Discovery of new quinoline and isatine derivatives as potential VEGFR-2 inhibitors: Design, Synthesis, Antiproliferative, Docking, and MD simulation studies

Mohammed S. Taghour[a], Hazem Elkady[a]*, Wagdy M. Eldehna[b], Nehal El-Deeb[c,d], Ahmed M. Kenawy[e], Abeer E Abd El-Wahab [f,d], Eslam B. Elkaeed[g], Bshra A. Alsfouk[h], Ahmed M. Metwaly[i,c]* Ibrahim H. Eissa[a]*

[a] Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo11884, Egypt.
[b] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Kafrelsheikh University, P.O. Box 33516, Kafrelsheikh, Egypt.
[c] Biopharmaceutical Products Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications (SRTA-City), Alexandria 21934, Egypt.
[d] Pharmaceutical and Fermentation Industries Development Center, City of Scientific Research and Technological Applications (SRTA city), Alexandria, New Borg El-Arab City, 21934, Egypt
[e] Nucleic Acids Research Department, Genetic Engineering and Biotechnology Research Institute. City of Scientific Research and Technological Applications (SRTA-City). Alexandria 21934, Egypt
[f] Medical Biotechnology Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications (SRTA-CITY), New Borg El-Arab City, Egypt.
[g]Department of Pharmaceutical Sciences, College of Pharmacy, AlMaarefa University, Riyadh 13713, Saudi Arabia.
[h] Department of Pharmaceutical Sciences, College of Pharmacy, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia.
[i] 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

Hazem Elkady:

Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo 11884, Egypt. **Email:** Hazemelkady@azhar.edu.eg

Content

1- Biological testing

a- Mammalian cell lines culture

CaCo-2, and A549 cell lines were cultured on DMEM media, meanwhile MDA-MB-231 and hepG-2 cell line were cultured on RBMI media. The cultured media were supplemented with 200 mM L-glutamine, 10.0% fetal bovine serum (Lonza), and 1.0% penicillin/streptomycin. Cells were seeded into 25.0 cm tissue culture flasks and incubated at 37°C in a 5.0% CO_2 incubator for 24 h or till confluency.

b- Safety assay

The safety profiles of the tested compounds were checked on one non-cancerous cell line (Vero and 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.

c- *In-vitro* anticancer activity

Anticancer activities of the tested compounds against CaCo-2, A549, MDA-MB-231, and hepG-2 cell lines were quantified using MTS assay kit (Promega) as described by the Manufacturer.

d- Selectivity index (SI)

The selectivity index values of the tested compounds on cancer cells were calculated as described by Koch et al with slight modifications; $SI=IC_{50}nc/IC_{50}cc$, where $IC_{50}nc$: the IC_{50} value of the tested compound on normal cells and $IC_{50}cc$: IC_{50} of the tested compound on cancer cell line**.**

e- Wound healing assay (Migration assay)

CaCo-2 cells were grown to 95.0% confluency in a complete DMEM medium and then the wounds were formed using a plastic tip. After washing with pre-warmed PBS, the cells were incubated in

the specific medium or the **9** treatment. After incubation at 37°C and 5.0% CO_2 for 24h, the cells were washed with PBS and the wounds distance was determined as the scratch width of the treated and untreated groups using ImageJ software.

f- Gene expression pattern alternation of cancer cell after 9 treatment

The molecular anticancer mode of action of **9** was investigated by screening their ability to affect the gene expression levels of Bcl2, Bcl-xl, TGF and Survivin genes using specific forward and reverse primers and RTq-PCR technique (Table 1) in CaCo-2 cells (chosen as the most sensitive cancer cell line). After cellular treatment, CaCo-2 cell line was cultured into 12 well plates (6×10^3 cell/ml) for 24 h. with the sub-IC50 concentration of **9**. After treatment, total RNA extraction was performed using RNA extraction kit (Qiagen, Germany). Then, 1 ug of the obtained RNA was used to synthesiz cDNA using cDNA synthesis kit (Promega Corp., Madison, WI) as recommended by the manufacturer. Simultaneously, GAPDH forward and reverse primers (Table 1) were used to amplify the house keeping gene as internal control for standardization of PCR products. 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.

g- Flow cytometry analysis for cell cycle

To determine the role of the synthesized compounds in cell cycle distribution, cell cycle analysis was performed using propidium iodide (PI) staining and flow cytometry analysis for compound **9**. Flow Cytometry Kit for Cell Cycle Analysis (ab139418_Propidium Iodide Flow Cytometry Kit/BD) was used in this test. HepG2 cells were treated with compound **9** (13.90 µM) for 24 h. Then, the cells were fixed in 70% ethanol at 4 °C for 12 h. After that, the cells were washed with cold PBS, incubated with 100 µl RNase A at 37 °C for 30 min, and stained with 400 µl PI in the dark at room temperature for further 30 min. The stained cells were measured using Epics XL-MCL™ Flow Cytometer (Beckman Coulter), and the data were analyzed using Flowing software (version 2.5.1, Turku Centre for Biotechnology, Turku, Finland).

-Sequence of the primers

Primer ID	Sequence
Bcl-F	5′-TATAAGCTGTCGCAGAGGGGCTA-3′
Bcl-R	5′-GTACTCAGTCATCCACAGGGCGAT-3′
Bcl-Xlf	5’CAGAGCTTTGAACAGGTAG-3′
Bcl-XlR	5’GCTCTCGGGTGCTGTATTG-3′
Surv-F	5′-TGCCCCGACGTTGCC-3′
Surv-R	5′-CAGTTCTTGAATGTAGAGATGCGGT-3′
TGF-F	5′CAAGGGCTACCATGCCAACT3′
TGF-R	5′AGGGCCAGGACCTTGCTG3′
β-actin-F	5′-GTGGGGCGCCCCAGGCACCA-3′
β-actin-R	5′-CTCCTTAATGTCACGCACGATTTC-3′

2- *In silico* studies

a- Docking studies

The docking studies were performed utilizing MOE 2019 software to explore the binding mode of the synthesized compounds towards VEGFR-2. The 3D crystal structure of the target macromolecules VEGFR-2 were downloaded from the protein databank, http://www.pdb.org (PDB ID; 2OH4). Sorafenib was used as reference ligand. To prepare the target protein, water molecules were removed, and the valances of atoms were corrected through protonation of the whole molecule. Then energy minimization was carried out by applying CHARMM and MMFF94 force fields. After that, the active binding site was defined and prepared for docking. The validation process was performed by redocking the co-crystallized ligand. The designed compounds together with sorafenib were drawn using ChemBioDraw Ultra 14.0 and saved as MDL-SD format. The sketched compounds were constructed from fragment libraries in MOE program, protonated,

followed by energy minimization then prepared for docking. Docking process was carried through Triangle matcher placement inserted in compute window, and the scoring function was London dG. Ten conformers (poses) for each molecule were generated using genetic algorithm searches. The free energies and binding modes of the designed molecules against VEGFR-2 were determined. The most ideal pose was selected according to its binding free energy as well as its binding mode with target molecule.

b- ADMET studies

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

c- Toxicity studies

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

d- **Molecular dynamics simulation**

Molecular dynamics simulation of the protein-ligand complexes was performed using GROMACS 2021 and Linux 5.4 package. The GROMOS96 54a7 forcefield was selected as the force field for proteins and the ligand topologies were generated from the PRODRG server. All the complexes were solvated using simple point charge (SPC) water molecules in a rectangular box. To make the simulation system electrically neutral, required number of Na^+ and Cl^- ions were added while 0.15 mol/L salt concentrations were set in all the systems. Using the steepest descent method, all the solvated systems were subjected to energy minimization for 5000 steps. Afterwards, NVT (constant number of particles, volume, and temperature) series, NPT (constant number of particles, pressure, and temperature) series, and the production run were conducted in the MD simulation. The NVT and the NPT series were conducted at a 300 K temperature and 1 atm pressure for the duration of 300 ps. V-rescale thermostat and Parrinello-Rahman barostat were

selected of the performed simulation. Finally, the production run was performed at 300 K for a duration of 100 ns (nanoseconds). Thereafter, a comparative analysis was performed measuring root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA) and hydrogen bonds to analyze their stability. The Xmgrace program was used to represent the analyses in the form of plots.

e- **MM/PBSA**

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

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

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

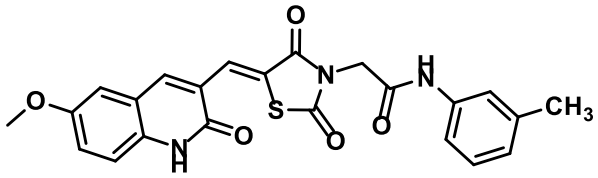
3- Chemistry and material

Melting points were measured with a Stuart melting point apparatus and were uncorrected. Infrared (IR) spectra were recorded as KBr disks using an FT-IR Spectrum BX apparatus (Perkin Elmer, CT, USA). NMR spectra were recorded on a Bruker NMR spectrometer (Bruker, Reinstetten, Germany).^{1}H spectra were run at 500 MHz and ^{13}C spectra were run at 125 MHz in deuterated dimethyl sulfoxide (DMSO-d_6). Chemical shifts are expressed in values (ppm) using the solvent peak as an internal standard. All coupling constant (J) values are given in Hz. The abbreviations used are as follows: s, singlet; d, doublet; m, multiplet. Analytical thin layer chromatography (TLC) on silica gel plates containing a UV indicator was employed routinely to follow the course of reactions and to check the purity of the products. All reagents and solvents were purified and dried by standard techniques. The microspheres were prepared with poly (D, L-lactide co-glycolide) PLGA (50:50, mol. wt 30,000–60,000), which was purchased from Sigma-

Aldrich (St. Louis, USA). The emulsifier, low molecular weight polyvinyl alcohol (PVA) was obtained from Alfa Aesar (Karlsruhe, Germany). Dichloromethane (DCM) was purchased from Avonchem (United Kingdom). Dimethyl sulfoxide (DMSO) was obtained from Loba Chemie (Mumbai, India). All ingredients used were of analytical grade. All cell lines have been purchased from the American Type Culture Collection (ATCC).

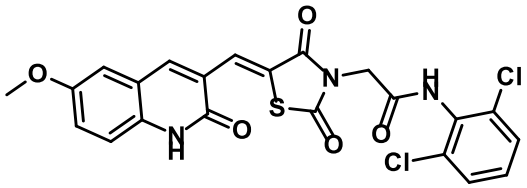
4- Characterization of the target compounds 8,9, 12, 13, and 14

4.1.(Z)-2-(5-((6-Methoxy-2-oxo-1,2-dihydroquinolin-3-yl)methylene)-2,4-dioxothiazolidin-3-yl)-N-(m-tolyl)acetamide 8



White powder **(**yield, 71%); m. p. = 255-257ºC; IR (KBr, cm^{-1}): 3270, 3225 (NH), 3084 (aromatic CH) 2992, 2910 (aliphatic CH), 1671 (C=O); ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.40 (s, 1H), 10.25 (s, 1H), 8.45 (s, 1H), 8.20 (s, 1H), 7.98 (t, *J* = 15.8 Hz, 2H), 7.40 (d, *J* = 2.5 Hz, 2H), 7.32 (m, 3H), 4.51 (s, 2H), 3.82 (s, 3H), 2.29 (s, 3H) ; ^{13}C NMR (101 MHz, DMSO-d_6) δ 190.39, 171.60, 161.56, 157.02, 155.00, 142.27, 138.62, 136.50, 132.64, 129.00, 128.83, 127.68, 126.20, 124.34, 124.18, 119.18, 117.30, 111.58, 56.04, 46.47, 21.30 for $C_{23}H_{19}N_3O_5S$ (449.48).

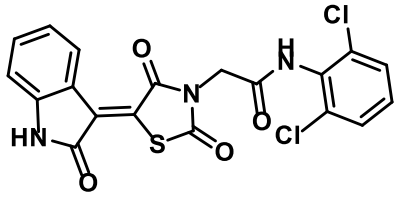
4.2. (Z)-N-(2,6-dichlorophenyl)-2-(5-((6-methoxy-2-oxo-1,2-dihydroquinolin-3-yl)methylene)-2,4-dioxothiazolidin-3-yl)acetamide 9.



