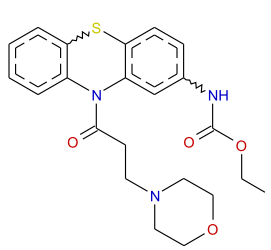
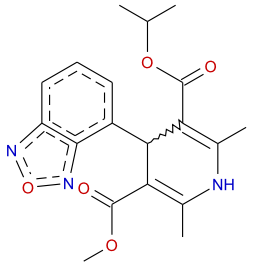
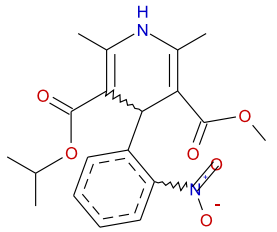
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.376
Enrichment: 1.28
Bayesian Score: 2.3
Mahalanobis Distance: 16.3
Mahalanobis Distance p-value: 2.15e-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	Moricizine	Isradipine	Nisoldipine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.555	0.599	0.604
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

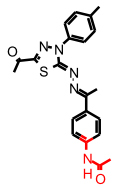
Model Applicability

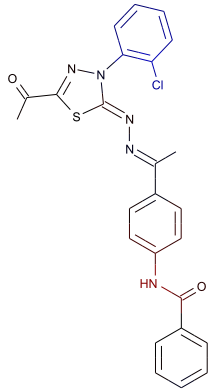
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC9 out of range. Value: 5.948. Training min, max, SD, explained variance: -5.0113, 5.5609, 1.7, 0.0303.

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



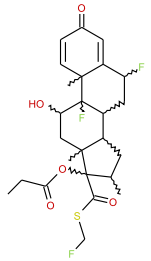
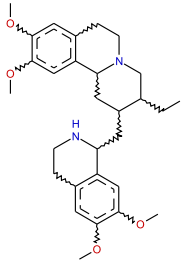
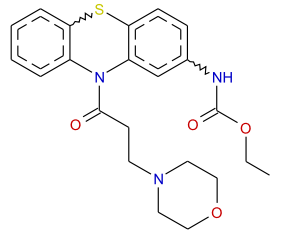
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.249
Enrichment: 0.846
Bayesian Score: -2.34
Mahalanobis Distance: 17.4
Mahalanobis Distance p-value: 7.24e-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	0.679
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

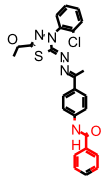
Model Applicability

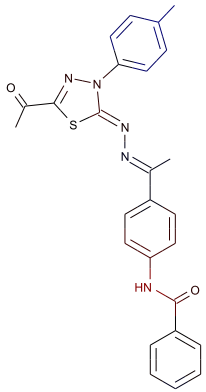
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

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



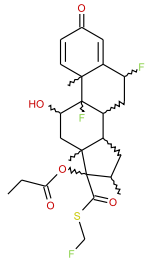
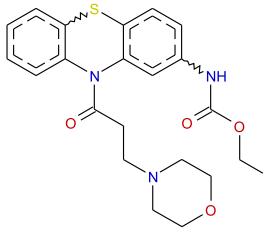
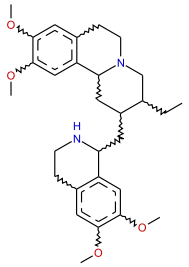
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.285
Enrichment: 0.97
Bayesian Score: -0.774
Mahalanobis Distance: 18.1
Mahalanobis Distance p-value: 1.74e-014

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

Structural Similar Compounds

Name	Fluticasone	Moricizine	Emetine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.614	0.655	0.669
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

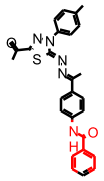
Model Applicability

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

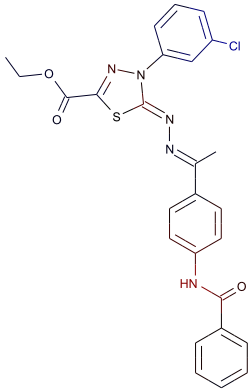
1. All properties and OPS components are within expected ranges.

Feature Contribution

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

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



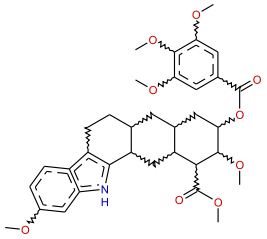
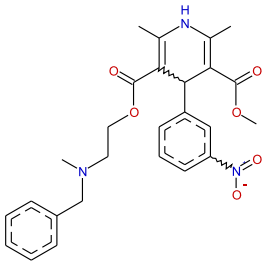
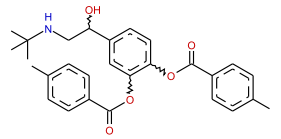
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.247
Enrichment: 0.84
Bayesian Score: -2.42
Mahalanobis Distance: 15.2
Mahalanobis Distance p-value: 4.65e-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	Reserpine	Nicardipine	Bitolterol
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.689	0.728	0.734
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

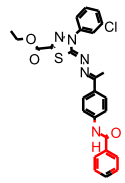
Model Applicability

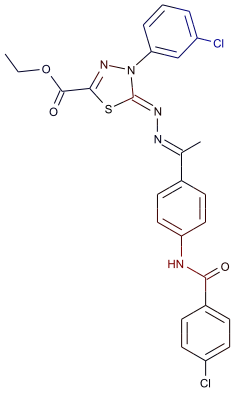
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



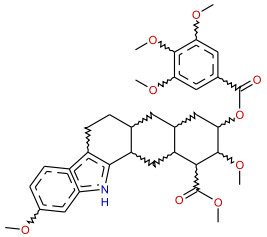
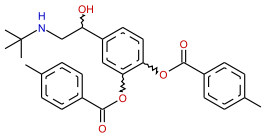
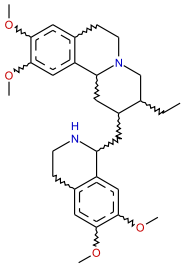
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.262
Enrichment: 0.89
Bayesian Score: -1.75
Mahalanobis Distance: 15.3
Mahalanobis Distance p-value: 1.98e-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	Reserpine	Bitolterol	Emetine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.671	0.767	0.780
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

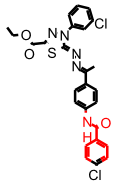
Model Applicability

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.

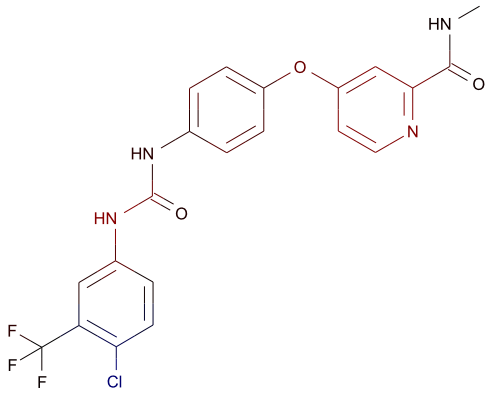
Feature Contribution

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

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

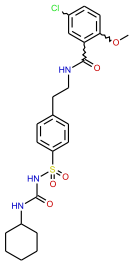
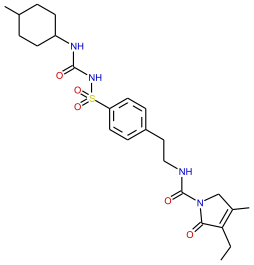
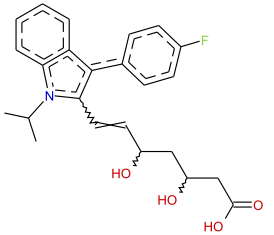
Bayesian Score: 4.21

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

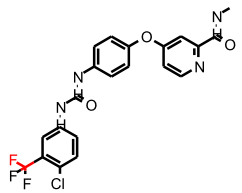
Model Applicability

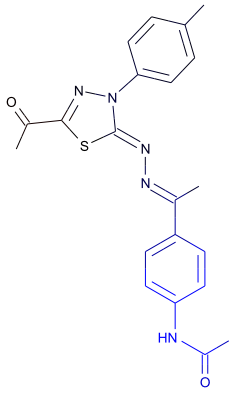
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



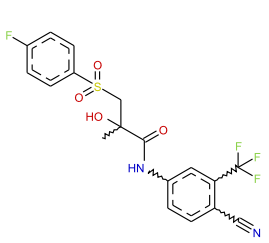
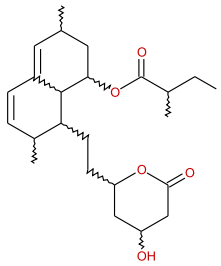
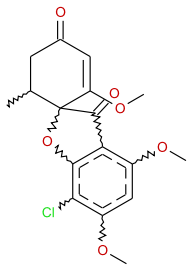
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.145
Enrichment: 0.48
Bayesian Score: -14.1
Mahalanobis Distance: 14.2
Mahalanobis Distance p-value: 1e-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	Bicalutamide	Lovastatin	Griseofulvin
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.740	0.752	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

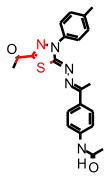
Model Applicability

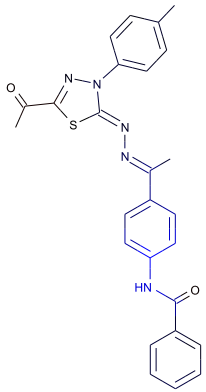
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: 2.4658. Training min, max, SD, explained variance: -3.4599, 2.3291, 1.246, 0.0290.

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



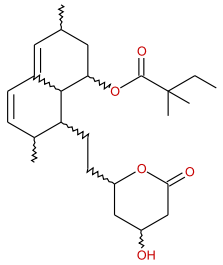
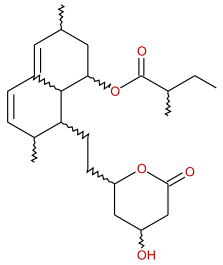
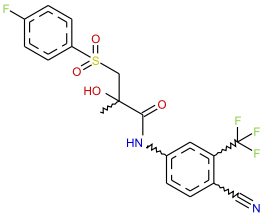
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.146
Enrichment: 0.484
Bayesian Score: -9.63
Mahalanobis Distance: 13.9
Mahalanobis Distance p-value: 1.87e-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	0.736	0.750	0.808
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

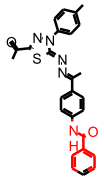
Model Applicability

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

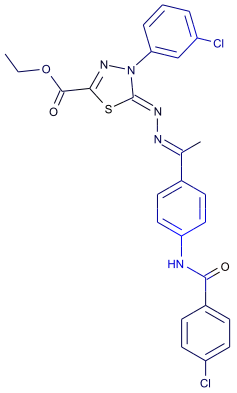
1. OPS PC12 out of range. Value: 3.8138. Training min, max, SD, explained variance: -3.4599, 2.3291, 1.246, 0.0290.

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

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{26}H_{21}Cl_2N_5O_3S$

Molecular Weight: 554.44763

ALogP: 6.547

Rotatable Bonds: 8

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.132

Enrichment: 0.439

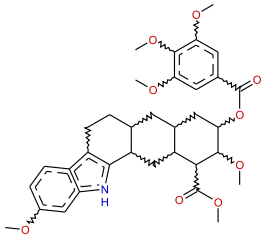
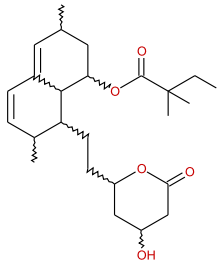
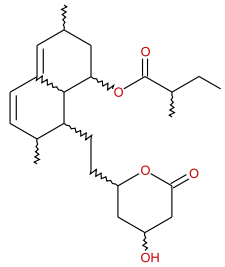
Bayesian Score: -15.5

Mahalanobis Distance: 13.7

Mahalanobis Distance p-value: 2.65e-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	0.697	0.918	0.944
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

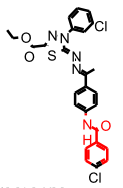
Model Applicability

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

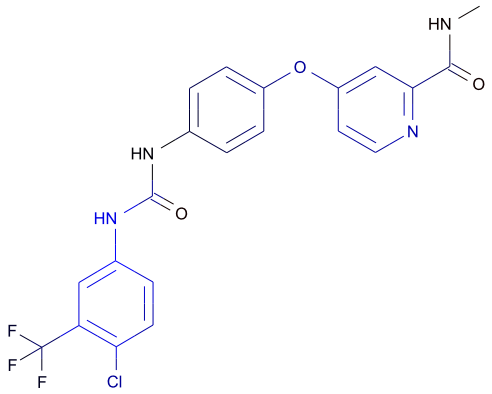
1. All properties and OPS components are within expected ranges.