White powder **(**yield, 73%); m. p. = 249-250ºC; IR (KBr, cm^{-1}): 3243, 3209 (NH), 2991, 2905 (aliphatic CH), 1684 (C=O); ^{1}H NMR (400 MHz, DMSO-d6) δ 10.65 (s, 1H), 10.25 (s, 1H), 8.45 (s, 1H), 7.47 (s, 1H), 7.43 (d, *J* = 2.7 Hz, 2H), 7.26 (d, *J* = 6.8 Hz, 2H), 7.21 (t, *J* = 5.4 Hz, 2H), 4.31 (s, 2H), 3.79 (s, 3H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 190.37, 161.55, 154.98, 142.24,

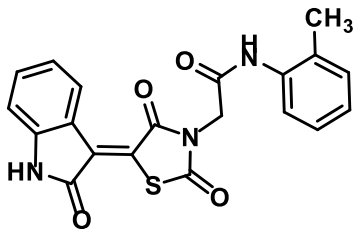
136.50, 134.90, 132.57, 129.36, 128.98, 126.19, 124.16, 119.16, 117.29, 111.57, 56.03, 46.90 for $C_{22}H_{15}Cl_2N_3O_5S$ (504.34).

4.3. (Z)-N-(2,6-Dichlorophenyl)-2-(2,4-dioxo-5-(2-oxoindolin-3-ylidene)thiazolidin-3-yl)acetamide 12.



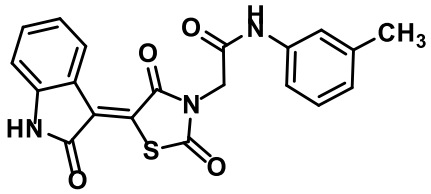
Yellowish crystals (yield 89%); mp: 166-168 °C; IR (KBr) *v* cm^{-1}: 3248, 3211 (NH), 3065 (aromatic CH), 2824 (aliphatic CH), 1740, 1688 (C≡O); ^{1}H NMR (400 MHz, DMSO) *δ* 11.31 (s, 1H, NH), 10.38 (s, 1H, NH), 8.80 (d, *J* = 7.9 Hz, 1H, Ar-H), 7.57 (d, *J* = 8.1 Hz, 2H, Ar-H), 7.43 (d, *J* = 7.6 Hz, 1H, Ar-H), 7.37 (d, *J* = 8.1 Hz, 1H, Ar-H), 7.13 (d, *J* = 7.7 Hz, 1H, Ar-H), 6.99 (d, *J* = 7.7 Hz, 1H, Ar-H), 4.59 (s, 2H, CH_2); ^{13}C NMR (101 MHz, DMSO) *δ* 170.09, 168.72, 165.52, 164.67, 144.59, 133.97 (2), 133.53, 132.56, 130.04, 129.63, 129.09 (2), 128.39, 128.07, 122.69, 120.23, 111.15, 43.43 for $C_{19}H_{11}Cl_2N_3O_4S$ (448.27).

***4.4.(Z)-2-(2,4-Dioxo-5-(2-oxoindolin-3-ylidene)thiazolidin-3-yl)-N-(o-tolyl)acetamide* 13.**



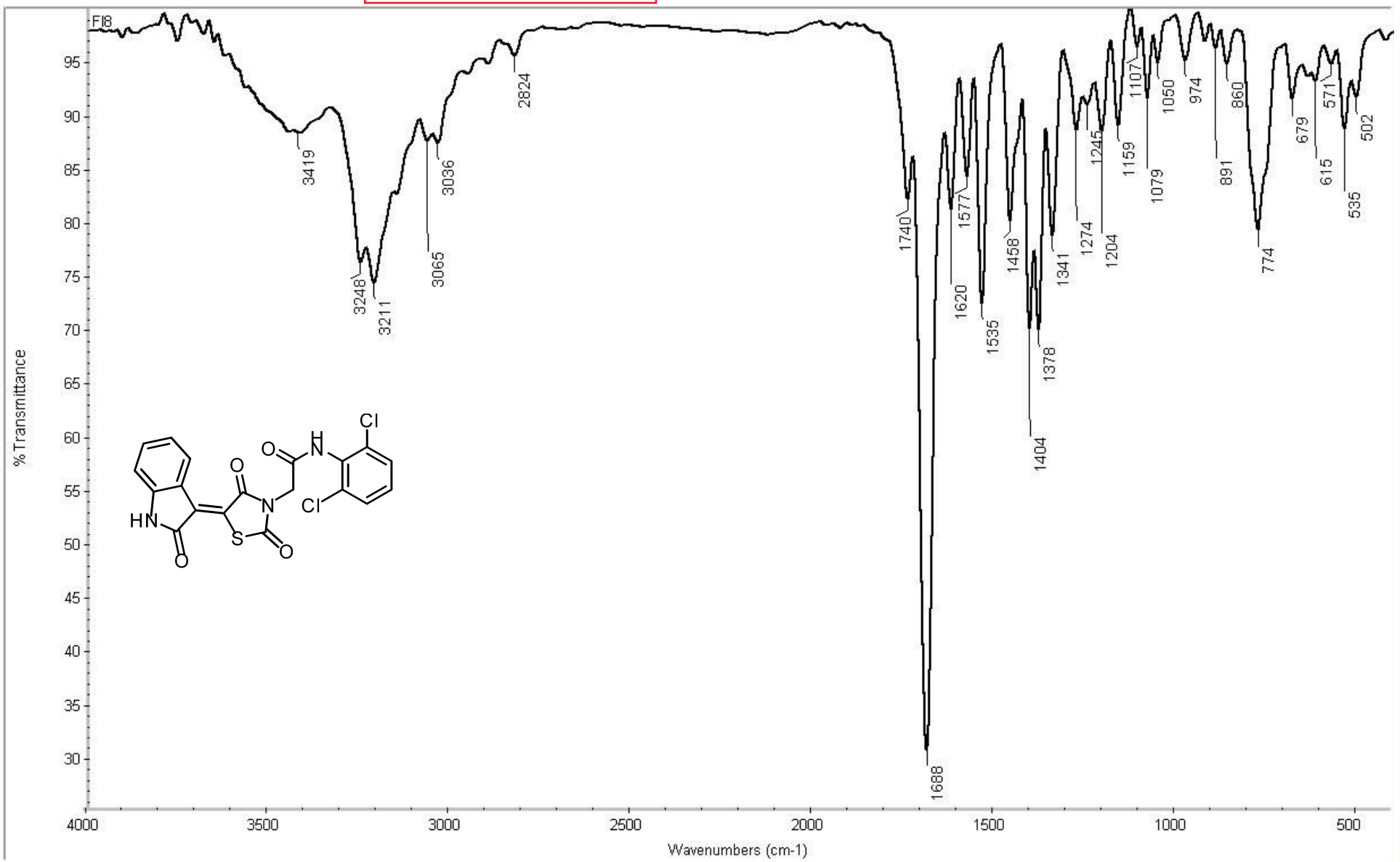
Off-white powder **(**yield, 76%); m. p. = 160-162ºC; IR (KBr, cm^{-1}): 3289, 3178 (NH), 3065 (aromatic CH) 2953 (aliphatic CH), 1745, 1695 (C=O); ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.32 (s, 1H), 9.80 (s, 1H), 8.81 (d, *J* = 8.0 Hz, 1H), 7.44 (d, *J* = 8.2 Hz, 1H), 7.38 (d, *J* = 7.8 Hz, 1H), 7.24 (d, *J* = 7.3 Hz, 1H), 7.12 (t, *J* = 7.9 Hz, 2H), 6.99 (d, *J* = 7.9 Hz, 1H), 4.61 (s, 2H), 2.24 (s, 3H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 170.30, 168.75, 165.74, 164.49, 144.58, 135.96, 133.55, 132.43, 130.91, 129.68, 128.40, 128.05, 126.56, 126.17, 125.48, 122.70, 120.24, 111.17, 43.90, 18.25 for $C_{20}H_{15}N_3O_4S$ (393.42).

4.5. (Z)-2-(2,4-Dioxo-5-(2-oxoindolin-3-ylidene)thiazolidin-3-yl)-N-(m-tolyl)acetamide 14.

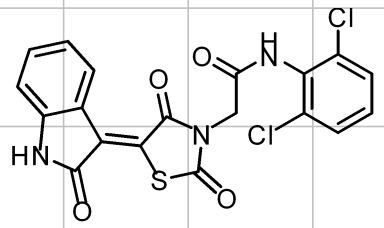


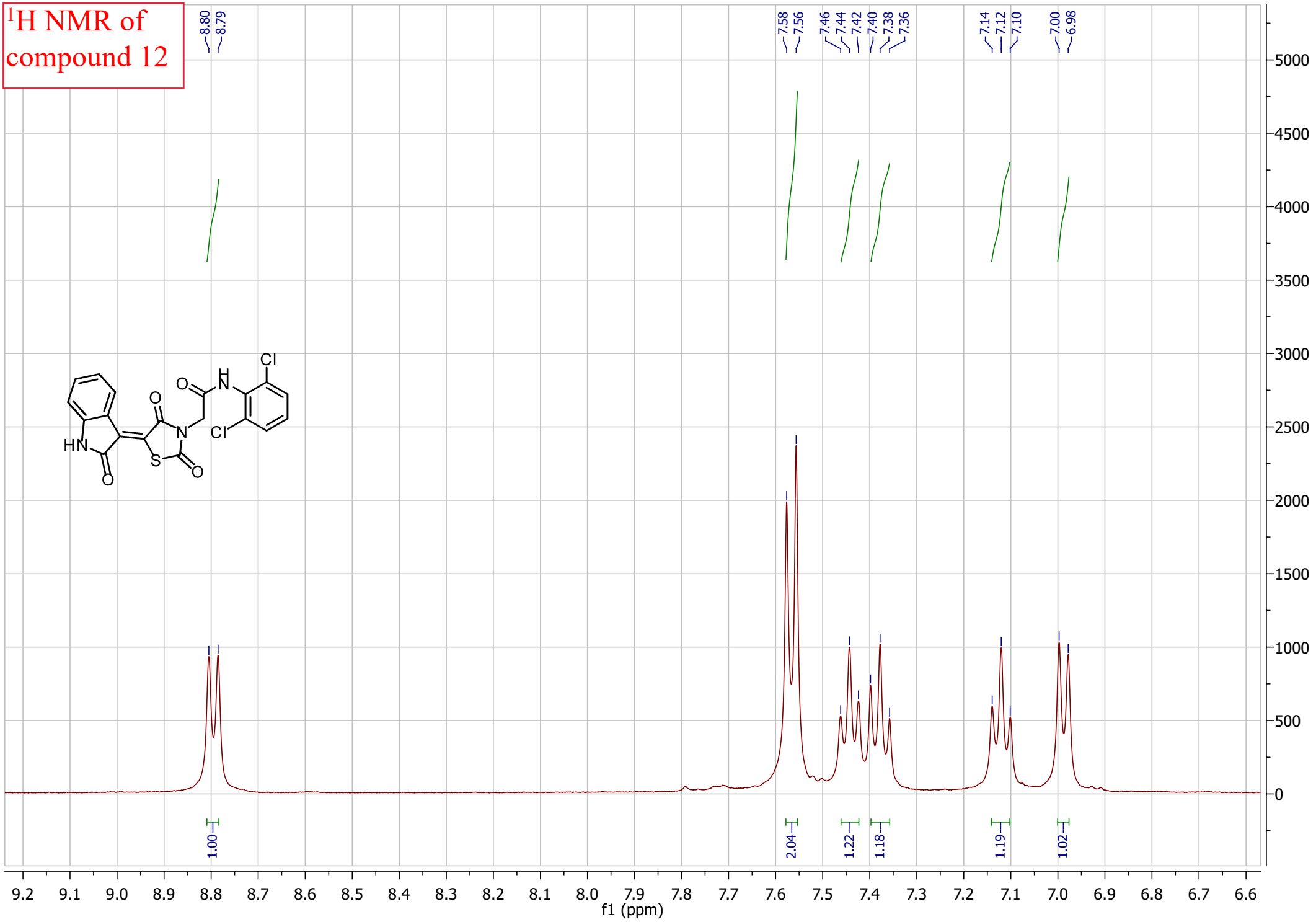
White powder **(**yield, 81%); m. p. = 158-160ºC; IR (KBr, cm^{-1}): 3214, 3148 (NH), 3065 (aromatic CH) 2943 (aliphatic CH), 1743, 1693 (C=O); ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.32 (s, 1H), 10.36 (s, 1H), 8.77 (d, *J* = 7.9 Hz, 1H), 7.49 – 7.40 (m, 2H), 7.35 (d, *J* = 8.1 Hz, 1H), 7.21 (t, *J* = 7.7 Hz, 1H), 7.09 (t, *J* = 7.7 Hz, 1H), 6.99 (d, *J* = 7.8 Hz, 1H), 6.91 (d, *J* = 7.6 Hz, 1H), 4.56 (s, 2H), 2.28 (s, 3H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 170.25, 168.71, 165.69, 164.15, 144.62, 138.76, 138.60, 133.55, 129.38, 129.20, 128.42, 128.18, 124.92, 122.65, 120.24, 120.22, 116.88, 111.15, 44.13, 21.62 for $C_{20}H_{15}N_3O_4S$ (393.42).

IR of compound 12

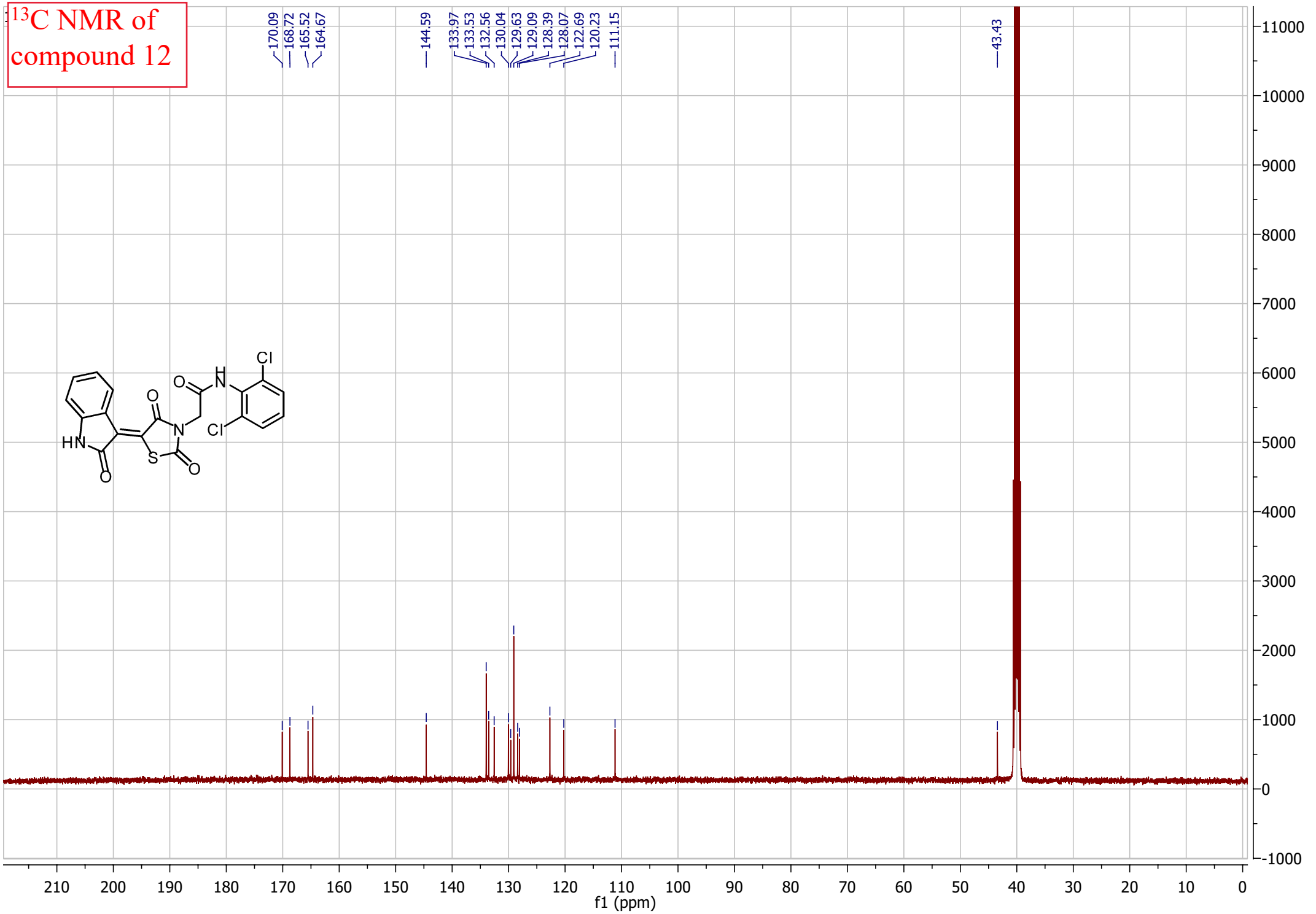


^{1}H NMR of compound 12

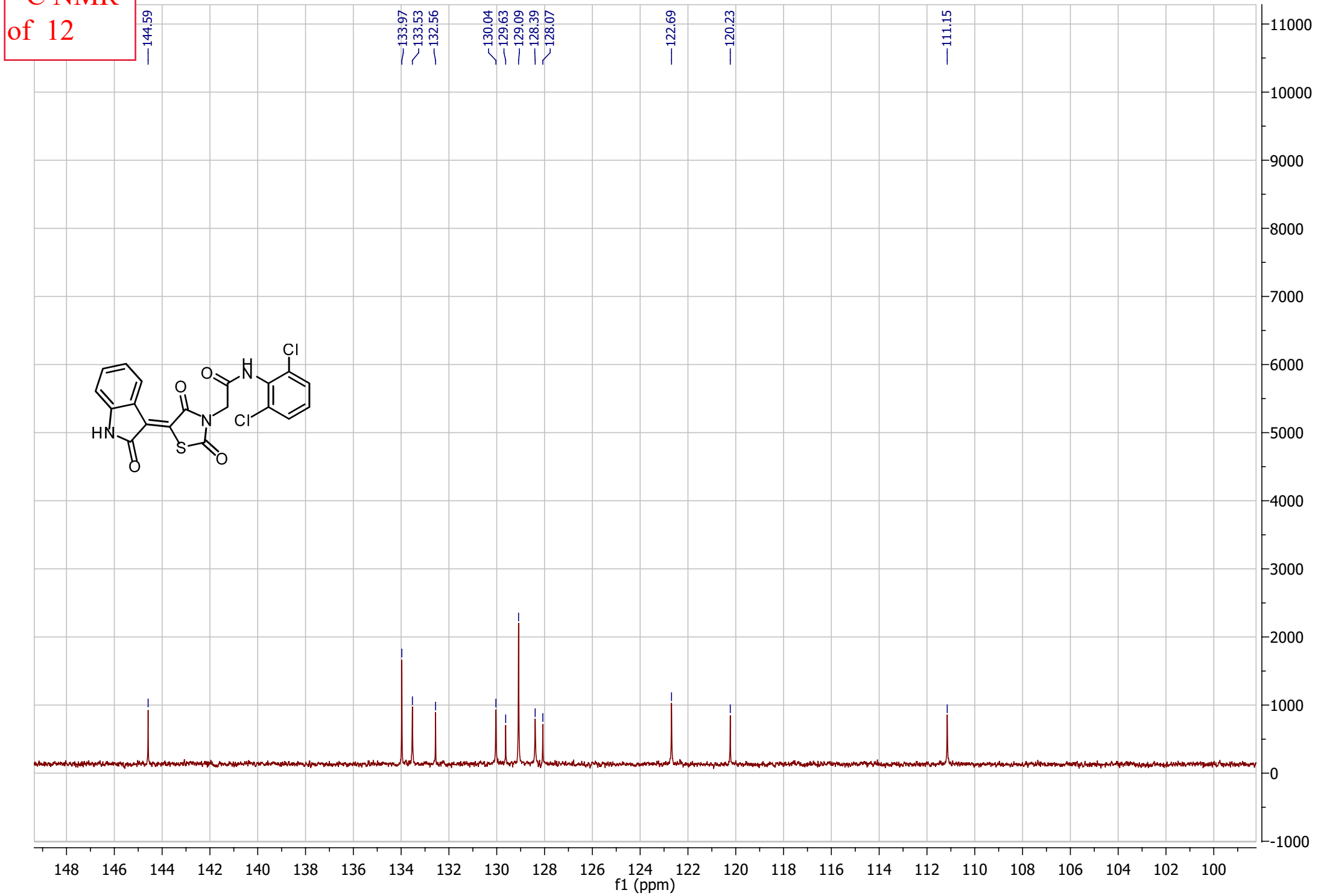



[1]H NMR of compound 12
8.80
8.79
7.58
7.56
7.46
7.44
7.42
7.40
7.38
7.36
7.14
7.12
7.10
7.00
6.98
HN
O
O
N
S
O
O
H
N
Cl
Cl
1.00
2.04
1.22
1.18
1.19
1.02
f1 (ppm)