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]:[c H]:[*]:[cH]:[cH]:1	0.239	2 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

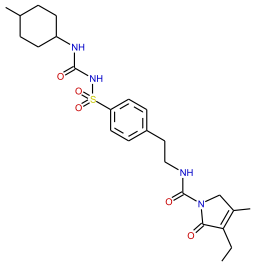
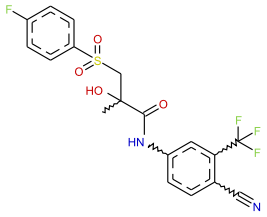
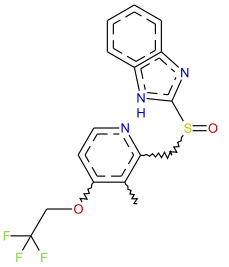
Mahalanobis Distance: 21.3

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

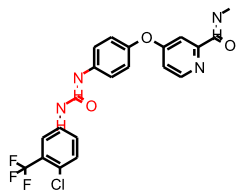
Model Applicability

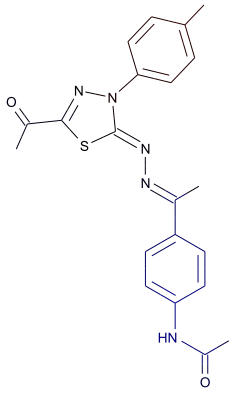
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



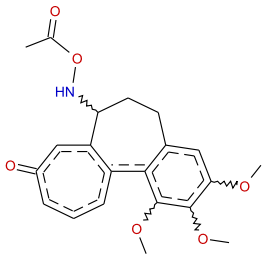
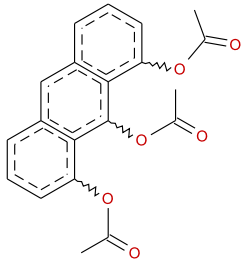
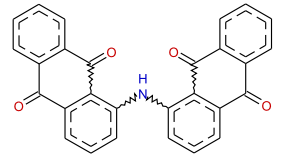
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.732
Enrichment: 1.06
Bayesian Score: -2.9
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 0.0561

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	ANTHRAQUINONE; 1;1'-IMINODI-
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild
Distance	0.621	0.738	0.741
Reference	AJOPAA 31;837;48	BJOPAL 53;819;69	28ZPAK-;125;72

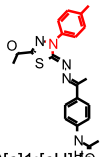
Model Applicability

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

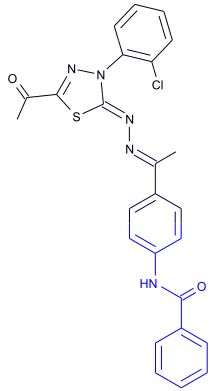
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	633795852	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.294	3 out of 3

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{25}H_{20}ClN_5O_2S$

Molecular Weight: 489.97659

ALogP: 5.409

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.0279

Enrichment: 0.0405

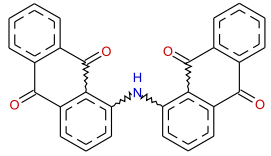
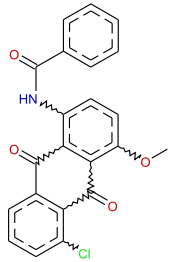
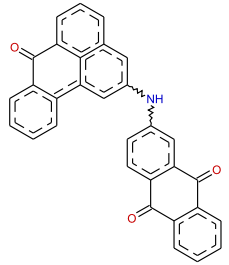
Bayesian Score: -13

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	0.806
Reference	28ZPAK-;125;72	28ZPAK-;90;72	28ZPAK-;126;72

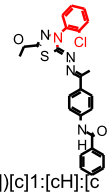
Model Applicability

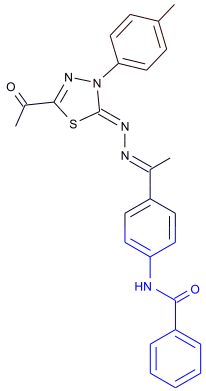
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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	-1881275701	 [*]N([*])[c]1:[cH]:[c H]:[cH]:[cH]:[c]:1Cl	0.186	1 out of 1



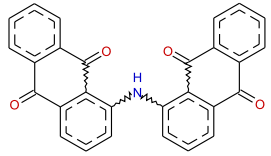
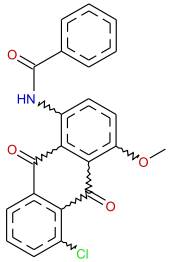
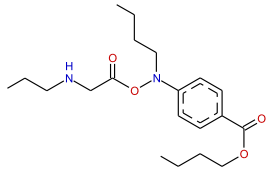
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.0584
Enrichment: 0.0848
Bayesian Score: -11.8
Mahalanobis Distance: 10.6
Mahalanobis Distance p-value: 0.0173

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Benzoic acid; p-(N-butyl-2-(butylamino)acetamido)-; butyl ester;
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.683	0.756	0.800
Reference	28ZPAK-;125;72	28ZPAK-;90;72	Arzneimittel-Forschung 8;609;58

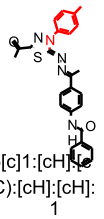
Model Applicability

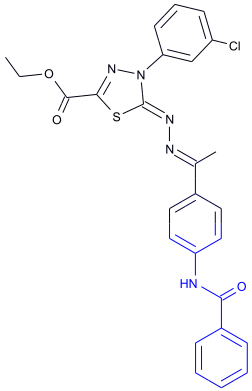
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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	633795852	 [*]N([*])[c]1:[cH]:[c H]:[c](C):[cH]:[cH]: 1	0.294	3 out of 3



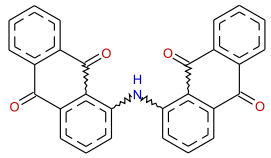
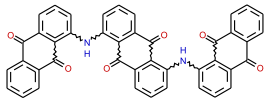
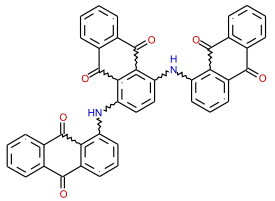
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.0346
Enrichment: 0.0503
Bayesian Score: -12.7
Mahalanobis Distance: 9.45
Mahalanobis Distance p-value: 0.28

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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-	Anthraquinone; 1;1'-(anthraquinon-1;5-ylenediimino)di-	Anthraquinone; 1;1'-(anthraquinon-1;4-ylenediimino)di-
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.824	0.854	0.855
Reference	28ZPAK-;125;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86

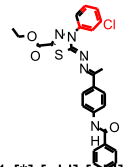
Model Applicability

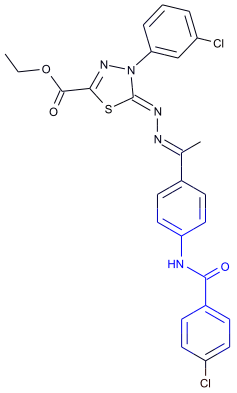
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-745491832	 [*][c]1:[*]:[cH]:[cH] :[c](Cl):[cH]:1	0.304	29 out of 32



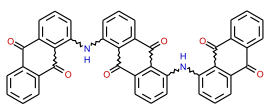
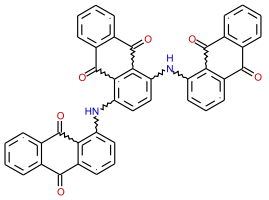
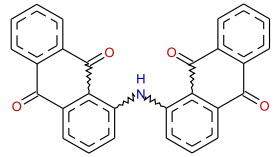
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.221
Enrichment: 0.321
Bayesian Score: -9.06
Mahalanobis Distance: 9.68
Mahalanobis Distance p-value: 0.189

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

Structural Similar Compounds

Name	Anthraquinone; 1;1'-(anthraquinon-1;5-ylenediimino)di-	Anthraquinone; 1;1'-(anthraquinon-1;4-ylenediimino)di-	ANTHRAQUINONE; 1;1'-IMINODI-
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.794	0.796	0.872
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	28ZPAK-;125;72

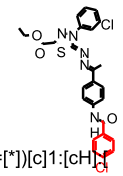
Model Applicability

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

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

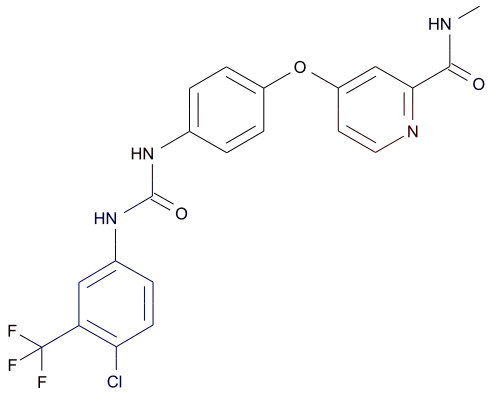
Feature Contribution

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

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

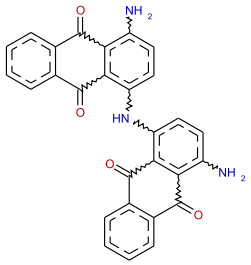
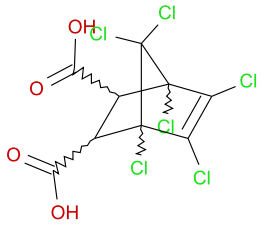
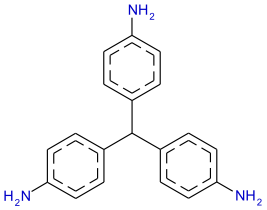
Bayesian Score: -1.8

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

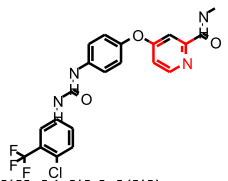
Model Applicability

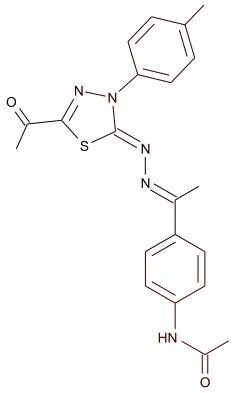
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



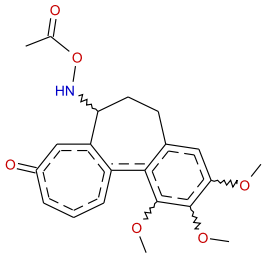
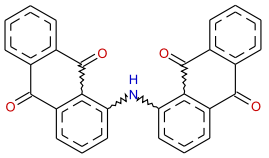
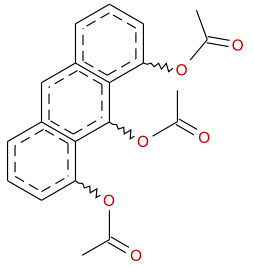
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 3.01
Mahalanobis Distance: 7.67
Mahalanobis Distance p-value: 0.972

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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-	1;8;9-ANTHRACENETRIOL; TRIACETATE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.611	0.725	0.736
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	BJOPAL 53;819;69

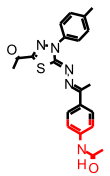
Model Applicability

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

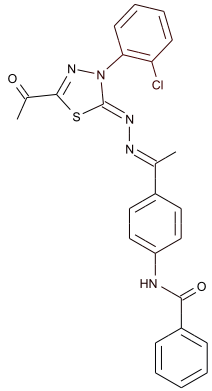
1. All properties and OPS components are within expected ranges.

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

TOPKAT_Ocular_Irritancy_None_vs_Irritant



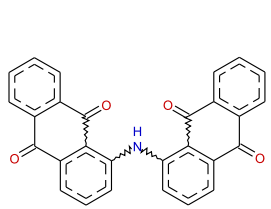
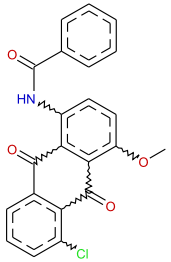
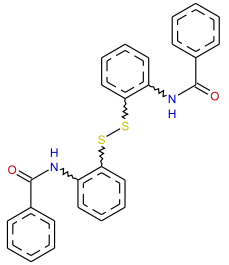
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.99
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	0.689	0.752	0.766
Reference	28ZPAK-;125;72	28ZPAK-;90;72	28ZPAK-;173;72

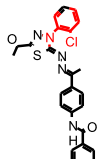
Model Applicability

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

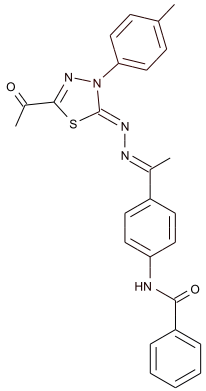
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-62776068	 [*]N([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	0.197	13 out of 13

TOPKAT_Ocular_Irritancy_None_vs_Irritant



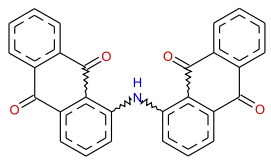
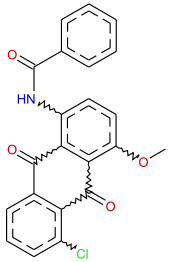
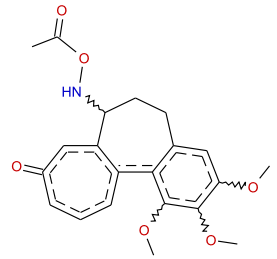
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.48
Mahalanobis Distance: 7.87
Mahalanobis Distance p-value: 0.948

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	COLCHICINE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.677	0.748	0.766
Reference	28ZPAK-;125;72	28ZPAK-;90;72	AJOPAA 31;837;48

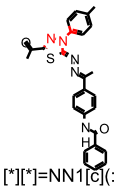
Model Applicability

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

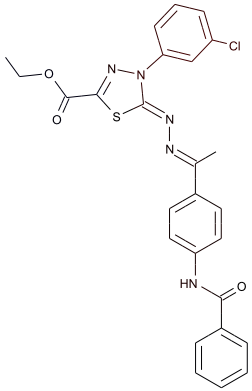
1. All properties and OPS components are within expected ranges.