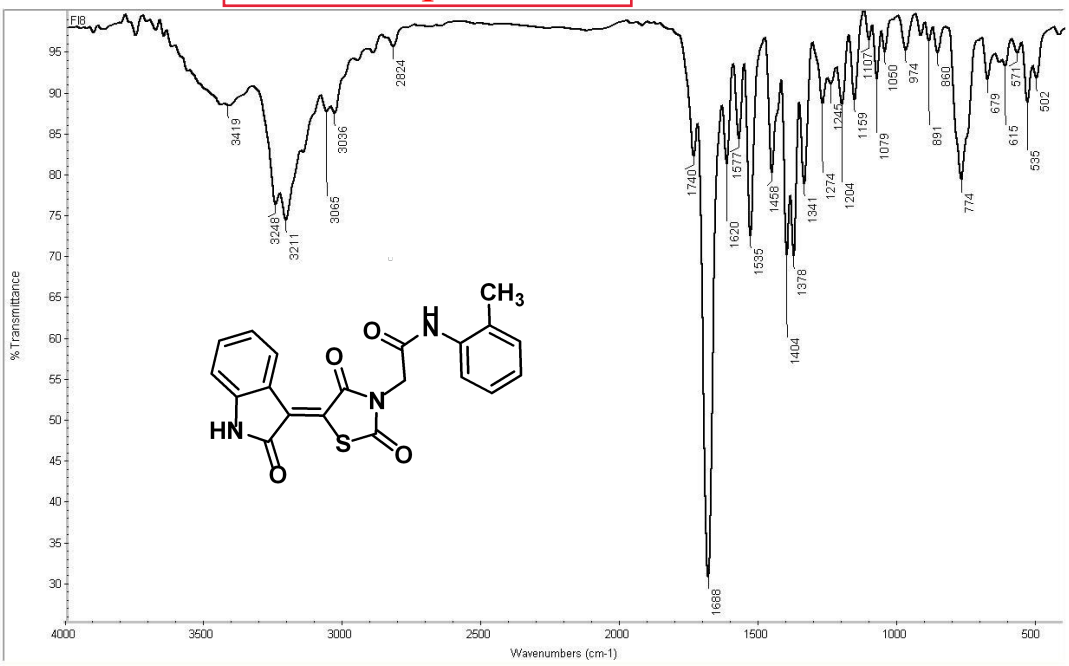
^{13}C NMR of compound 12

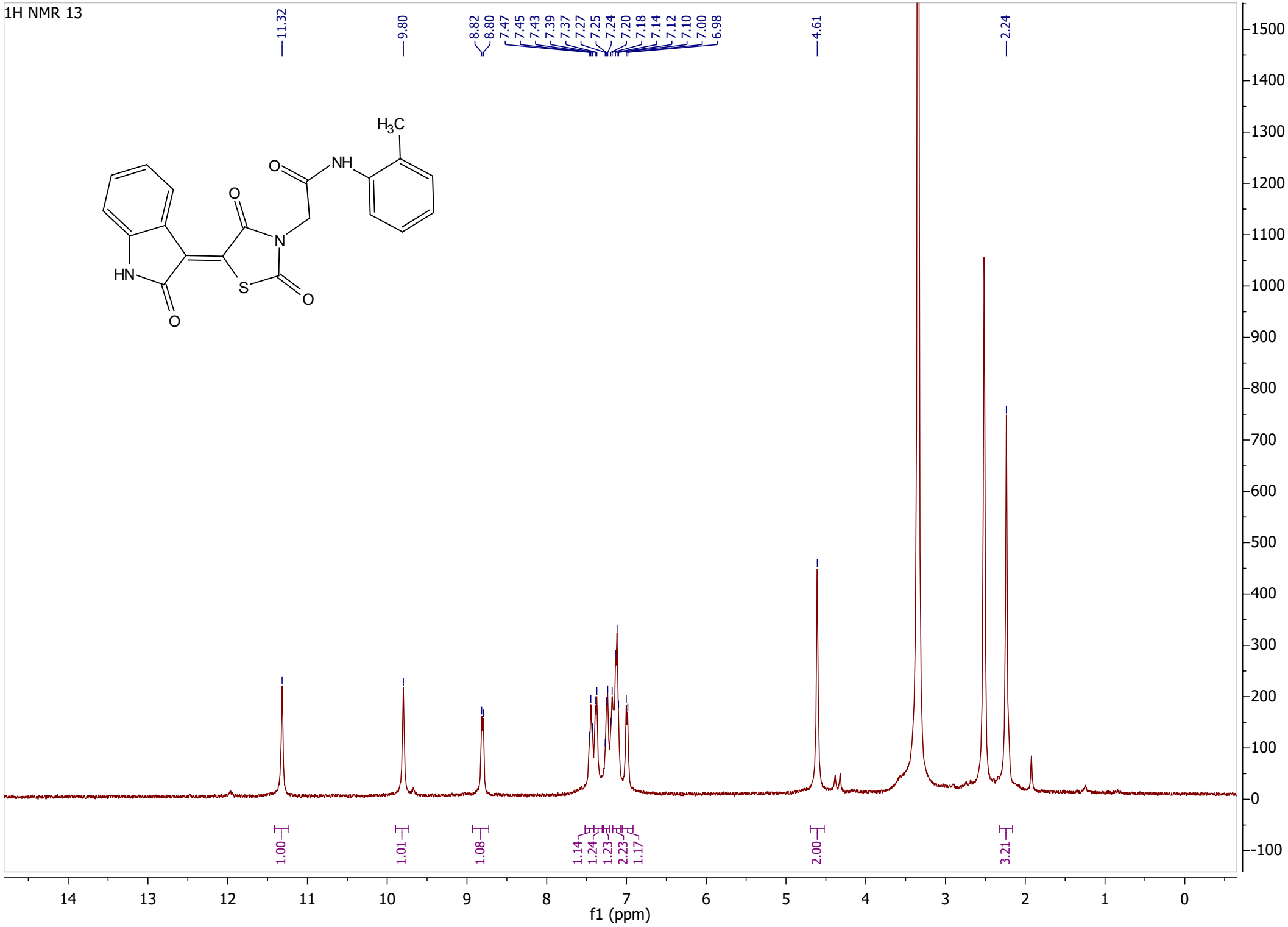


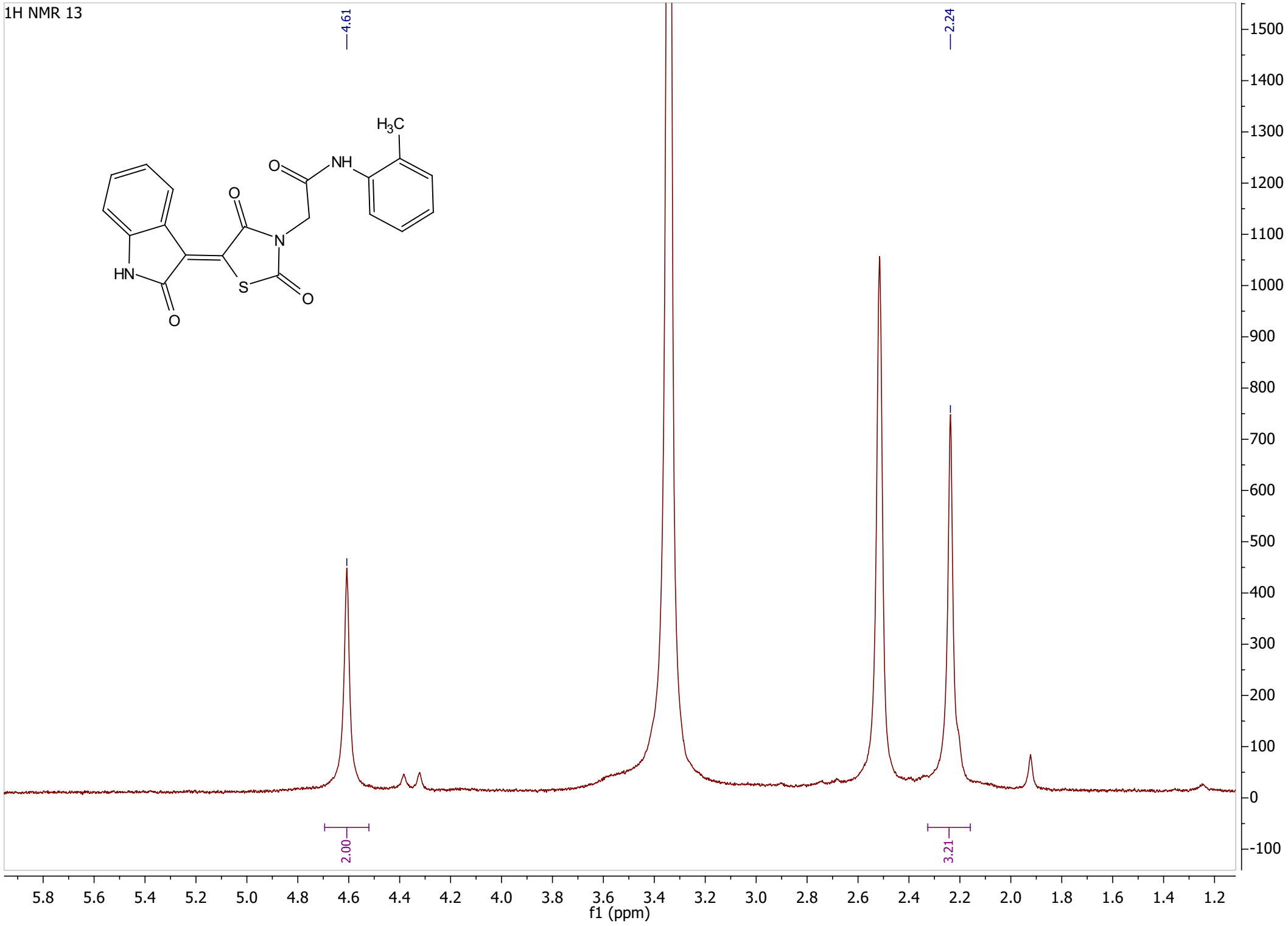
^{13}C NMR
of 12

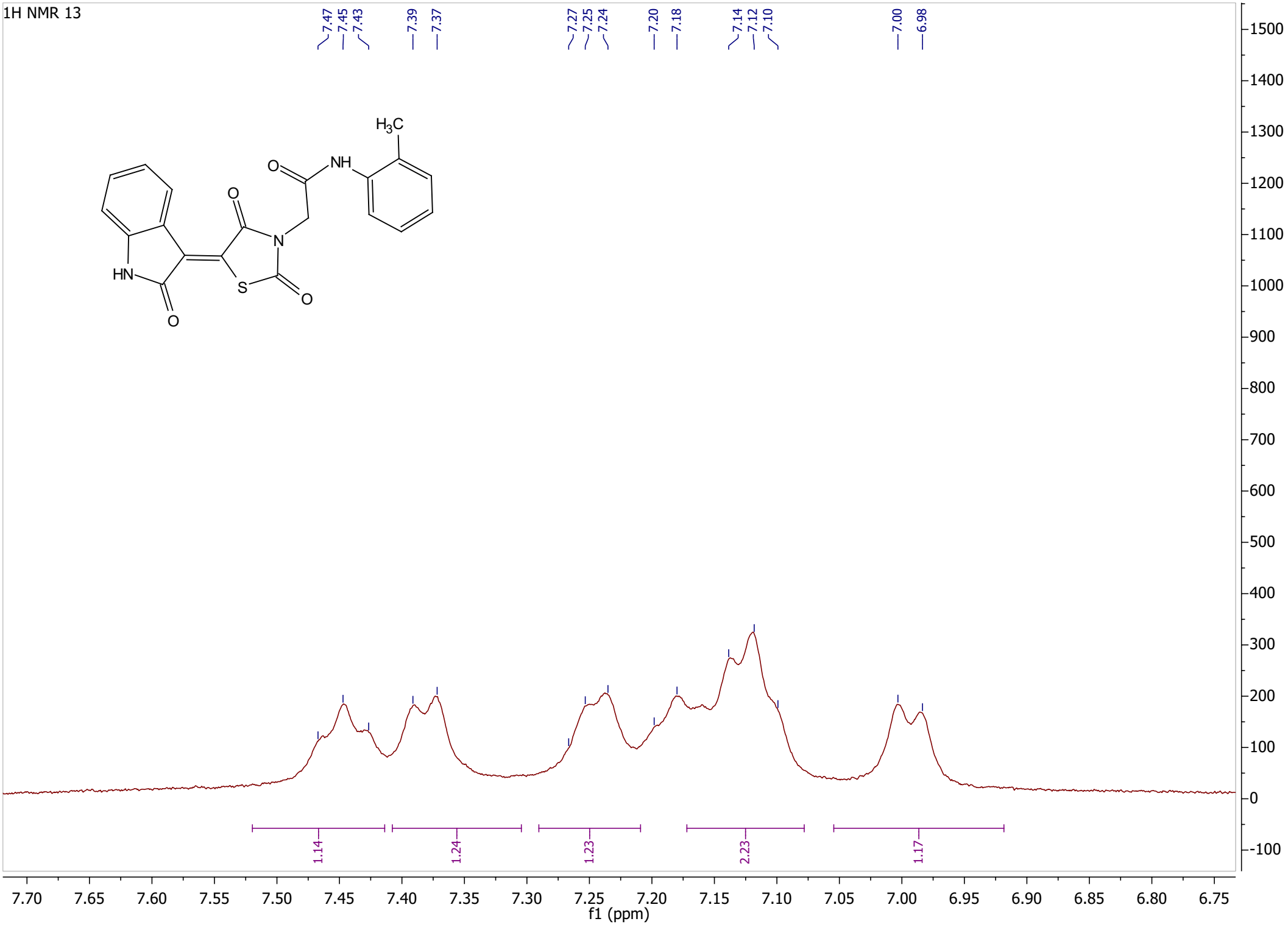

144.59
133.97
133.53
132.56
130.04
129.63
129.09
128.39
128.07
122.69
120.23
111.15
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
148
146
144
142
140
138
136
134
132
130
128
126
124
122
120
118
116
114
112
110
108
106
104
102
100
f1 (ppm)
HN
O
O
N
S
O
O
N
H
Cl
Cl

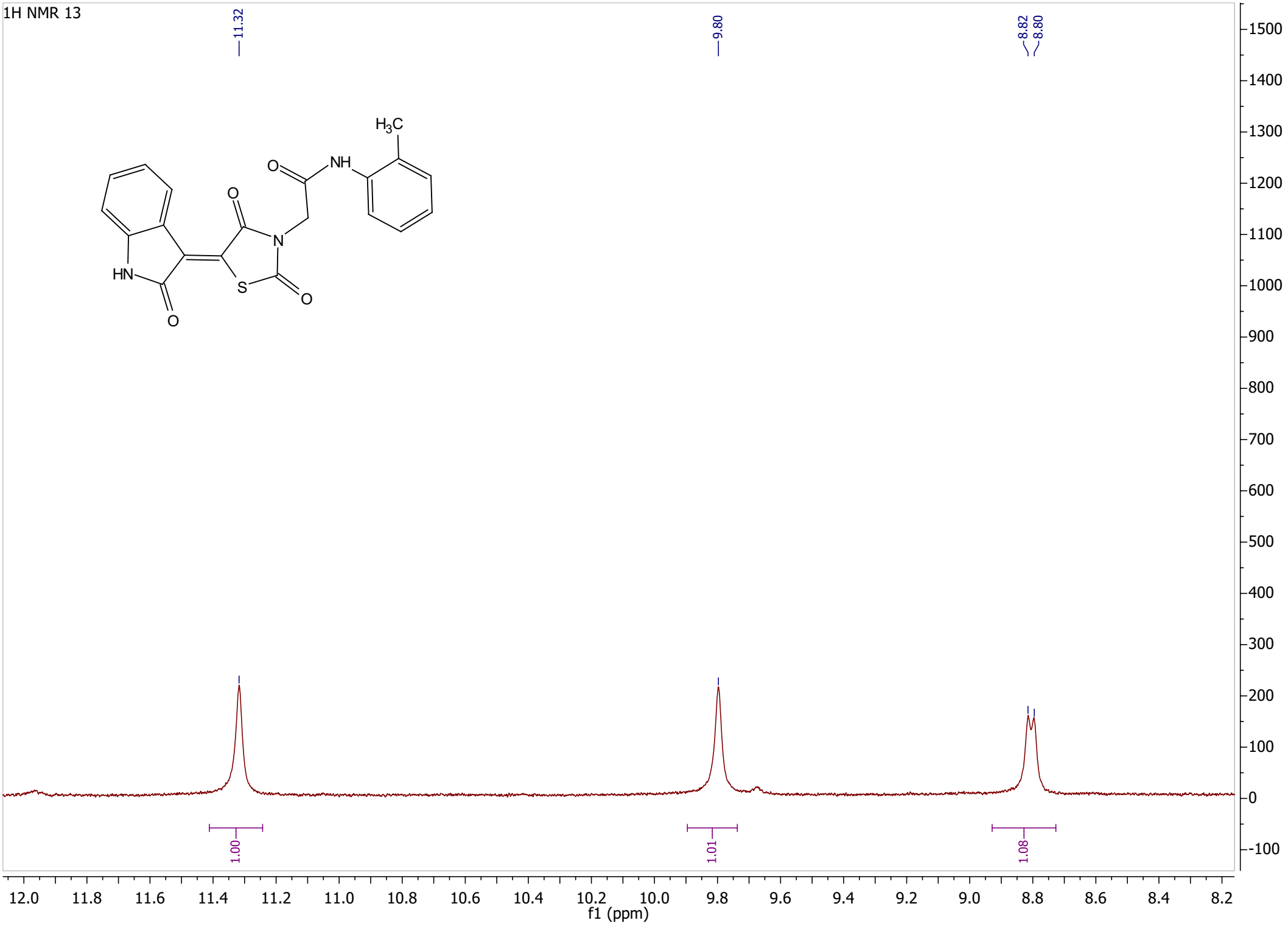
IR of compound 13

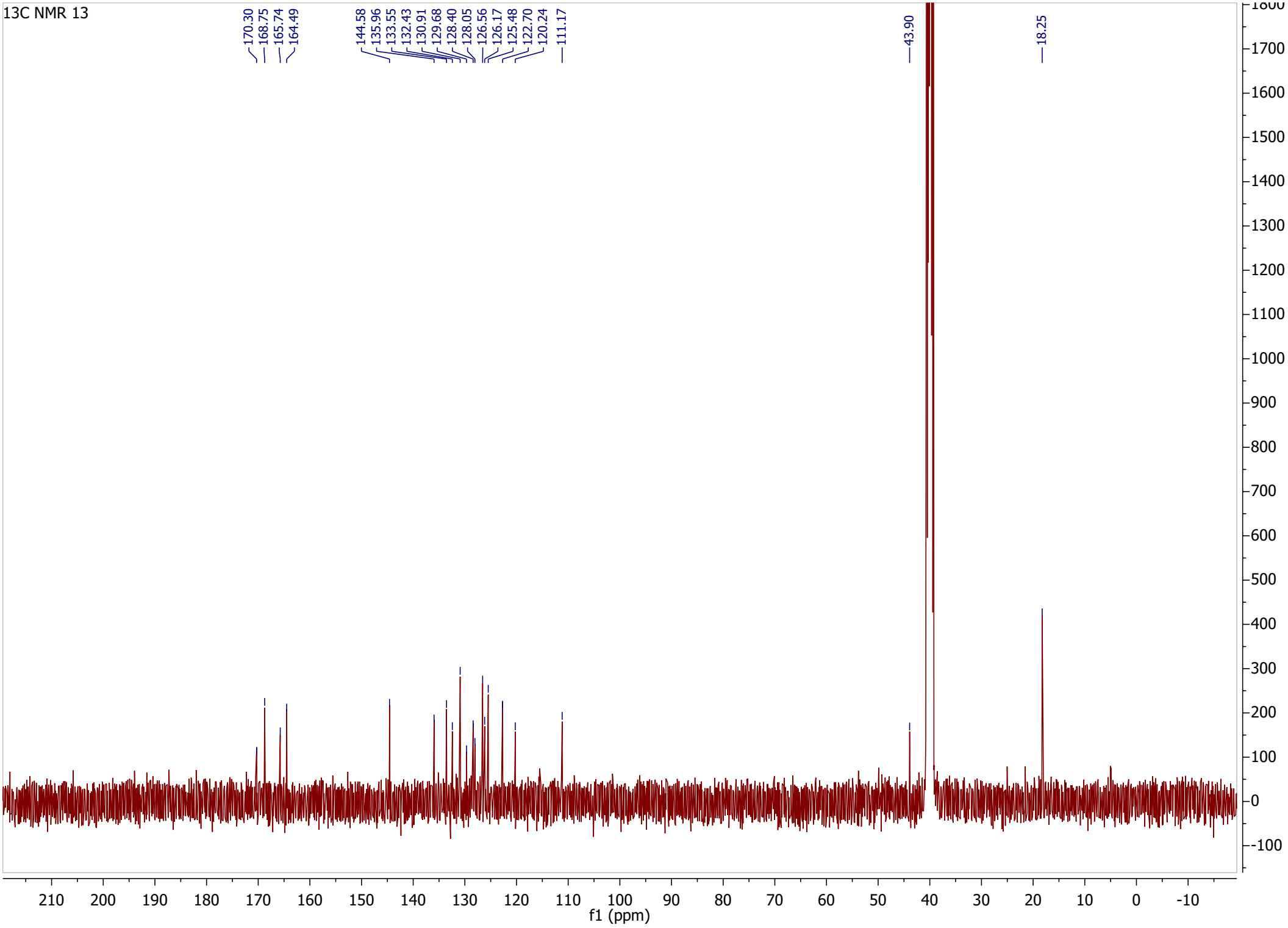


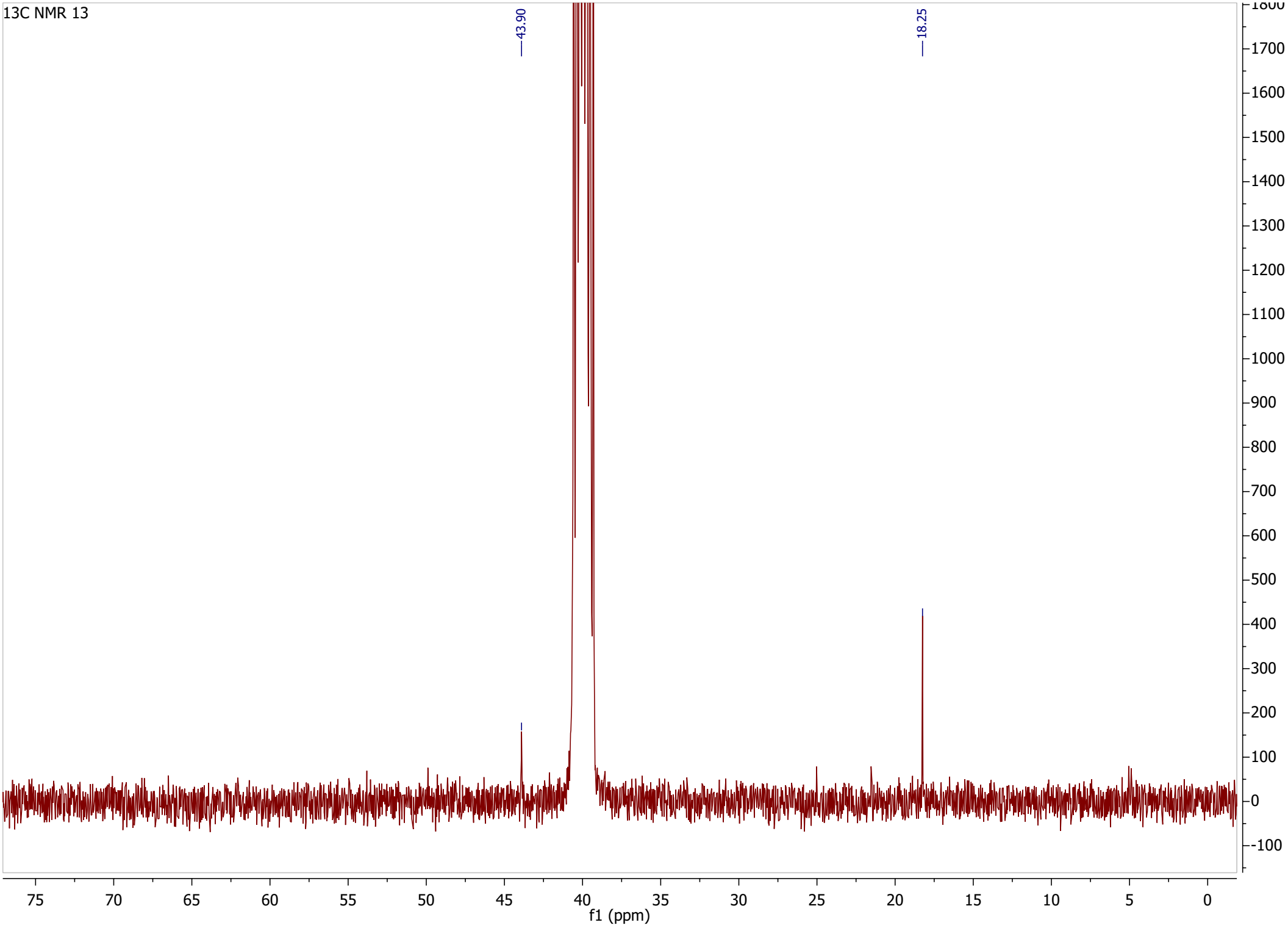

1H NMR 13
11.32
9.80
8.82
8.80
7.47
7.45
7.43
7.39
7.37
7.27
7.25
7.24
7.20
7.18
7.14
7.12
7.10
7.00
6.98
4.61
2.24
H3C
NH
O
O
N
HN
S
O
O
1.00
1.01
1.08
1.14
1.24
1.23
2.23
1.17
2.00
3.21
f1 (ppm)

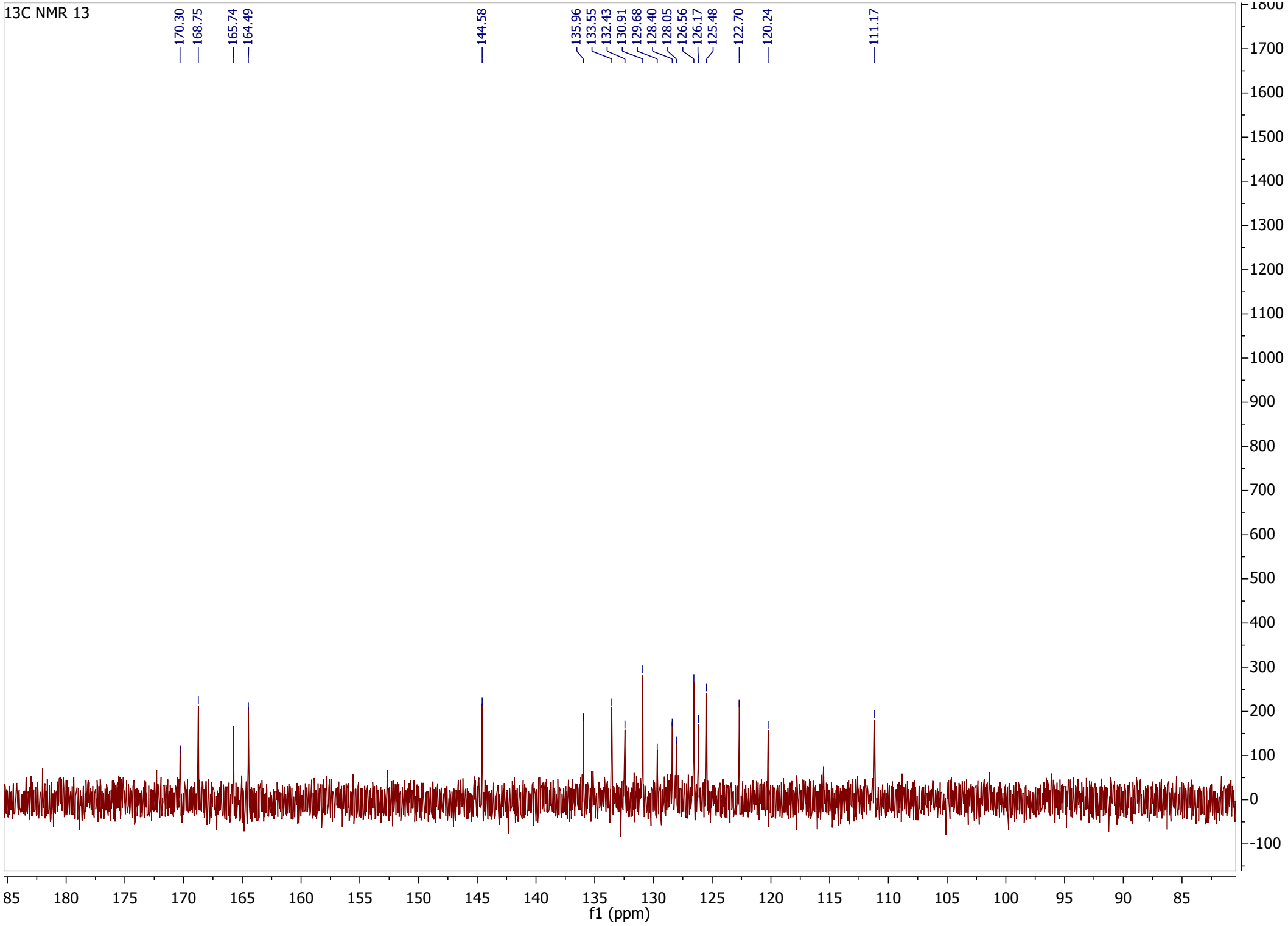

1H NMR 13
4.61
2.24
2.00
3.21
H3C
NH
O
O
N
HN
S
O
O
5.8
5.6
5.4
5.2
5.0
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
f1 (ppm)
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
-100