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

TOPKAT_Ocular_Irritancy_None_vs_Irritant



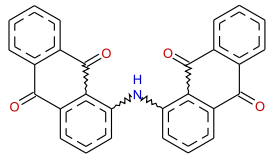
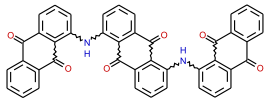
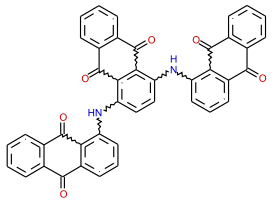
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.6
Mahalanobis Distance: 9.15
Mahalanobis Distance p-value: 0.431

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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-	Anthraquinone; 1;1'-(anthraquinon-1;5-ylenediimino)di-	Anthraquinone; 1;1'-(anthraquinon-1;4-ylenediimino)di-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.810	0.837	0.838
Reference	28ZPAK-;125;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86

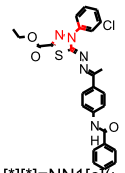
Model Applicability

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

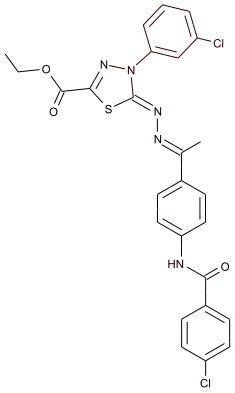
1. All properties and OPS components are within expected ranges.

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

TOPKAT_Ocular_Irritancy_None_vs_Irritant



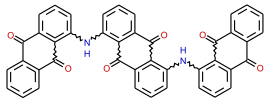
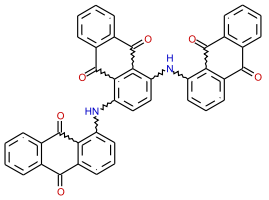
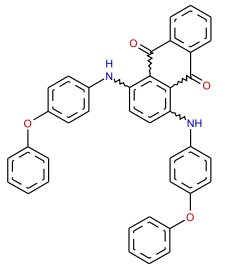
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 2.09
Mahalanobis Distance: 9.37
Mahalanobis Distance p-value: 0.317

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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'-(anthraquinon-1;5-ylenediimino)di-	Anthraquinone; 1;1'-(anthraquinon-1;4-ylenediimino)di-	ANILINE;N;N'-1;4-ANTHRAQUINONYLENEBIS(4-PHENOXY-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.785	0.786	0.849
Reference	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 736;86	28ZPAK-;114;72

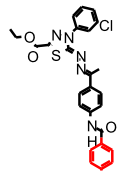
Model Applicability

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.

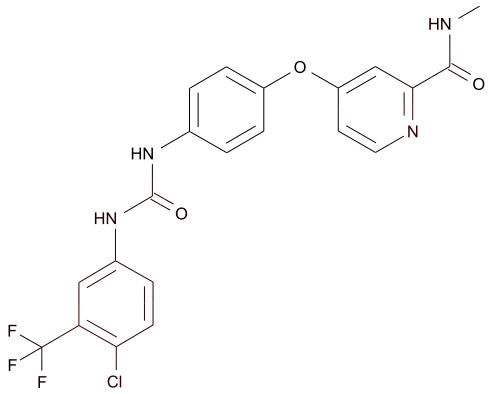
Feature Contribution

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

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

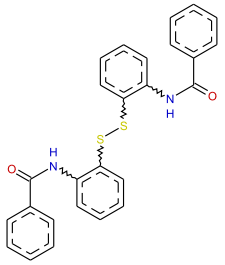
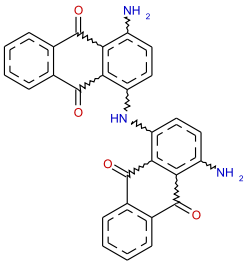
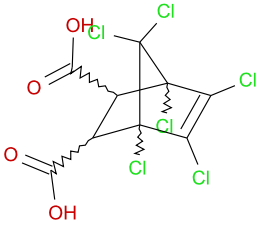
Bayesian Score: 3.04

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

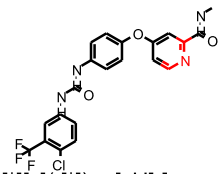
Model Applicability

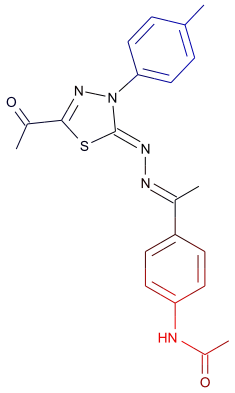
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



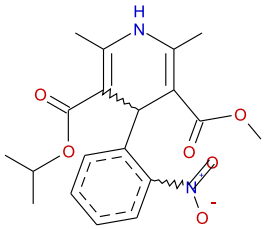
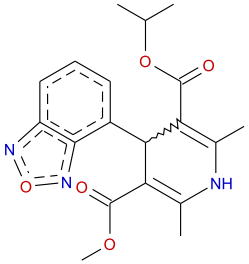
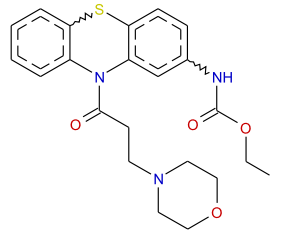
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.323
Enrichment: 1
Bayesian Score: 1.15
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 0.246

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	0.574	0.580	0.597
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.

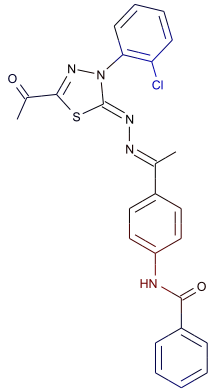
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{25}H_{20}ClN_5O_2S$

Molecular Weight: 489.97659

ALogP: 5.409

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.232

Enrichment: 0.721

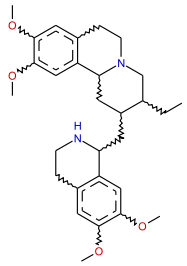
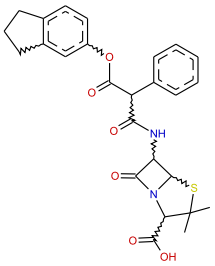
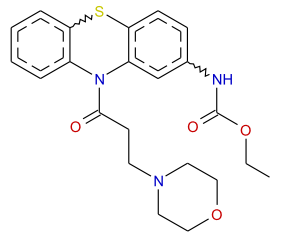
Bayesian Score: -4.08

Mahalanobis Distance: 11.6

Mahalanobis Distance p-value: 0.0148

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	0.717
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

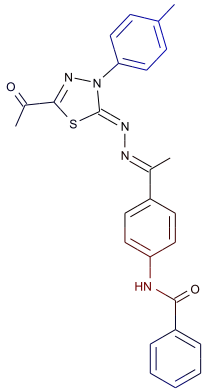
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



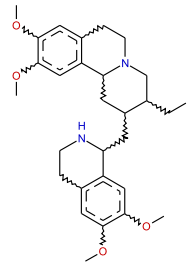
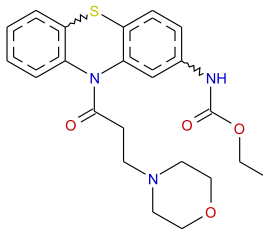
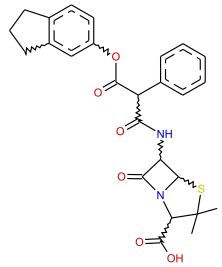
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.236
Enrichment: 0.732
Bayesian Score: -3.8
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.0374

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Moricizine	Carbenicillin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.696	0.697	0.703
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.

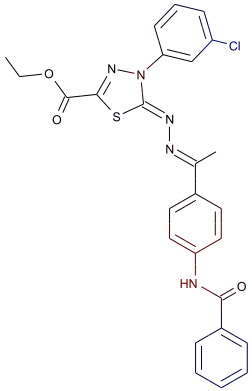
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{26}H_{22}ClN_5O_3S$

Molecular Weight: 520.00258

ALogP: 5.883

Rotatable Bonds: 8

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.266

Enrichment: 0.827

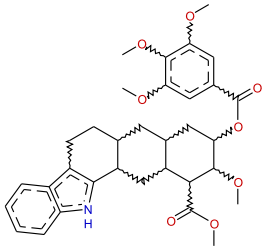
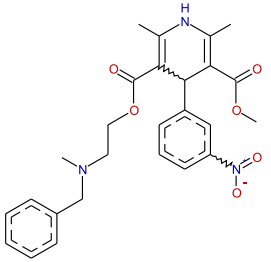
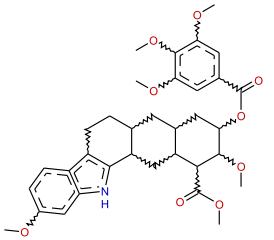
Bayesian Score: -1.76

Mahalanobis Distance: 18.4

Mahalanobis Distance p-value: 7.3e-017

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

Structural Similar Compounds

Name	Deserpidine	Nicardipine	Reserpine
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.618	0.689	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

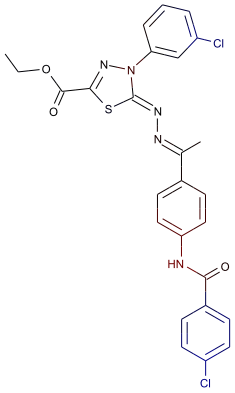
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. 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



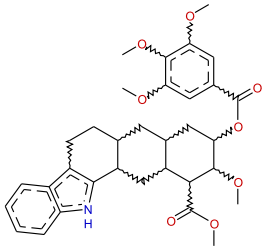
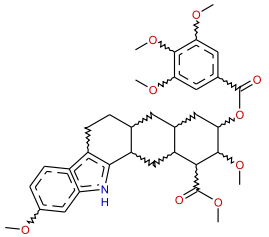
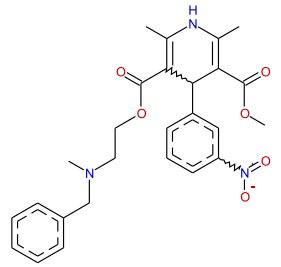
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.251
Enrichment: 0.779
Bayesian Score: -2.73
Mahalanobis Distance: 12.6
Mahalanobis Distance p-value: 0.000622

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Reserpine	Nicardipine
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.612	0.690	0.768
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.

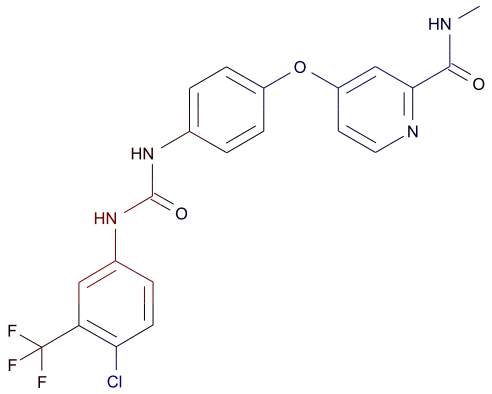
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]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

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

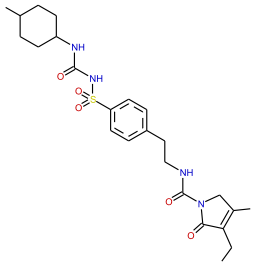
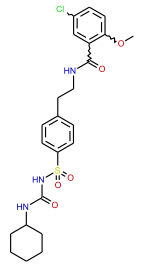
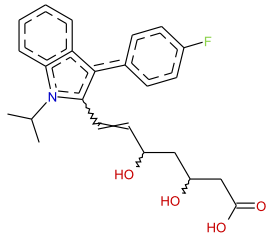
Mahalanobis Distance: 12.2

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

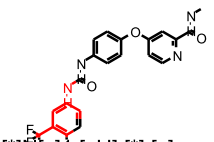
Model Applicability

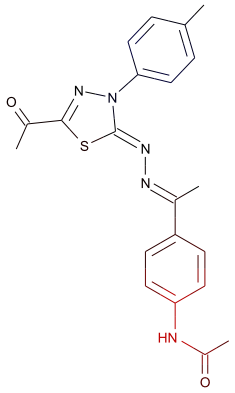
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



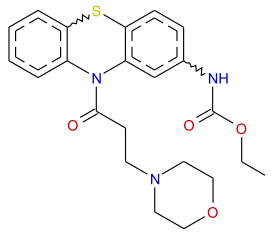
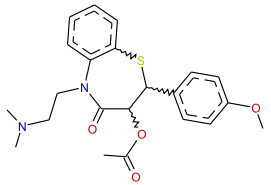
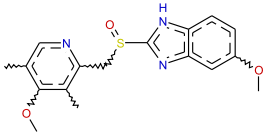
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.687
Enrichment: 1.84
Bayesian Score: 5.17
Mahalanobis Distance: 14.3
Mahalanobis Distance p-value: 1.38e-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	Moricizine	Diltiazem	Omeprazole
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.552	0.669	0.695
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

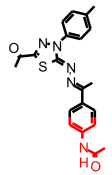
Model Applicability

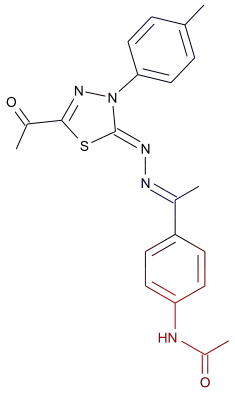
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: 3.0332. Training min, max, SD, explained variance: -2.8991, 3.0113, 1.313, 0.0255.