1H NMR 13
7.47
7.45
7.43
7.39
7.37
7.27
7.25
7.24
7.20
7.18
7.14
7.12
7.10
7.00
6.98
H3C
NH
O
O
N
HN
S
O
O
1.14
1.24
1.23
2.23
1.17
f1 (ppm)

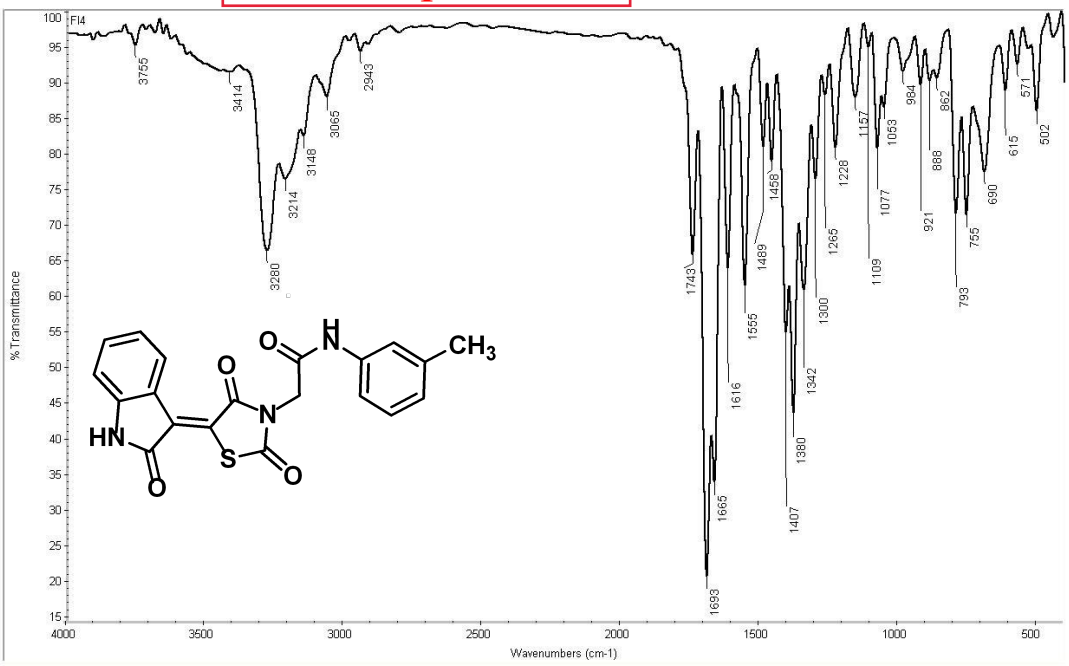

1H NMR 13
11.32
9.80
8.82
8.80
H3C
NH
O
O
N
HN
S
O
O
1.00
1.01
1.08
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
-100
12.0
11.8
11.6
11.4
11.2
11.0
10.8
10.6
10.4
10.2
10.0
9.8
9.6
9.4
9.2
9.0
8.8
8.6
8.4
8.2
f1 (ppm)

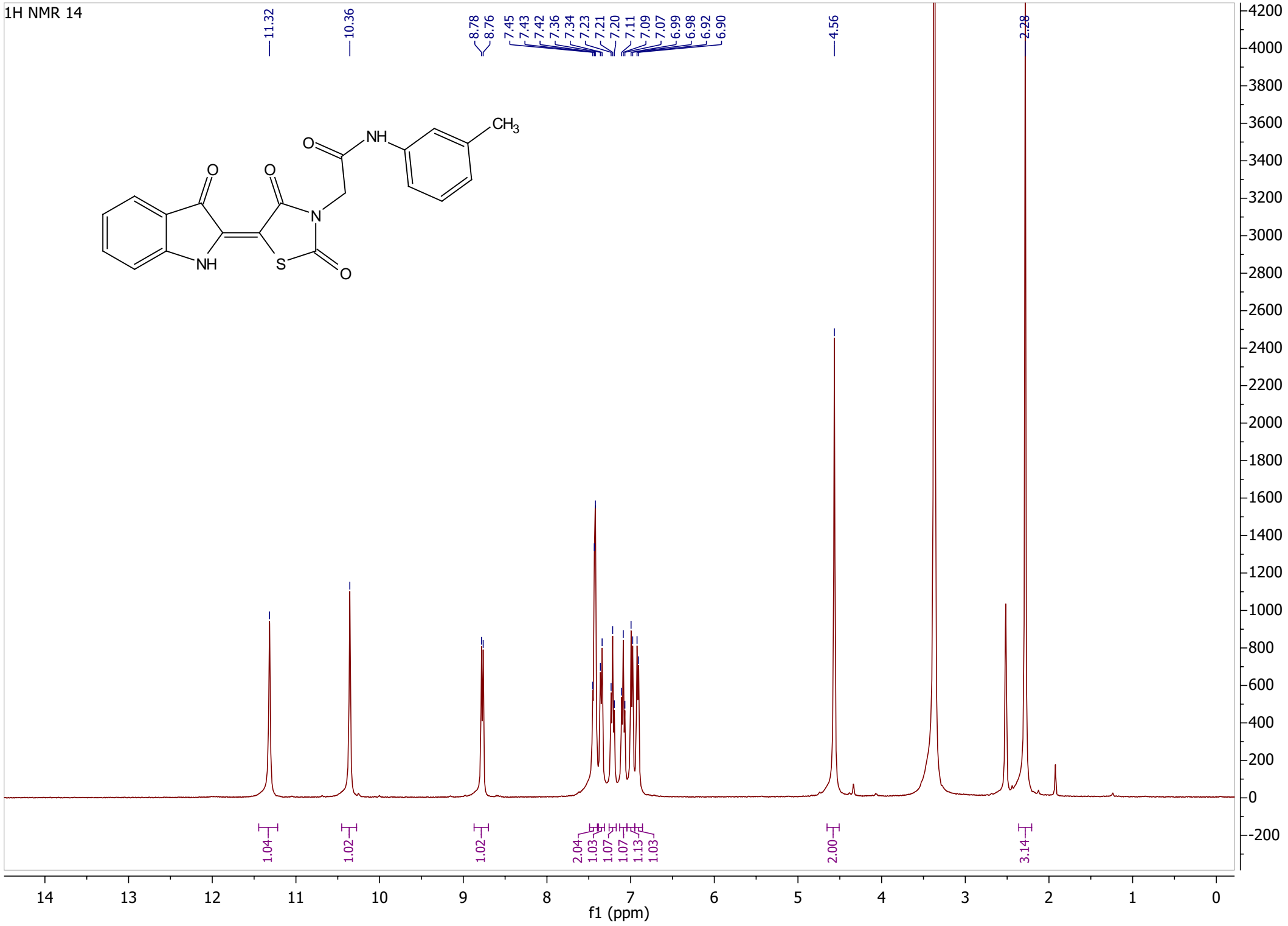

13C NMR 13
170.30
168.75
165.74
164.49
144.58
135.96
133.55
132.43
130.91
129.68
128.40
128.05
126.56
126.17
125.48
122.70
120.24
111.17
43.90
18.25
f1 (ppm)

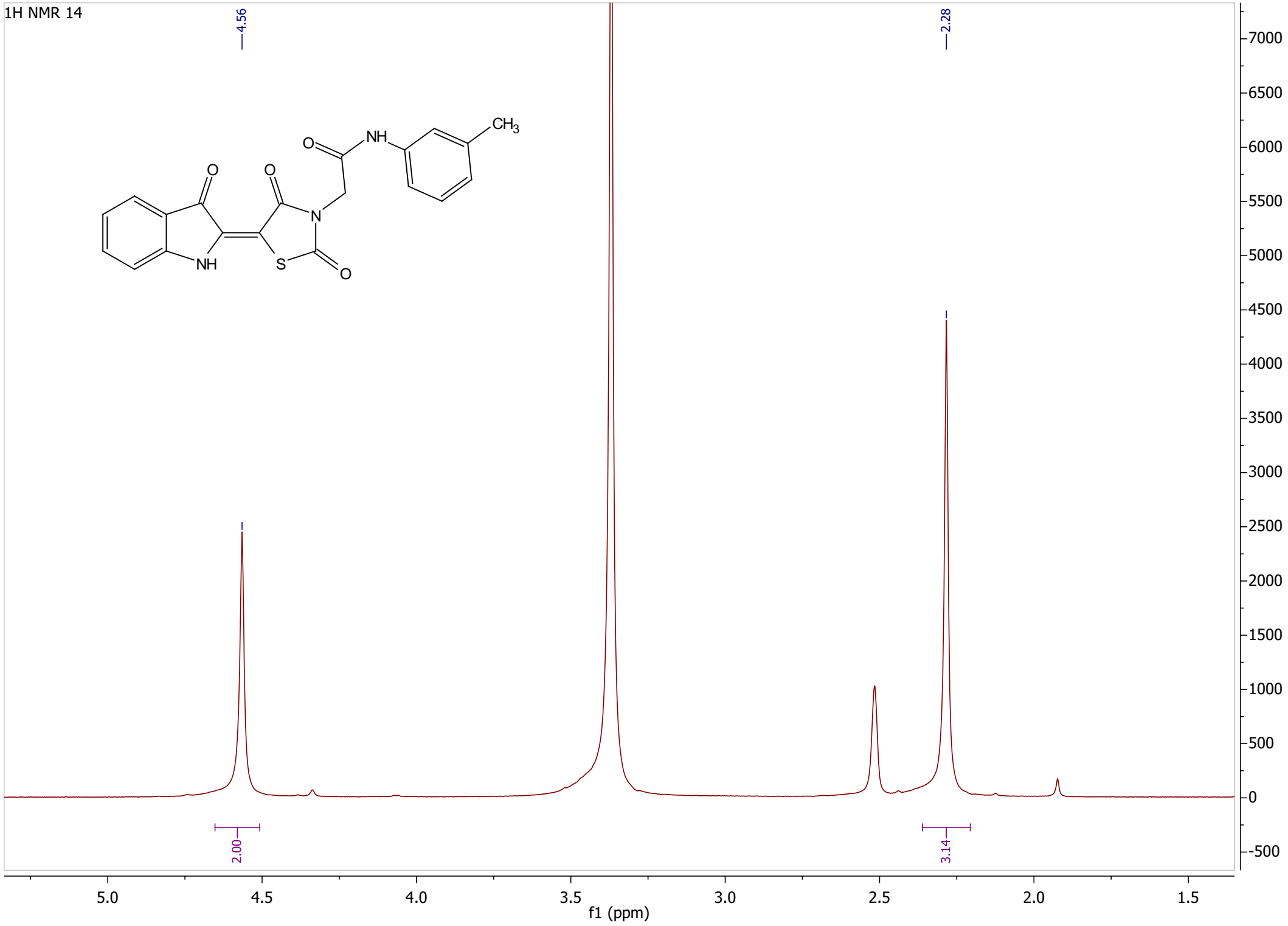

13C NMR 13
43.90
18.25
f1 (ppm)