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



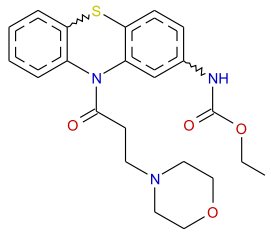
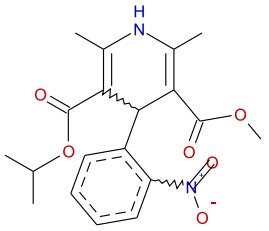
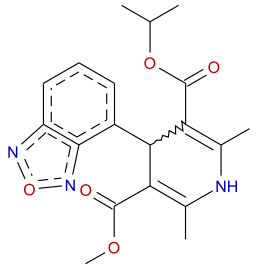
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.444
Enrichment: 1.33
Bayesian Score: 2.79
Mahalanobis Distance: 13.2
Mahalanobis Distance p-value: 0.000888

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Nisoldipine	Isradipine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.575	0.609	0.616
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

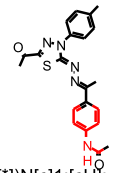
Model Applicability

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

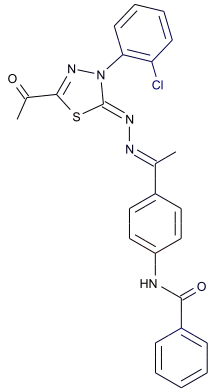
1. All properties and OPS components are within expected ranges.

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

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



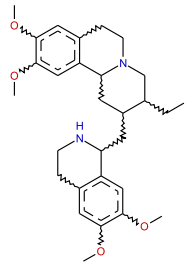
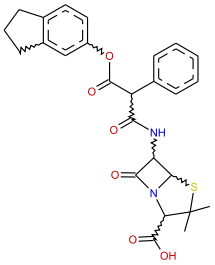
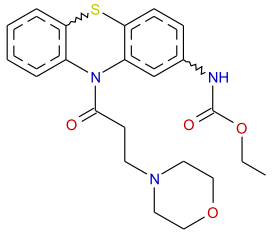
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.296
Enrichment: 0.886
Bayesian Score: -2.29
Mahalanobis Distance: 15.9
Mahalanobis Distance p-value: 7.31e-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	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

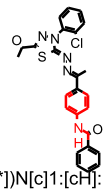
Model Applicability

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

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

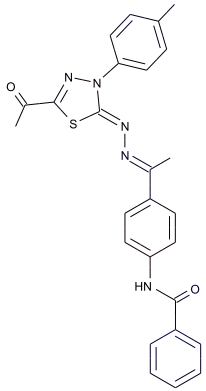
Feature Contribution

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

20b

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



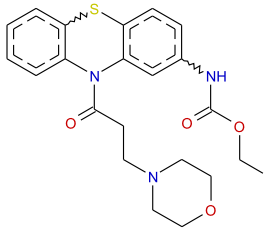
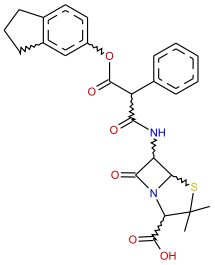
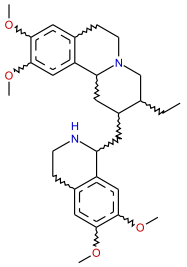
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.334
Enrichment: 0.998
Bayesian Score: -0.816
Mahalanobis Distance: 16.4
Mahalanobis Distance p-value: 7.7e-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	Moricizine	Carbenicillin	Emetine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.680	0.685	0.686
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

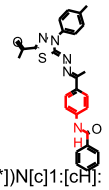
Model Applicability

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

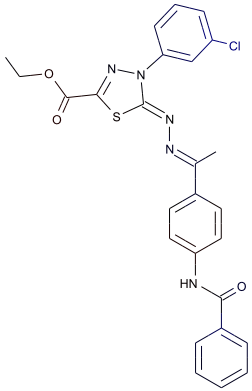
1. All properties and OPS components are within expected ranges.

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

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{26}H_{22}ClN_5O_3S$

Molecular Weight: 520.00258

ALogP: 5.883

Rotatable Bonds: 8

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.29

Enrichment: 0.868

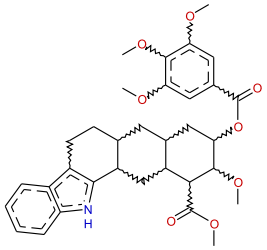
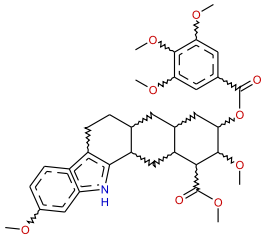
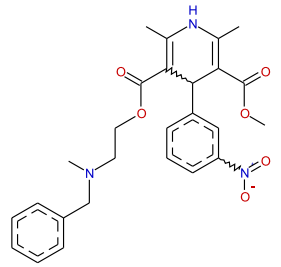
Bayesian Score: -2.53

Mahalanobis Distance: 16.9

Mahalanobis Distance p-value: 4.49e-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	Deserpidine	Reserpine	Nicardipine
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.610	0.700	0.733
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

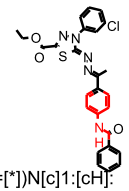
Model Applicability

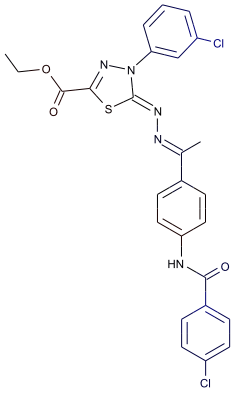
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC3 out of range. Value: 8.3204. Training min, max, SD, explained variance: -5.918, 7.9008, 2.435, 0.0553.

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



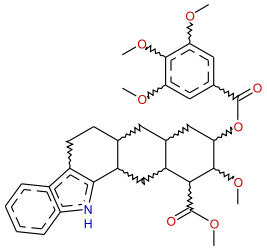
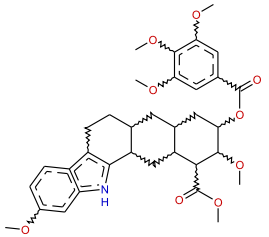
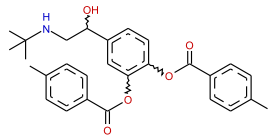
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.275
Enrichment: 0.822
Bayesian Score: -3.21
Mahalanobis Distance: 17.2
Mahalanobis Distance p-value: 9.39e-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	Deserpidine	Reserpine	Bitolterol
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.601	0.680	0.782
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

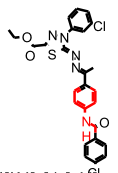
Model Applicability

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.

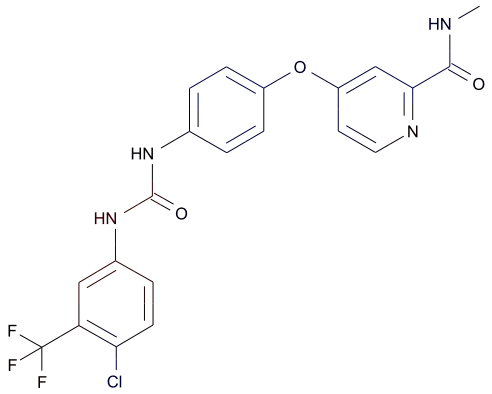
Feature Contribution

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

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

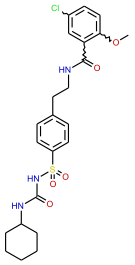
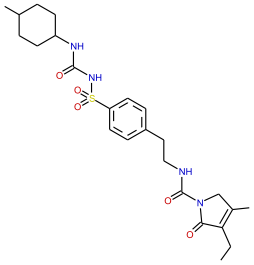
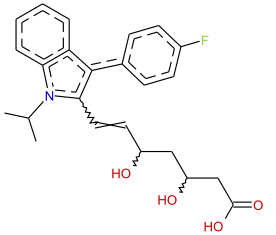
Bayesian Score: -2.4

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

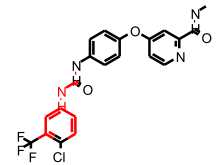
Model Applicability

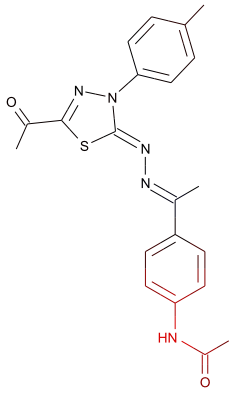
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



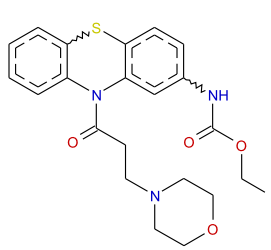
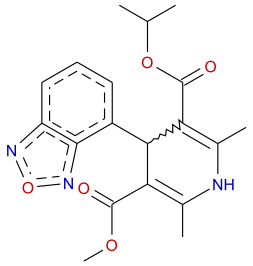
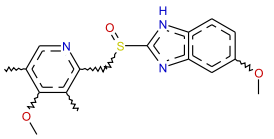
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.562
Enrichment: 1.36
Bayesian Score: 4.21
Mahalanobis Distance: 16.7
Mahalanobis Distance p-value: 3.13e-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	Isradipine	Omeprazole
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.594	0.636	0.706
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

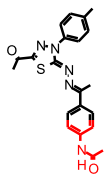
Model Applicability

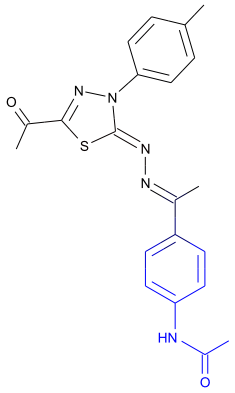
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

Top features for positive contribution

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



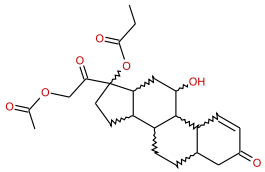
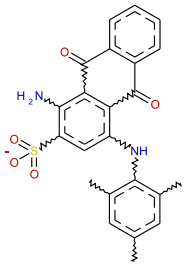
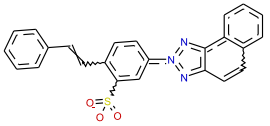
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 8.91e-005
Enrichment: 9.67e-005
Bayesian Score: -9
Mahalanobis Distance: 9.99
Mahalanobis Distance p-value: 0.0644

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Benzenesulfonic acid, 5-(2H-naphtho(1,2-d)triazol-2-yl)-2-(2-phenyl ethenyl)-, sodium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.654	0.771	0.782
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	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973

Model Applicability

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

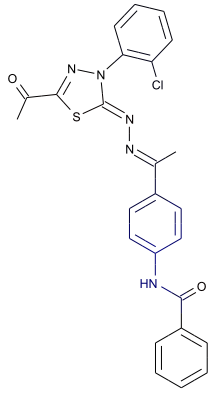
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{25}H_{20}ClN_5O_2S$

Molecular Weight: 489.97659

ALogP: 5.409

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.876

Enrichment: 0.952

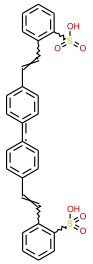
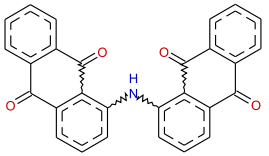
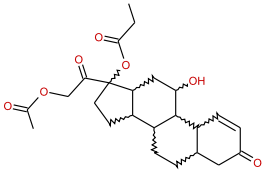
Bayesian Score: -2.99

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	0.795	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

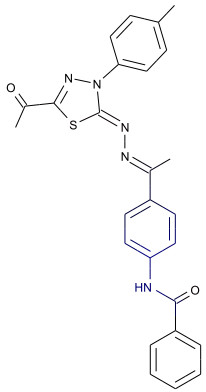
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



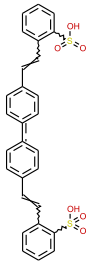
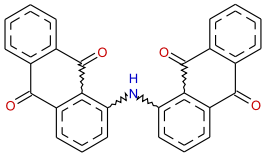
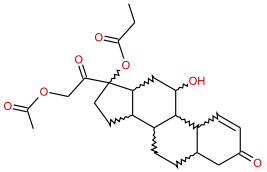
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 0.912
Enrichment: 0.99
Bayesian Score: -2.61
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 0.0276

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	0.737	0.782	0.789
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.

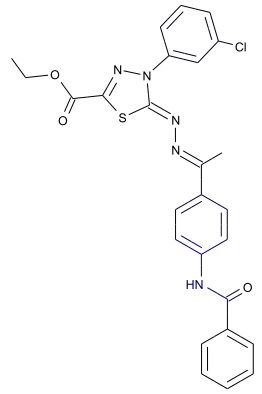
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{26}H_{22}ClN_5O_3S$

Molecular Weight: 520.00258

ALogP: 5.883

Rotatable Bonds: 8

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.879

Enrichment: 0.955

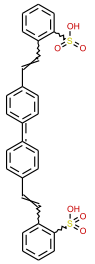
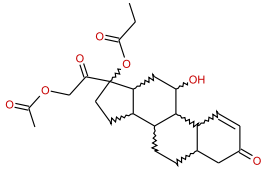
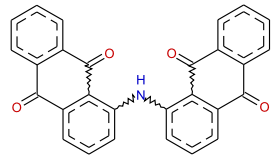
Bayesian Score: -2.97

Mahalanobis Distance: 9.79

Mahalanobis Distance p-value: 0.103

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	0.755	0.897	0.953
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.