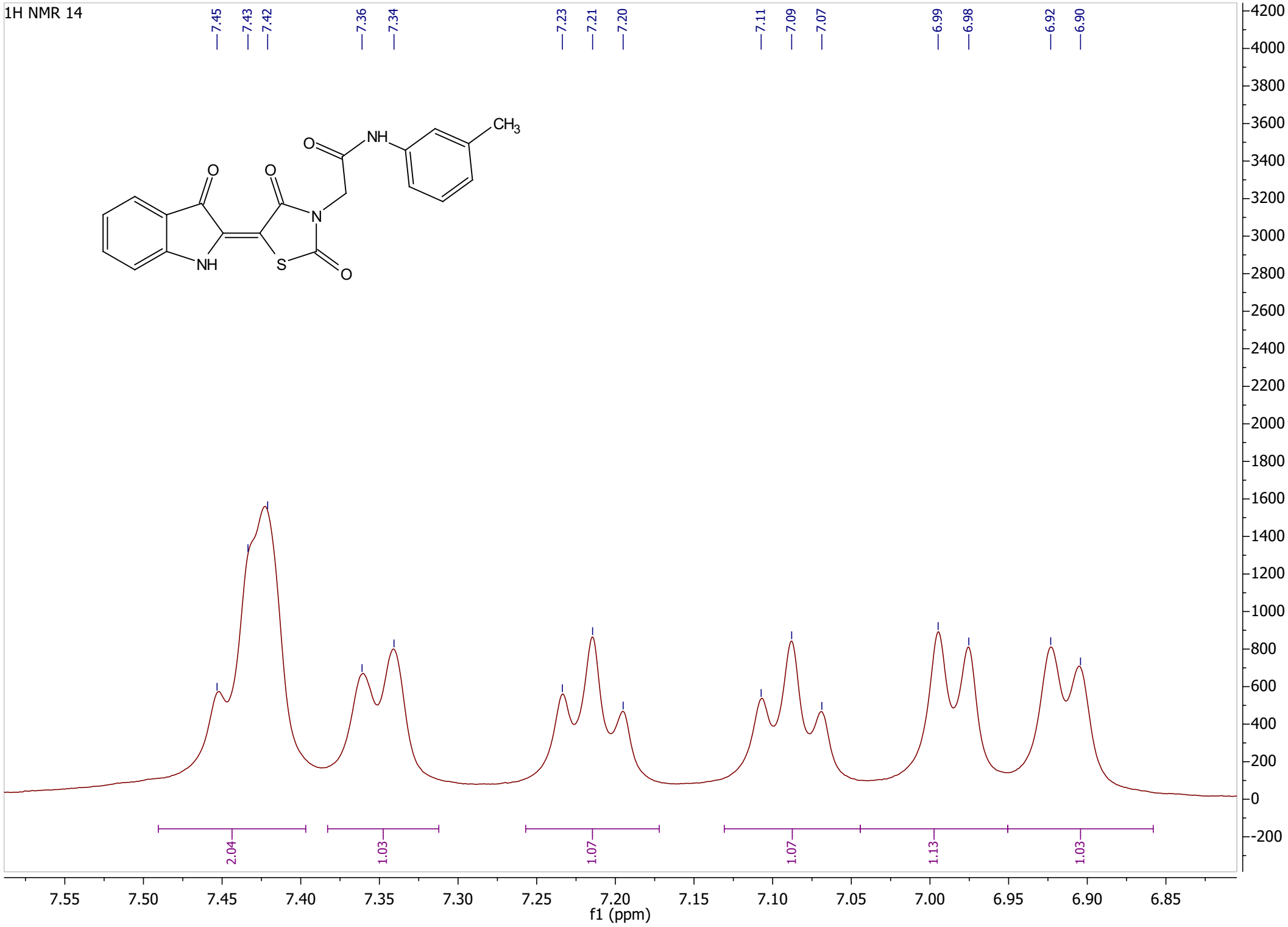

13C NMR 13
170.30
168.75
165.74
164.49
144.58
135.96
133.55
132.43
130.91
129.68
128.40
128.05
126.56
126.17
125.48
122.70
120.24
111.17
f1 (ppm)