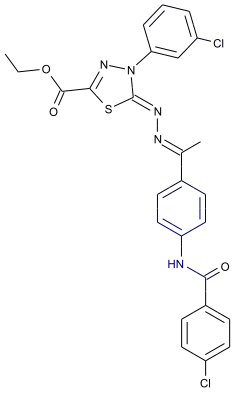
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{26}H_{21}Cl_2N_5O_3S$

Molecular Weight: 554.44763

ALogP: 6.547

Rotatable Bonds: 8

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.873

Enrichment: 0.948

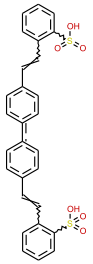
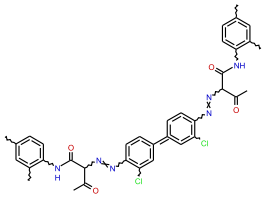
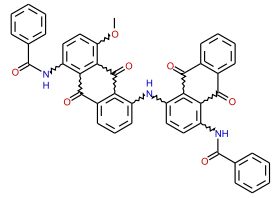
Bayesian Score: -3.02

Mahalanobis Distance: 10

Mahalanobis Distance p-value: 0.0588

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or 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	Butanamide, 2,2'-((3,3'-dichloro(1,1'-biphenyl)-4,4'-diyl)bis(azo)) bis(N-(2,4- dimethylphenyl)-3-oxo-	Anthraquinone, 3-methoxy-5,4'-iminobis(1-benzamido-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.774	0.915	0.979
Reference	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: OTS0555058	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,114,1

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.

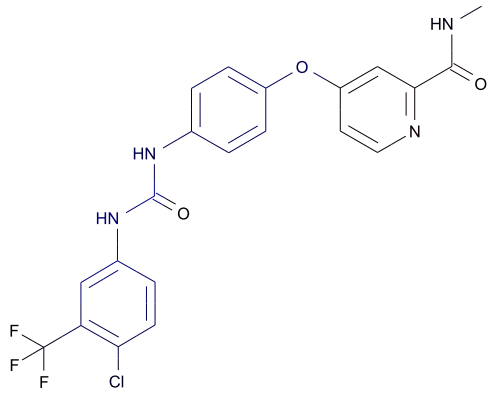
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

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

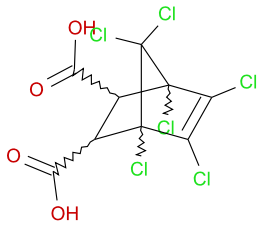
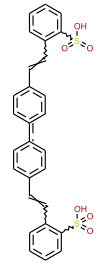
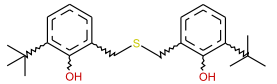
Bayesian Score: -5.23

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

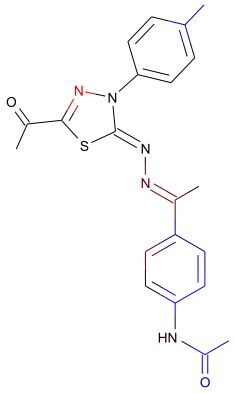
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



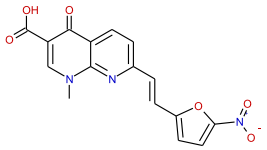
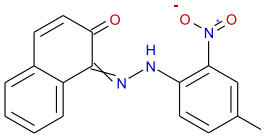
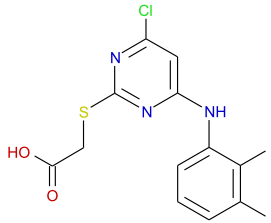
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 24.9
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.4
Mahalanobis Distance p-value: 0.000279

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	C.I. pigment red 3	[4-Chloro-6-(2,3-xylidino)-2-pyri-midinylthio]acetic acid s
Structure			
Actual Endpoint (-log C)	4.62839	0.937339	4.47685
Predicted Endpoint (-log C)	3.93264	3.17837	3.8529
Distance	0.663	0.676	0.676
Reference	CPDB	CPDB	CPDB

Model Applicability

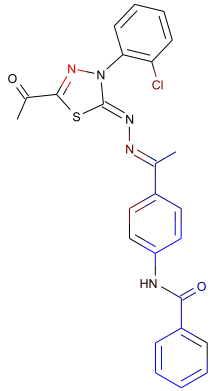
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



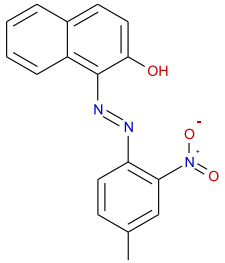
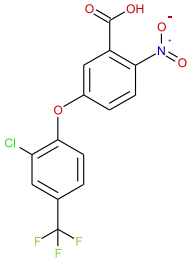
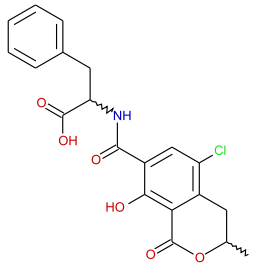
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 8.42
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 1.87e-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	Ochratoxin A
Structure			
Actual Endpoint (-log C)	0.937339	3.40908	4.79932
Predicted Endpoint (-log C)	3.26294	3.10974	3.6353
Distance	0.788	0.822	0.825
Reference	CPDB	CPDB	CPDB

Model Applicability

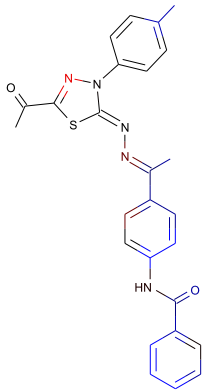
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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](:[*]):[*]
8. Unknown ECFP_2 feature: -1236953626: [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score



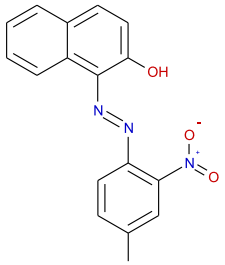
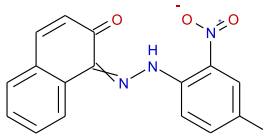
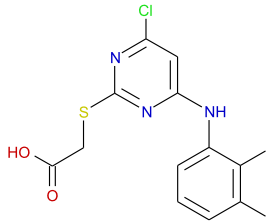
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 18.9
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 1.8e-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	C.I. pigment red 3	[4-Chloro-6-(2,3-xylidino)-2-pyri-midinylthio]acetic acid s
Structure			
Actual Endpoint (-log C)	0.937339	0.937339	4.47685
Predicted Endpoint (-log C)	3.26294	3.17837	3.8529
Distance	0.759	0.807	0.818
Reference	CPDB	CPDB	CPDB

Model Applicability

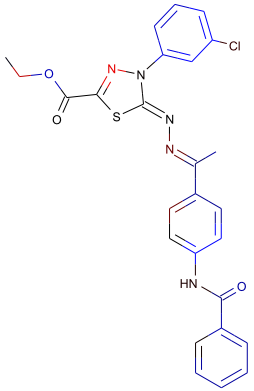
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 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



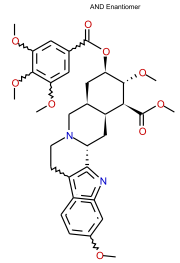
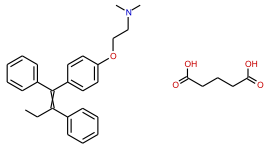
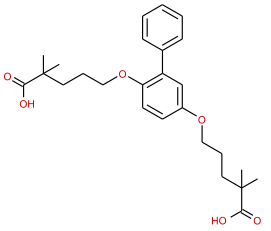
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 17.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.6
Mahalanobis Distance p-value: 7.31e-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	Tamoxifen citrate	5,5´-(1,1´-Biphenyl)-2,5-dyl-bis (oxy)(2,2-dimethylpentanoic acid)
Structure			
Actual Endpoint (-log C)	5.08368	5.05965	3.90166
Predicted Endpoint (-log C)	5.08273	4.24168	2.75893
Distance	0.786	0.867	0.873
Reference	CPDB	CPDB	CPDB

Model Applicability

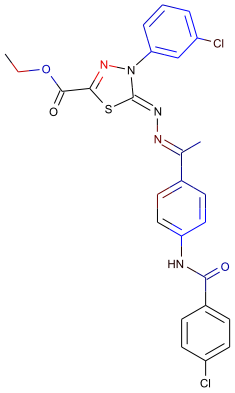
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



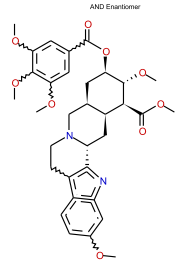
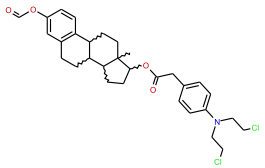
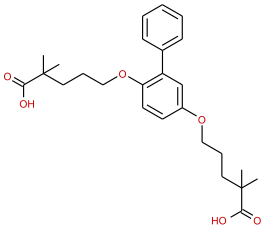
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 9.61
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15.6
Mahalanobis Distance p-value: 7.55e-015

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	223	Estradiol mustard	5,5´-(1,1´-Biphenyl)-2,5-dyl-bis (oxy)(2,2-dimethylpentanoic acid)
Structure			
Actual Endpoint (-log C)	5.08368	5.58568	3.90166
Predicted Endpoint (-log C)	5.08273	5.97715	2.75893
Distance	0.795	0.890	0.897
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.

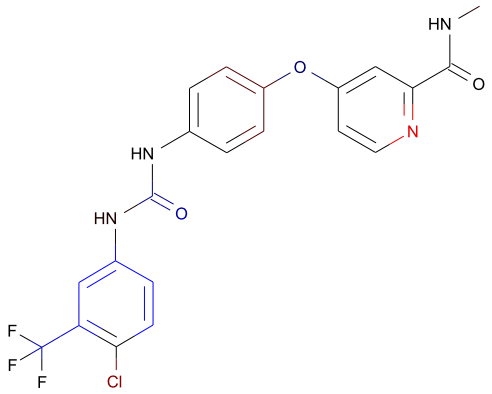
1. OPS PC20 out of range. Value: 3.8772. Training min, max, SD, explained variance: -4.3384, 3.4394, 1.14, 0.0162.
2. OPS PC24 out of range. Value: -4.6329. Training min, max, SD, explained variance: -4.4826, 3.8729, 1.034, 0.0133.
3. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
4. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
6. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
7. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
8. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

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

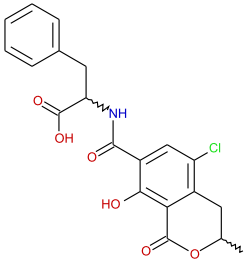
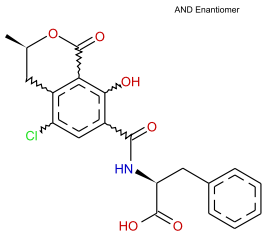
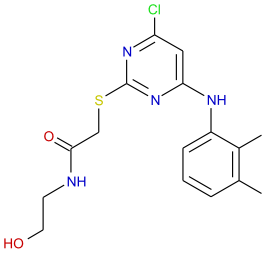
Mahalanobis Distance: 12.4

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

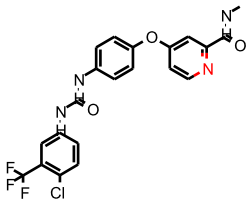
Model Applicability

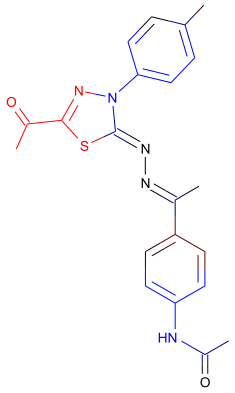
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



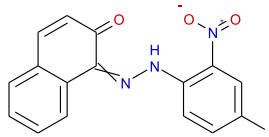
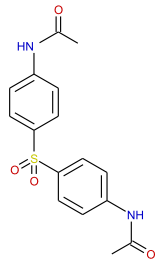
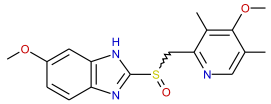
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 10.5
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.2
Mahalanobis Distance p-value: 1.49e-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	C.I. pigment red 3	4,4´-Sulfonylbisacetanilide	Omeprazole
Structure			
Actual Endpoint (-log C)	2.41938	3.77655	3.4628
Predicted Endpoint (-log C)	4.26375	3.55337	4.7324
Distance	0.634	0.650	0.656
Reference	CPDB	CPDB	CPDB

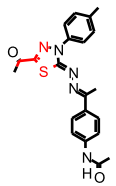
Model Applicability

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

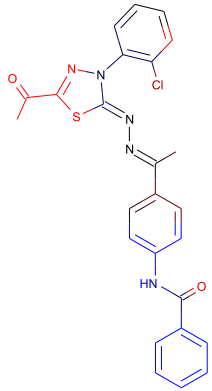
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{25}H_{20}ClN_5O_2S$

Molecular Weight: 489.97659

ALogP: 5.409

Rotatable Bonds: 6

Acceptors: 7

Donors: 1

Model Prediction

Prediction: 6.85

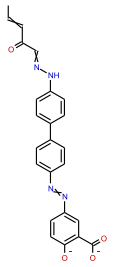
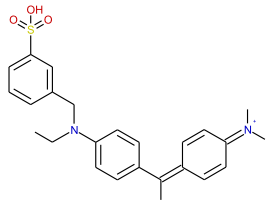
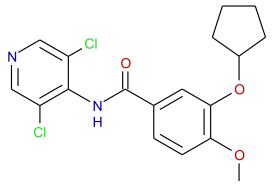
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 14.7

Mahalanobis Distance p-value: 3.34e-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	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	0.749
Reference	CPDB	CPDB	CPDB

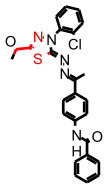
Model Applicability

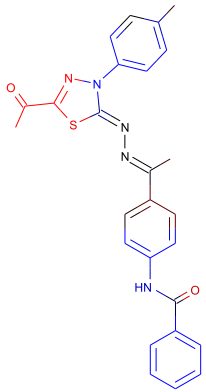
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



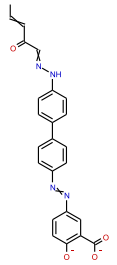
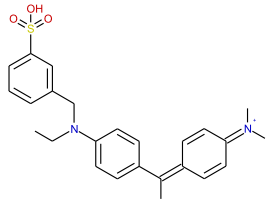
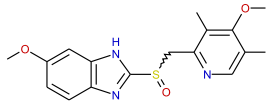
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 15.5
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.3
Mahalanobis Distance p-value: 2.75e-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	Omeprazole
Structure			
Actual Endpoint (-log C)	5.31387	2.8543	3.4628
Predicted Endpoint (-log C)	4.30266	3.40838	4.7324
Distance	0.646	0.713	0.751
Reference	CPDB	CPDB	CPDB

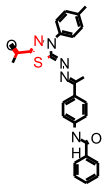
Model Applicability

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

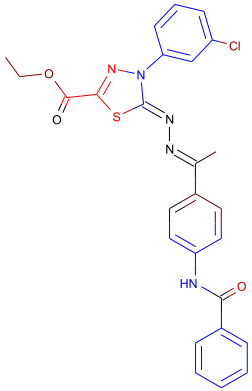
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357

TOPKAT_Carcinogenic_Potency_TD50_Rat



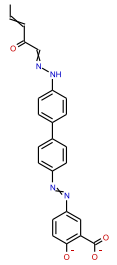
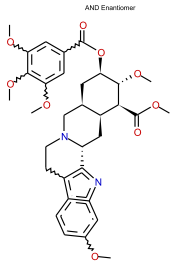
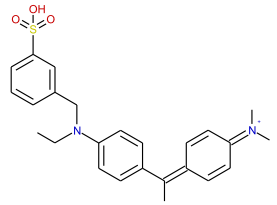
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 15.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.6
Mahalanobis Distance p-value: 2.14e-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	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	0.668	0.799	0.799
Reference	CPDB	CPDB	CPDB

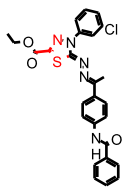
Model Applicability

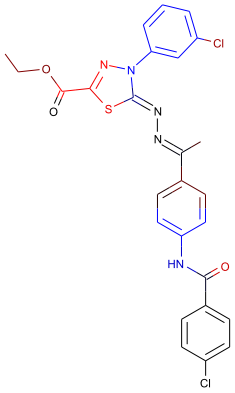
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



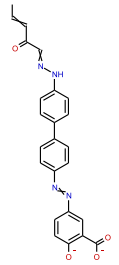
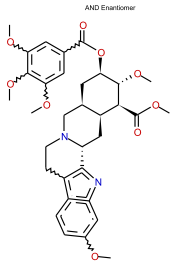
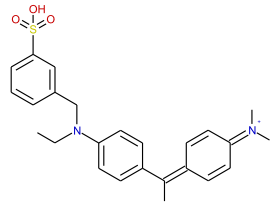
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 5.66
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.6
Mahalanobis Distance p-value: 2.54e-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	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	0.717	0.803	0.822
Reference	CPDB	CPDB	CPDB

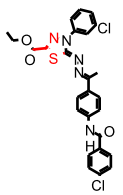
Model Applicability

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

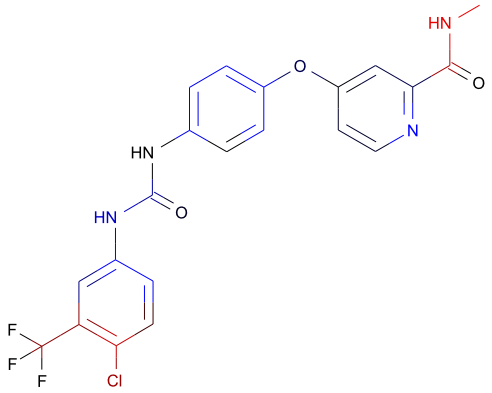
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357

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

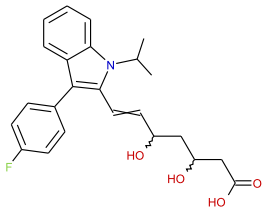
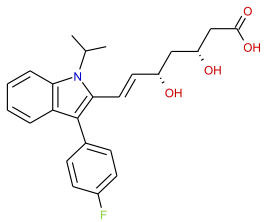
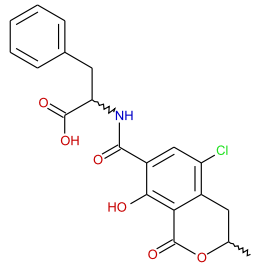
Mahalanobis Distance: 20.4

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

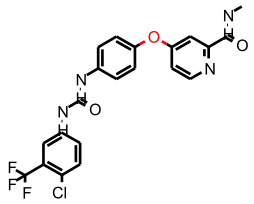
Model Applicability

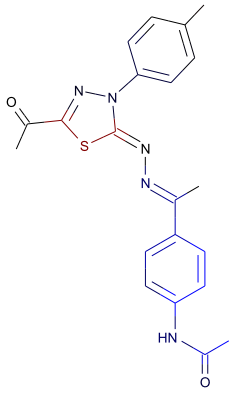
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



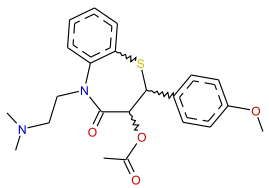
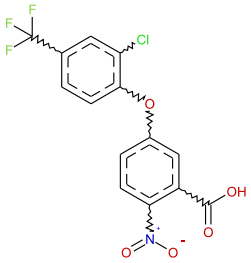
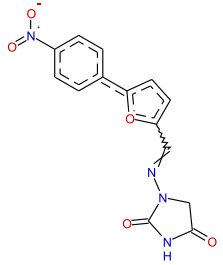
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0755
Unit: g/kg_body_weight
Mahalanobis Distance: 30.7
Mahalanobis Distance p-value: 7.29e-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	SODIUM ACIFLUORFEN	DANTROLENE.NA
Structure			
Actual Endpoint (-log C)	4.21961	4.16036	4.19625
Predicted Endpoint (-log C)	4.005	4.65915	4.62637
Distance	0.600	0.624	0.651
Reference	NDA-18602	EPA COVER SHEET 0192;891101;(1)	NDA-17443

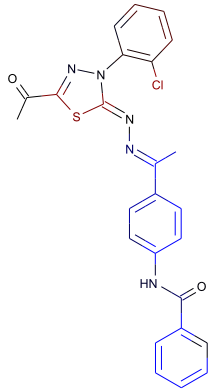
Model Applicability

Unknown features are fingerprint features in 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: -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: -179515162: [*]:[cH]:[c](C):[cH]:[*]

Feature Contribution

Top features for positive contribution



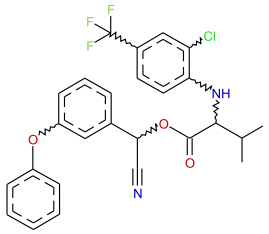
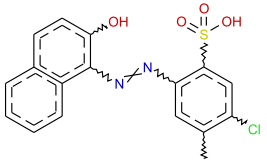
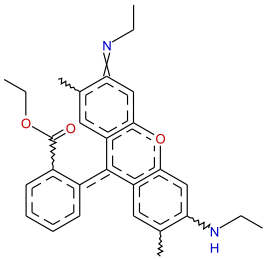
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0791
Unit: g/kg_body_weight
Mahalanobis Distance: 32.1
Mahalanobis Distance p-value: 1.54e-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.

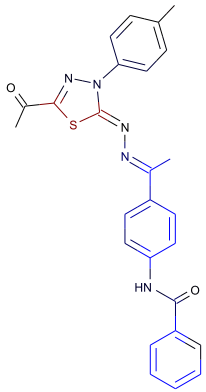
Structural Similar Compounds

Name	FLUVALINATE	D & C RED 9	RHODAMINE 6G
Structure			
Actual Endpoint (-log C)	5.30356	3.87715	4.54906
Predicted Endpoint (-log C)	4.89944	3.6546	4.6787
Distance	0.702	0.717	0.724
Reference	EPA COVER SHEET 0281;880630;(1)	NTP REPORT # 225	NTP 364 39

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: 4.6507. 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: -1236953626: [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl



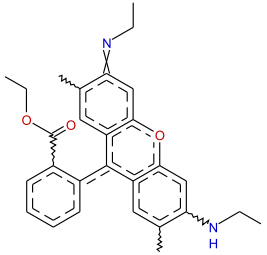
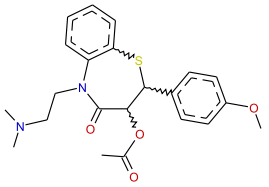
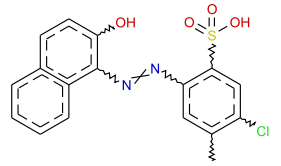
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0906
Unit: g/kg_body_weight
Mahalanobis Distance: 30.9
Mahalanobis Distance p-value: 3e-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

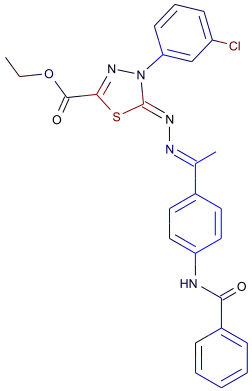
Name	RHODAMINE 6G	DILTIAZEM	D & C RED 9
Structure			
Actual Endpoint (-log C)	4.54906	4.21961	3.87715
Predicted Endpoint (-log C)	4.6787	4.005	3.6546
Distance	0.716	0.717	0.721
Reference	NTP 364 39	NDA-18602	NTP REPORT # 225

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: 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: -179515162: [*]:[cH]:[c](C):[cH]:[*]

Feature Contribution



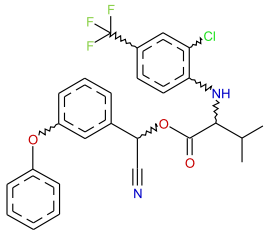
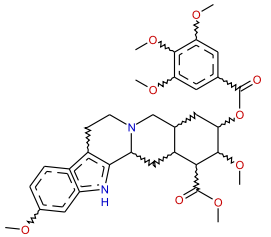
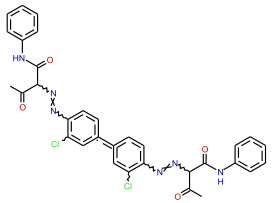
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.059
Unit: g/kg_body_weight
Mahalanobis Distance: 30.1
Mahalanobis Distance p-value: 6.28e-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.

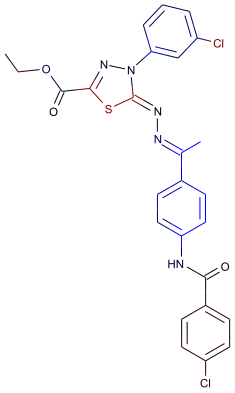
Structural Similar Compounds

Name	FLUVALINATE	RESERPINE	DIARYLANILIDE YELLOW
Structure			
Actual Endpoint (-log C)	5.30356	6.38645	2.70208
Predicted Endpoint (-log C)	4.89944	5.548	3.76154
Distance	0.694	0.705	0.726
Reference	EPA COVER SHEET 0281;880630;(1)	NTP 193 22	NTP 30 C-4

Model Applicability

Unknown features are fingerprint features in 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: 1430169877: [*]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: 1430791942: [*]OC(=O)C(=[*])[*]
16. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
17. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
18. Unknown ECFP_6 feature: -949601813: [*]OCC
19. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl



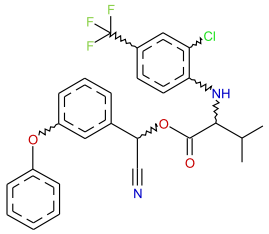
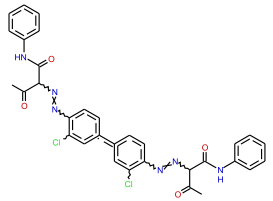
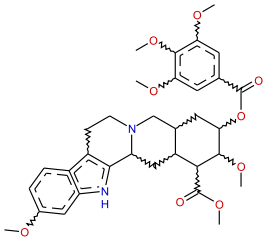
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0336
Unit: g/kg_body_weight
Mahalanobis Distance: 30.3
Mahalanobis Distance p-value: 2.68e-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	FLUVALINATE	DIARYLANILIDE YELLOW	RESERPINE
Structure			
Actual Endpoint (-log C)	5.30356	2.70208	6.38645
Predicted Endpoint (-log C)	4.89944	3.76154	5.548
Distance	0.692	0.694	0.710
Reference	EPA COVER SHEET 0281;880630;(1)	NTP 30 C-4	NTP 193 22

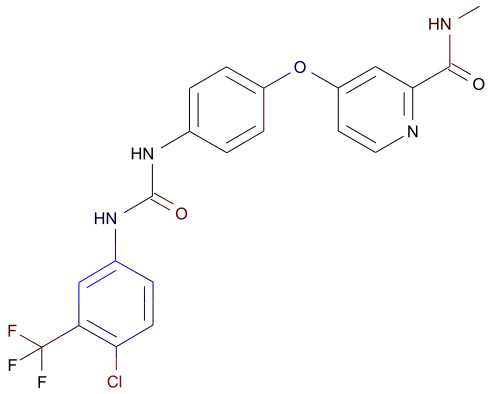
Model Applicability

Unknown features are fingerprint features in 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: 1430169877: [*]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: 1430791942: [*]OC(=O)C(=[*])[*]
16. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
17. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
18. Unknown ECFP_6 feature: -949601813: [*]OCC
19. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

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

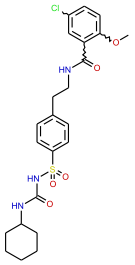
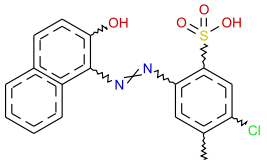
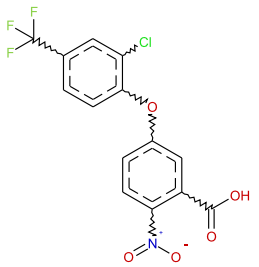
Unit: g/kg_body_weight

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)

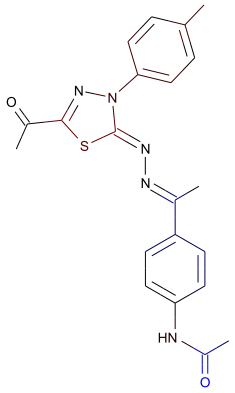
Model Applicability

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



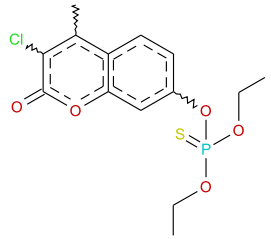
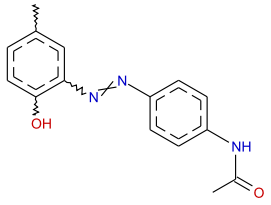
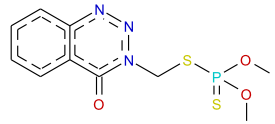
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0441
Unit: g/kg_body_weight
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 2.48e-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	COUMAPHOS	DISPERSE YELLOW 3	AZINPHOSMETHYL
Structure			
Actual Endpoint (-log C)	5.60537	2.77703	4.65515
Predicted Endpoint (-log C)	4.15004	2.80195	4.22281
Distance	0.608	0.649	0.670
Reference	NCI/NTP TR-96	NCI/NTP TR-222	NCI/NTP TR-69

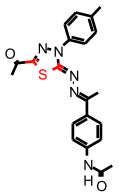
Model Applicability

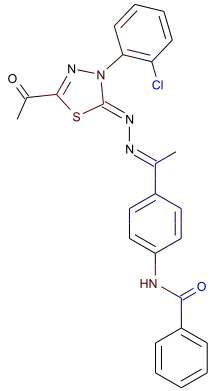
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



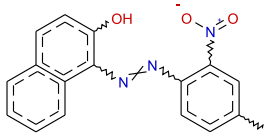
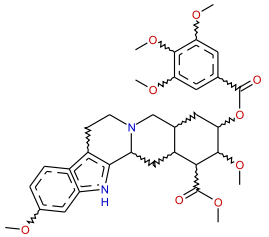
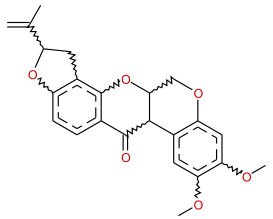
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

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

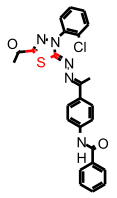
Model Applicability

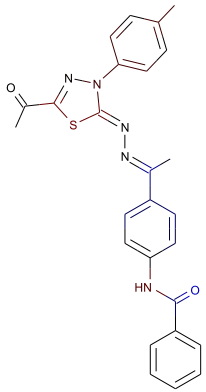
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



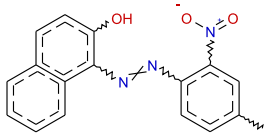
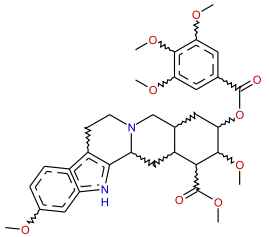
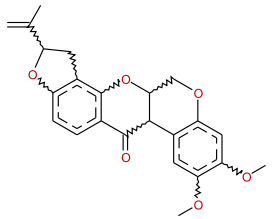
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.0423
Unit: g/kg_body_weight
Mahalanobis Distance: 9.28
Mahalanobis Distance p-value: 0.000577

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	0.714	0.788	0.789
Reference	NCI/NTP TR-407	NCI/NTP TR-193	NCI/NTP TR-320

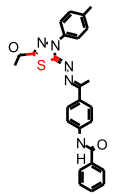
Model Applicability

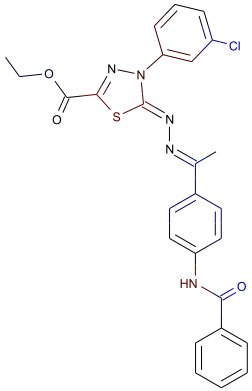
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



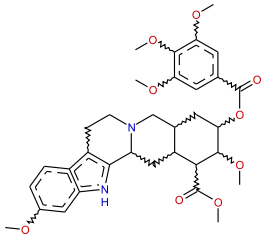
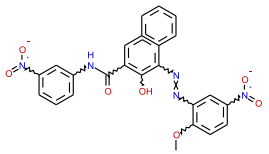
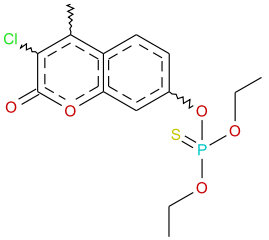
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0811
Unit: g/kg_body_weight
Mahalanobis Distance: 9.03
Mahalanobis Distance p-value: 0.0013

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	COUMAPHOS
Structure			
Actual Endpoint (-log C)	6.13118	2.30052	5.60537
Predicted Endpoint (-log C)	4.38304	3.55333	4.15004
Distance	0.670	0.810	0.868
Reference	NCI/NTP TR-193	NCI/NTP TR-411	NCI/NTP TR-96

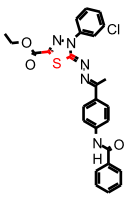
Model Applicability

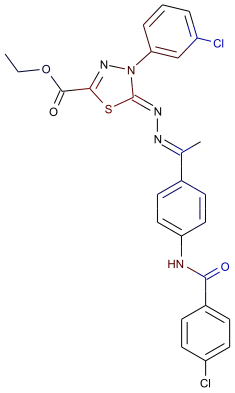
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 5.4937. 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



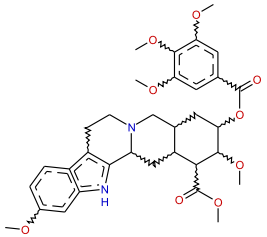
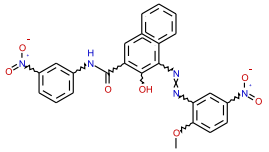
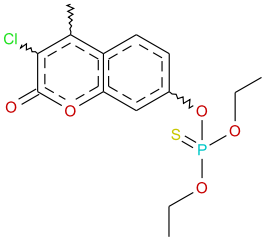
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0644
Unit: g/kg_body_weight
Mahalanobis Distance: 8.99
Mahalanobis Distance p-value: 0.00146

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	COUMAPHOS
Structure			
Actual Endpoint (-log C)	6.13118	2.30052	5.60537
Predicted Endpoint (-log C)	4.38304	3.55333	4.15004
Distance	0.682	0.827	0.915
Reference	NCI/NTP TR-193	NCI/NTP TR-411	NCI/NTP TR-96

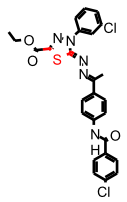
Model Applicability

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: 5.6115. Training min, max, SD, explained variance: -3.3892, 5.0834, 1.644, 0.0611.

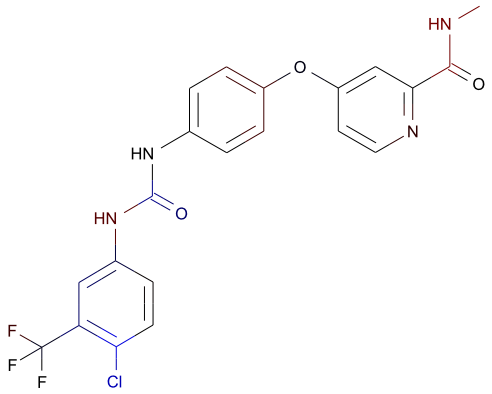
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095

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

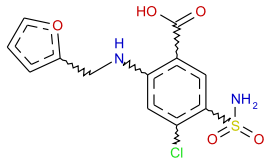
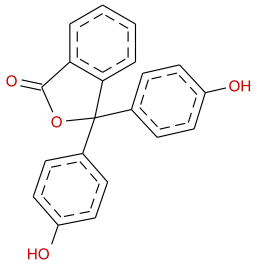
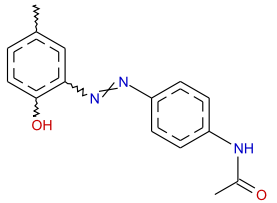
Unit: g/kg_body_weight

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

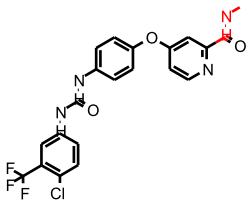
Model Applicability

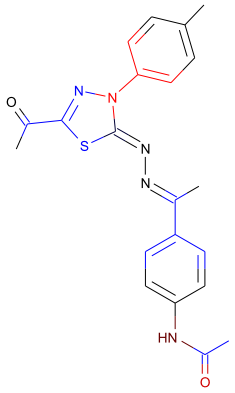
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



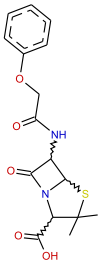
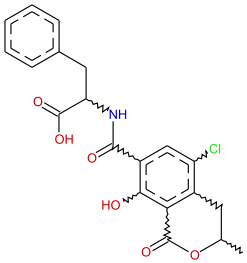
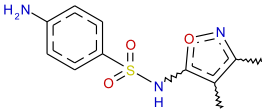
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.00165
Unit: g/kg_body_weight
Mahalanobis Distance: 9.37
Mahalanobis Distance p-value: 4.04e-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	PENICILLIN VK	OCHRATOXIN	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	2.54455	6.28396	2.82494
Predicted Endpoint (-log C)	3.9702	5.12358	3.0705
Distance	0.812	0.841	0.905
Reference	NCI/NTP TR-336	NCI/NTP TR-358	NCI/NTP TR-138

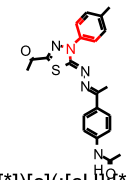
Model Applicability

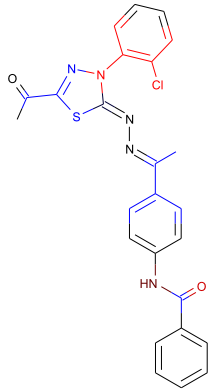
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Num_H_Acceptors out of range. Value: 7. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
2. OPS PC6 out of range. Value: -2.7424. 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



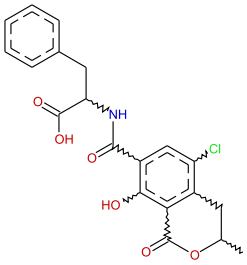
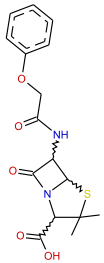
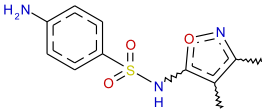
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

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

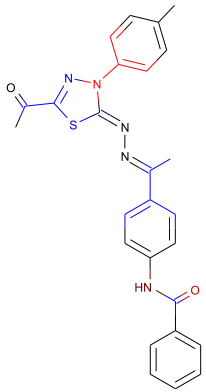
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



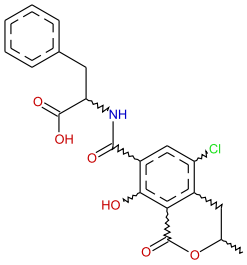
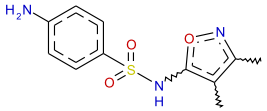
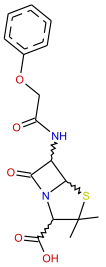
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.000884
Unit: g/kg_body_weight
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 1.4e-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	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	0.946	1.128	1.134
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

Model Applicability

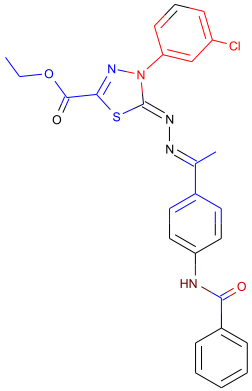
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 469.56. 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 PC6 out of range. Value: -3.2324. 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



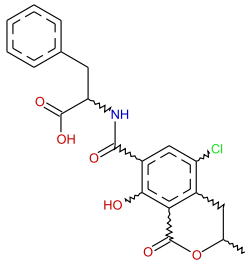
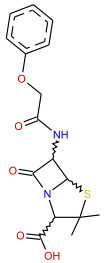
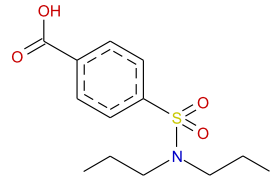
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 7.5e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 10
Mahalanobis Distance p-value: 5.06e-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	OCHRATOXIN	PENICILLIN VK	PROBENECID
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.85333
Predicted Endpoint (-log C)	5.12358	3.9702	2.4258
Distance	1.000	1.239	1.326
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-395

Model Applicability

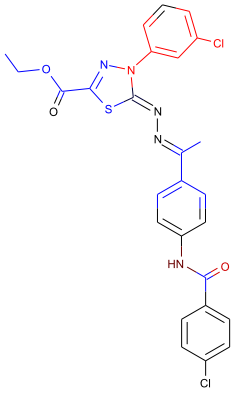
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. 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 PC3 out of range. Value: 5.2566. Training min, max, SD, explained variance: -4.6235, 5.1158, 1.773, 0.0972.
5. OPS PC5 out of range. Value: -3.8416. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
6. OPS PC10 out of range. Value: 2.749. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
7. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
8. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
9. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score



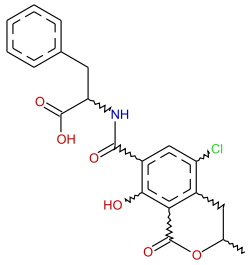
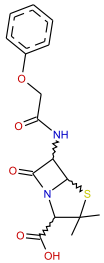
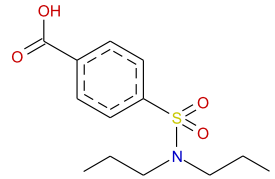
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 5.17e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 2.23e-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	OCHRATOXIN	PENICILLIN VK	PROBENECID
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.85333
Predicted Endpoint (-log C)	5.12358	3.9702	2.4258
Distance	1.059	1.313	1.396
Reference	NCI/NTP TR-358	NCI/NTP TR-336	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. Molecular_Weight out of range. Value: 554.45. 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 PC3 out of range. Value: 5.4618. Training min, max, SD, explained variance: -4.6235, 5.1158, 1.773, 0.0972.
5. OPS PC5 out of range. Value: -3.8598. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
6. OPS PC10 out of range. Value: 2.6215. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
7. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
8. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
9. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

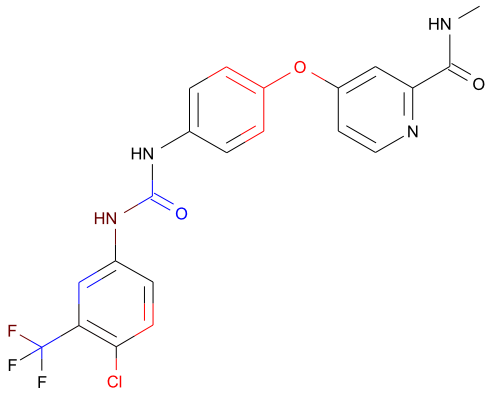
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

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

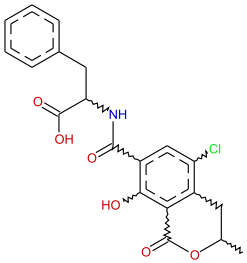
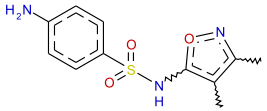
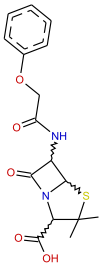
Unit: g/kg_body_weight

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

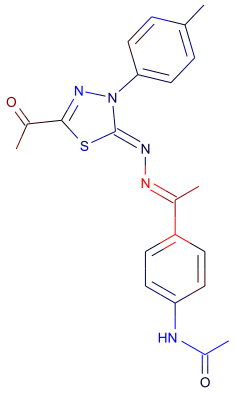
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



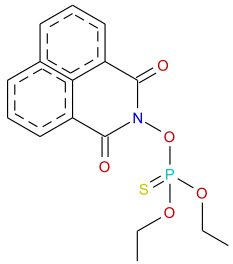
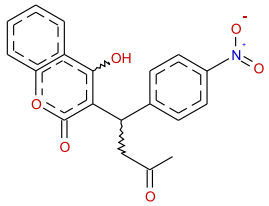
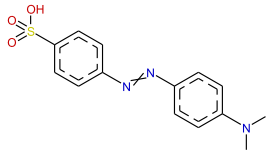
$C_{21}H_{21}N_5O_2S$
Molecular Weight: 407.48873
ALogP: 3.566
Rotatable Bonds: 5
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.663
Unit: g/kg_body_weight
Mahalanobis Distance: 23.8
Mahalanobis Distance p-value: 1.88e-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	NAPHTHALIMIDE; N-HYDROXY-; O;O-DIETHYL PHOSPHOROTHIOATE	ACENOCOUMARIN	METHYL ORANGE; SODIUM SALT (Na STRIPPED)
Structure			
Actual Endpoint (-log C)	2.864	2.838	3.707
Predicted Endpoint (-log C)	3.38692	3.3152	2.64236
Distance	0.625	0.627	0.636
Reference	TXAPA9 21;315;72	29ZVAB -;3;69	85JCAE -;1306;86

Model Applicability

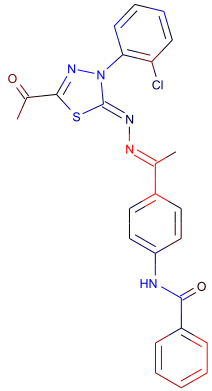
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



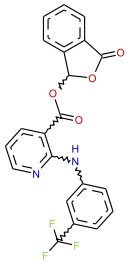
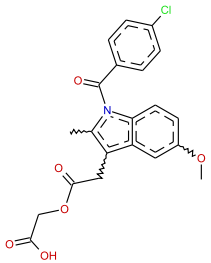
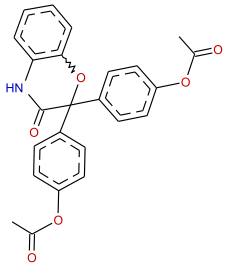
$C_{25}H_{20}ClN_5O_2S$
Molecular Weight: 489.97659
ALogP: 5.409
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.703
Unit: g/kg_body_weight
Mahalanobis Distance: 23.8
Mahalanobis Distance p-value: 1.14e-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	0.637	0.637	0.658
Reference	FRPSAX 36;372;81	ARZNAD 30;1398;80	NIIRDN 6;609;82

Model Applicability

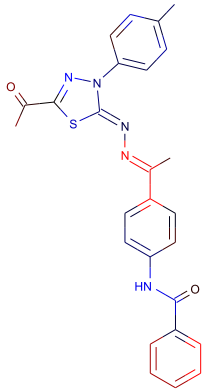
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



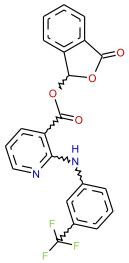
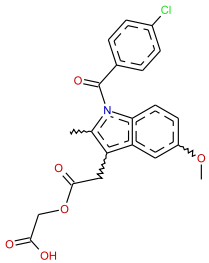
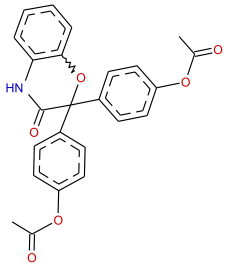
$C_{26}H_{23}N_5O_2S$
Molecular Weight: 469.55811
ALogP: 5.231
Rotatable Bonds: 6
Acceptors: 7
Donors: 1

Model Prediction

Prediction: 0.457
Unit: g/kg_body_weight
Mahalanobis Distance: 23.7
Mahalanobis Distance p-value: 5.01e-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	0.622	0.625	0.628
Reference	FRPSAX 36;372;81	ARZNAD 30;1398;80	NIIRDN 6;609;82

Model Applicability

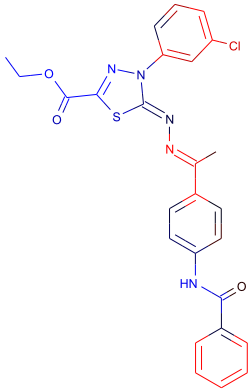
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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](:[*]):[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score



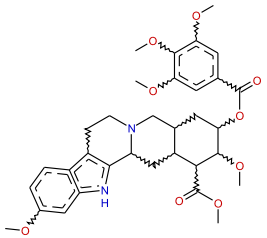
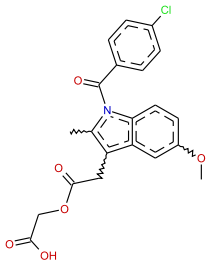
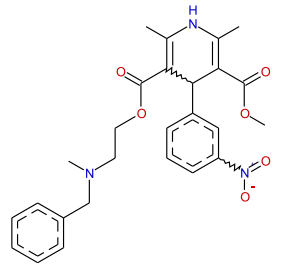
$C_{26}H_{22}ClN_5O_3S$
Molecular Weight: 520.00258
ALogP: 5.883
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.903
Unit: g/kg_body_weight
Mahalanobis Distance: 24.6
Mahalanobis Distance p-value: 8.56e-029

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	RESERPINE	ACEMETACIN	NICARDIPINE
Structure			
Actual Endpoint (-log C)	3.161	4.235	3.176
Predicted Endpoint (-log C)	2.72801	3.39415	3.41318
Distance	0.721	0.726	0.751
Reference	PSSCBG 11;555;80	ARZNAD 30;1398;80	ARZNAD 35;915;85

Model Applicability

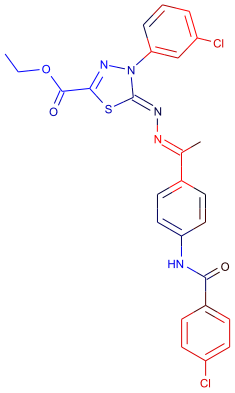
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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



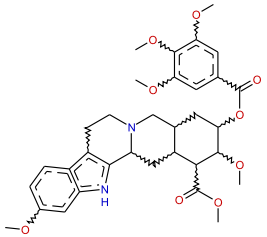
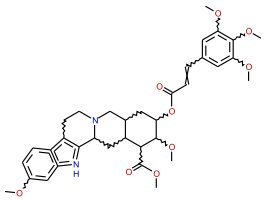
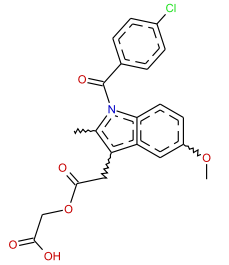
$C_{26}H_{21}Cl_2N_5O_3S$
Molecular Weight: 554.44763
ALogP: 6.547
Rotatable Bonds: 8
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.897
Unit: g/kg_body_weight
Mahalanobis Distance: 24.8
Mahalanobis Distance p-value: 7.55e-030

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Structural Similar Compounds

Name	RESERPINE	ANAPREL	ACEMETACIN
Structure			
Actual Endpoint (-log C)	3.161	2.803	4.235
Predicted Endpoint (-log C)	2.72801	2.99154	3.39415
Distance	0.737	0.748	0.790
Reference	PSSCBG 11;555;80	NIIRDN 6;898;82	ARZNAD 30;1398;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](:[*]):[*]
11. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

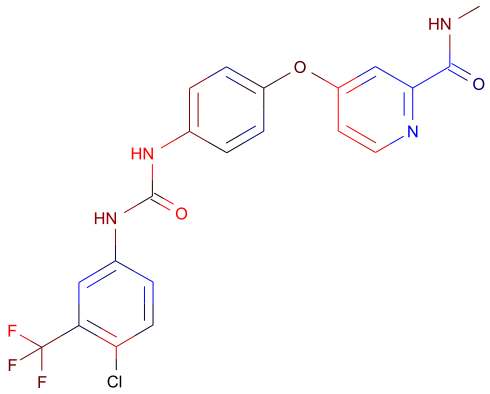
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

Sorafenib

TOPKAT_Rat_Oral_LD50



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

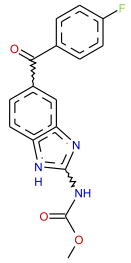
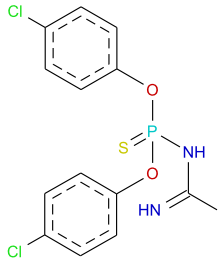
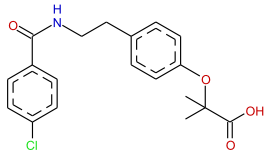
ALogP: 4.175

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