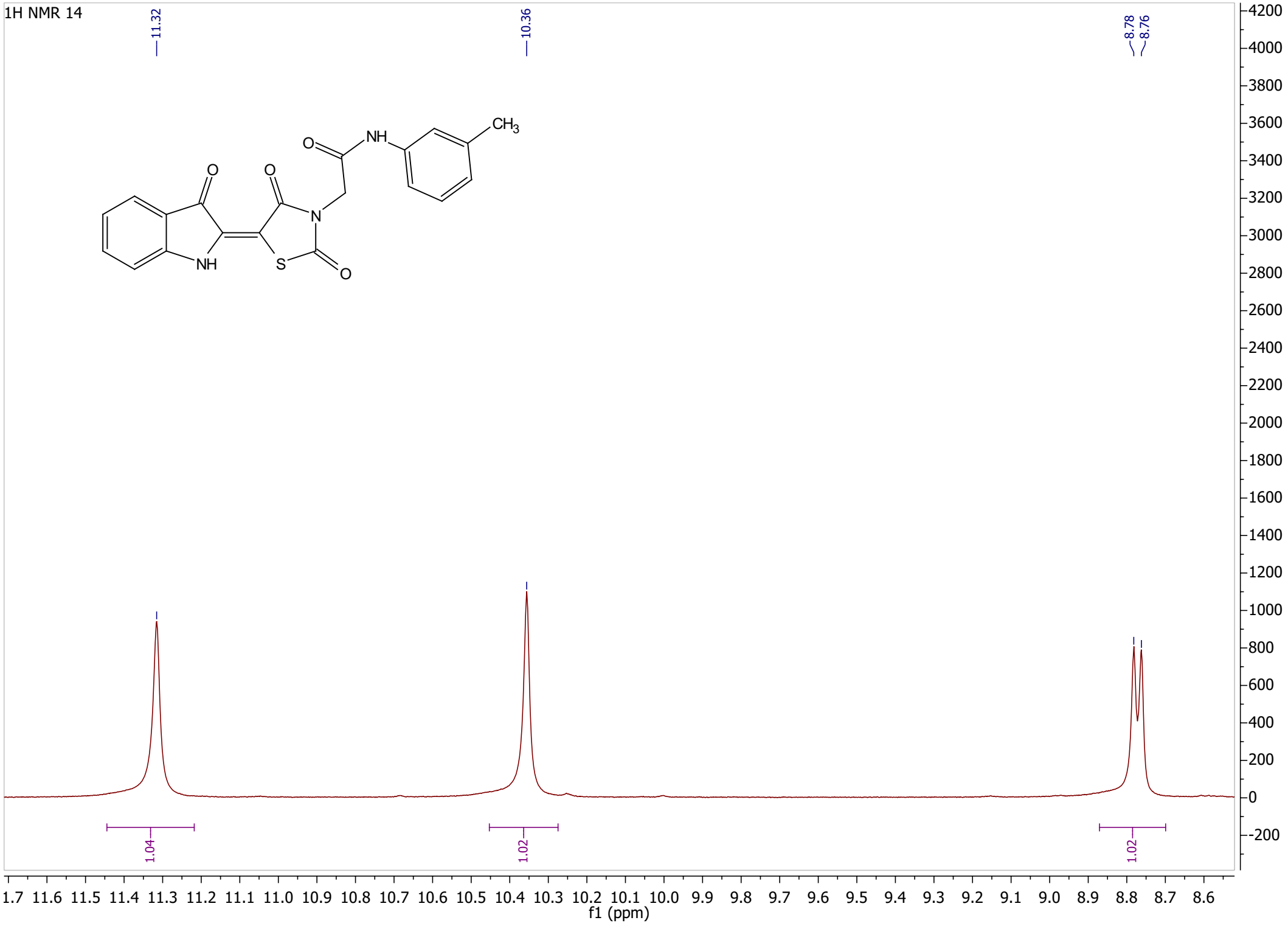
IR of compound 14

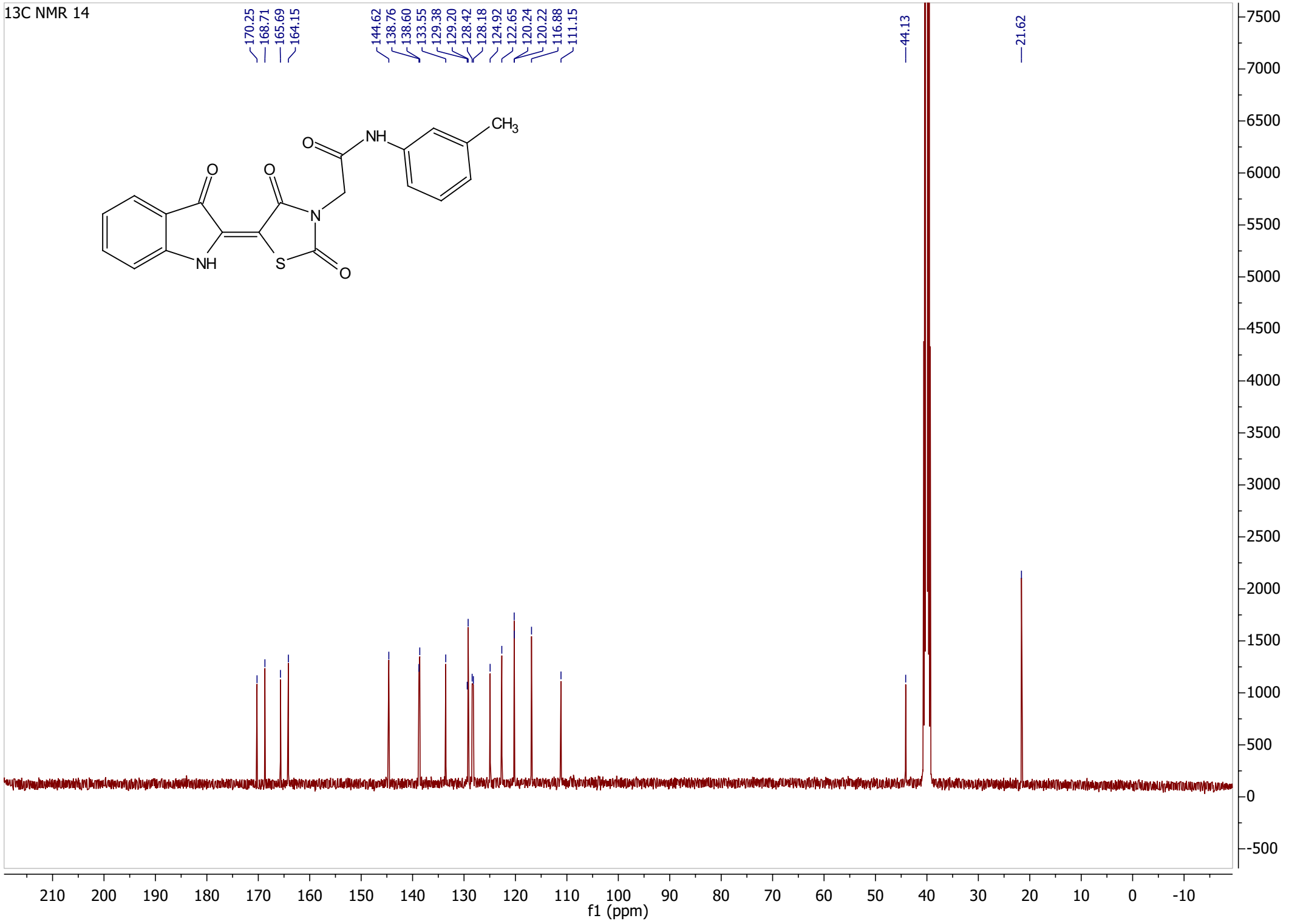


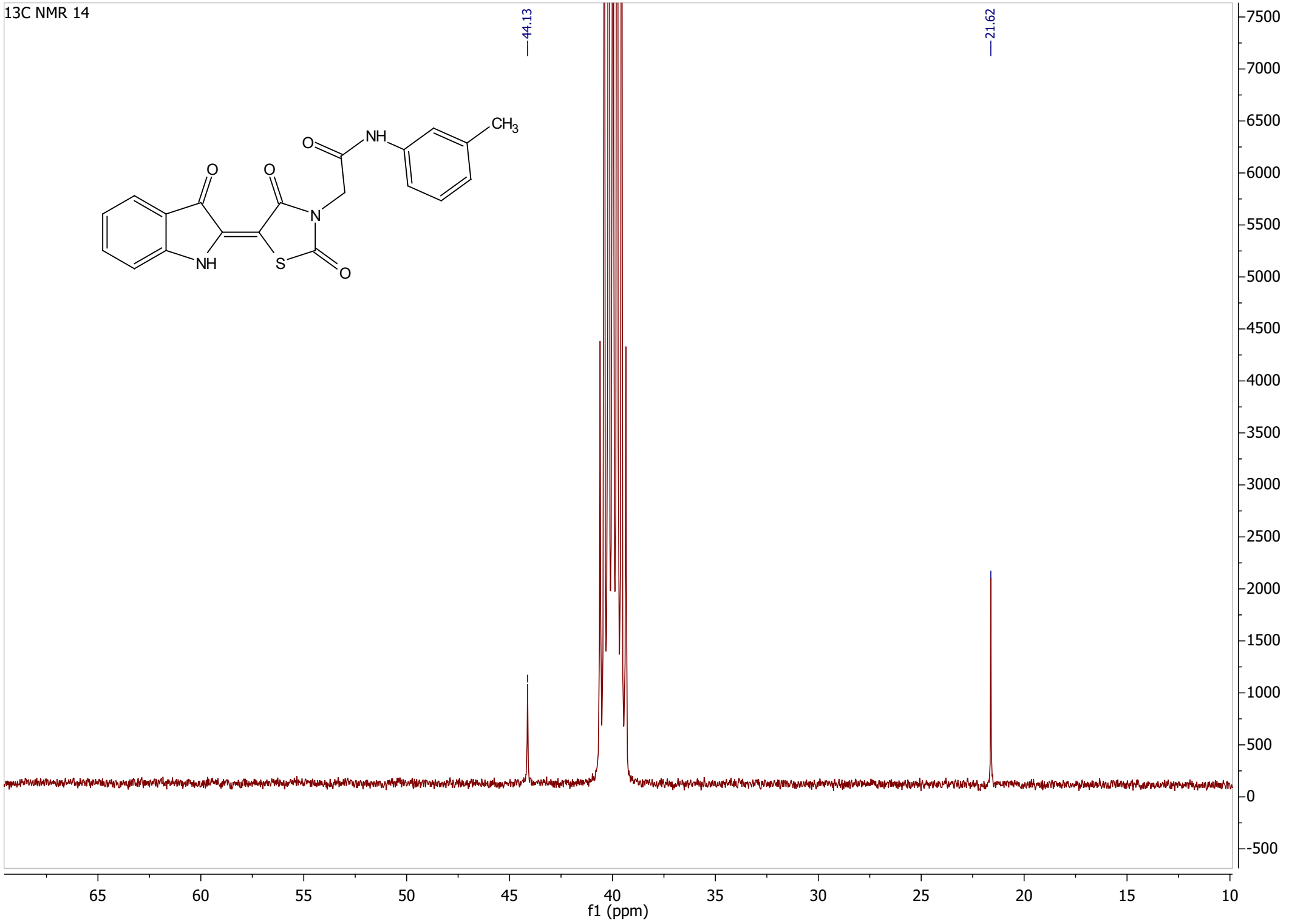

1H NMR 14
11.32
10.36
8.78
8.76
7.45
7.43
7.42
7.36
7.34
7.23
7.21
7.20
7.11
7.09
7.07
6.99
6.98
6.92
6.90
4.56
2.28
O
O
O
NH
CH3
N
NH
S
O
1.04
1.02
1.02
2.04
1.03
1.07
1.07
1.13
1.03
2.00
3.14
f1 (ppm)

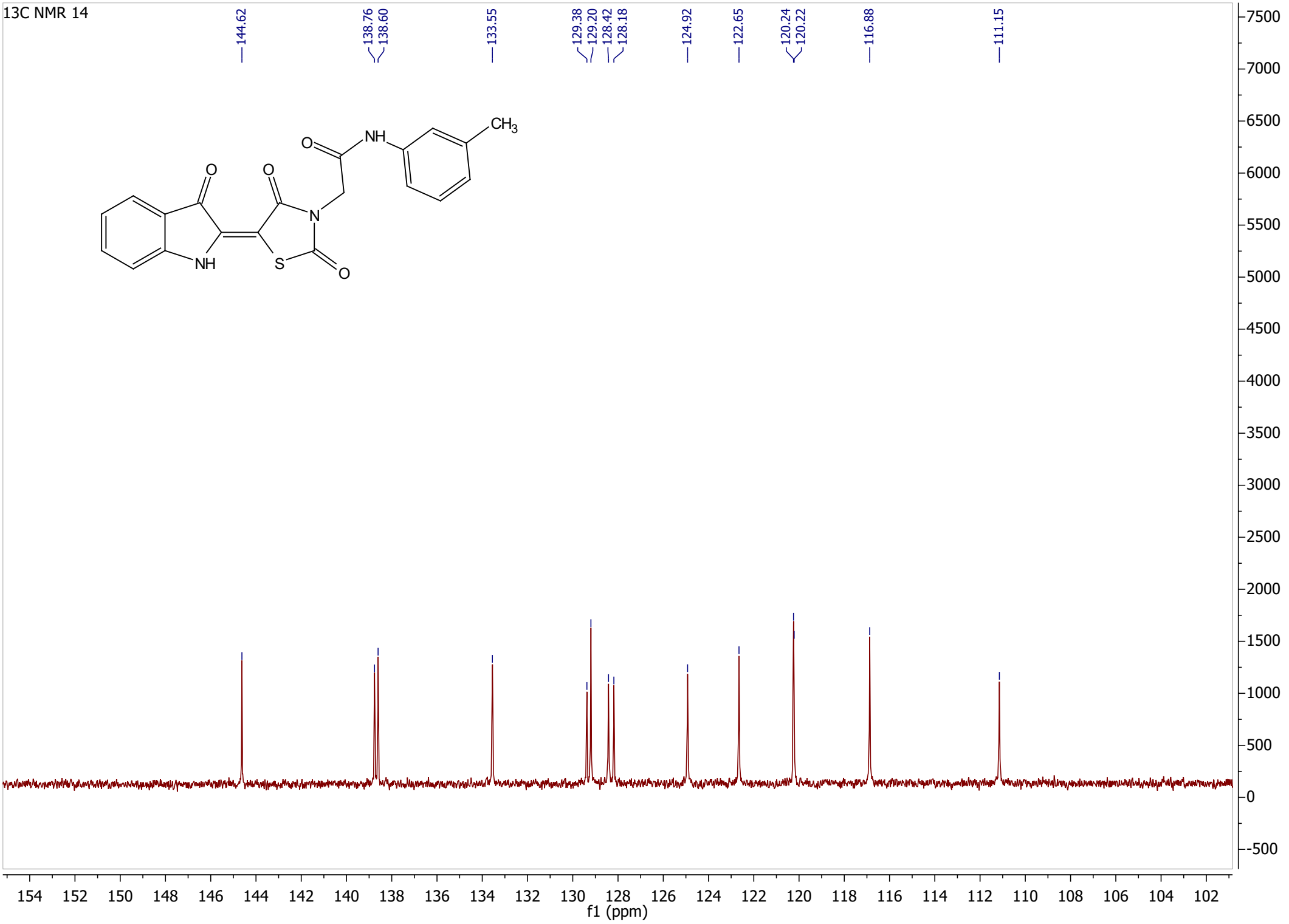

1H NMR 14
4.56
2.28
O
O
NH
CH3
O
N
NH
S
O
2.00
3.14
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
-500
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
f1 (ppm)

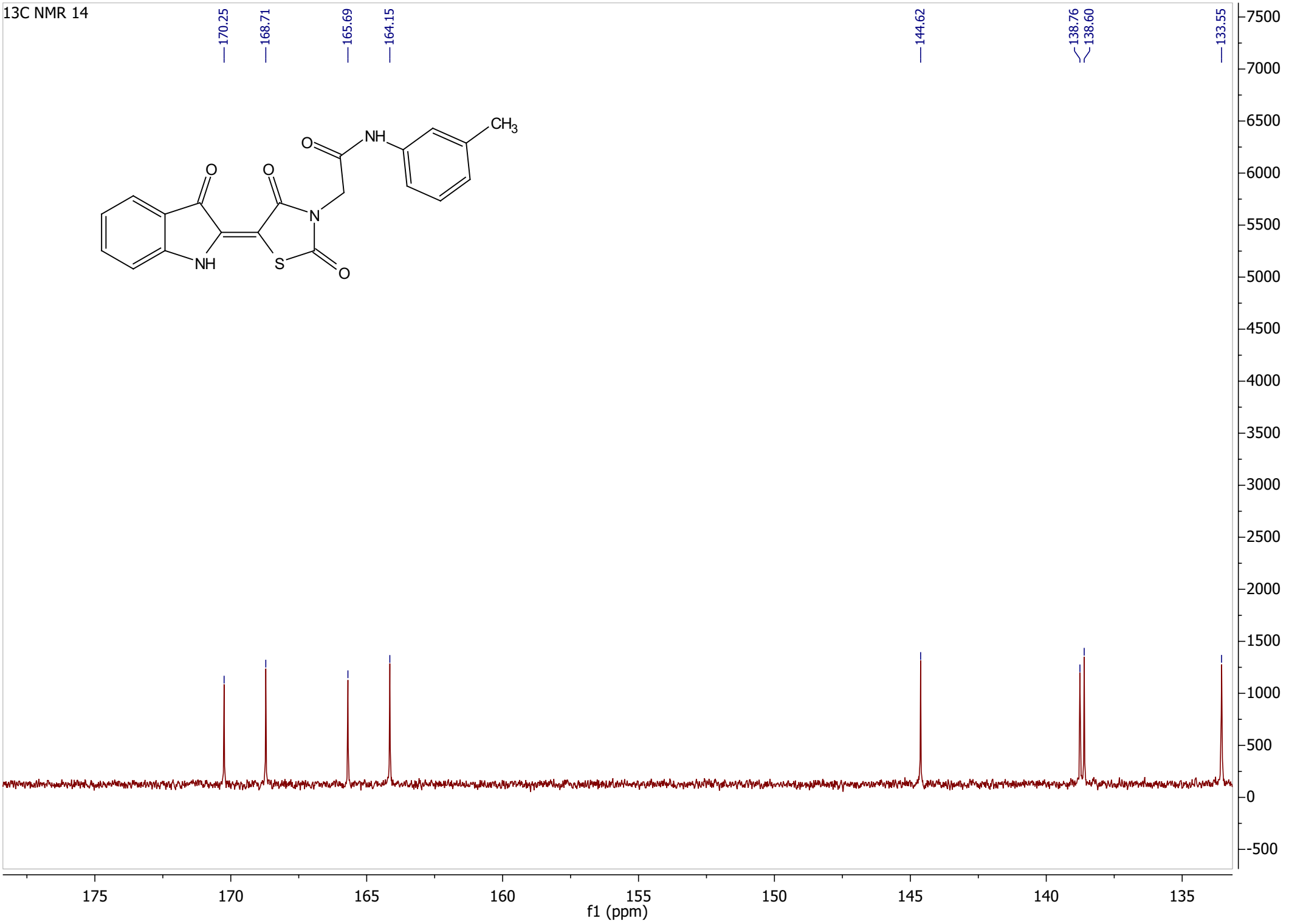

1H NMR 14
7.45
7.43
7.42
7.36
7.34
7.23
7.21
7.20
7.11
7.09
7.07
6.99
6.98
6.92
6.90
O
O
NH
CH3
O
N
NH
S
O
2.04
1.03
1.07
1.07
1.13
1.03
7.55
7.50
7.45
7.40
7.35
7.30
7.25
7.20
7.15
7.10
7.05
7.00
6.95
6.90
6.85
f1 (ppm)
4200
4000
3800
3600
3400
3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
-200


1H NMR 14
11.32
10.36
8.78
8.76
O
O
NH
O
NH
CH3
N
S
O
1.04
1.02
1.02
4200
4000
3800
3600
3400
3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
-200
1.7 11.6 11.5 11.4 11.3 11.2 11.1 11.0 10.9 10.8 10.7 10.6 10.5 10.4 10.3 10.2 10.1 10.0 9.9 9.8 9.7 9.6 9.5 9.4 9.3 9.2 9.1 9.0 8.9 8.8 8.7 8.6
f1 (ppm)

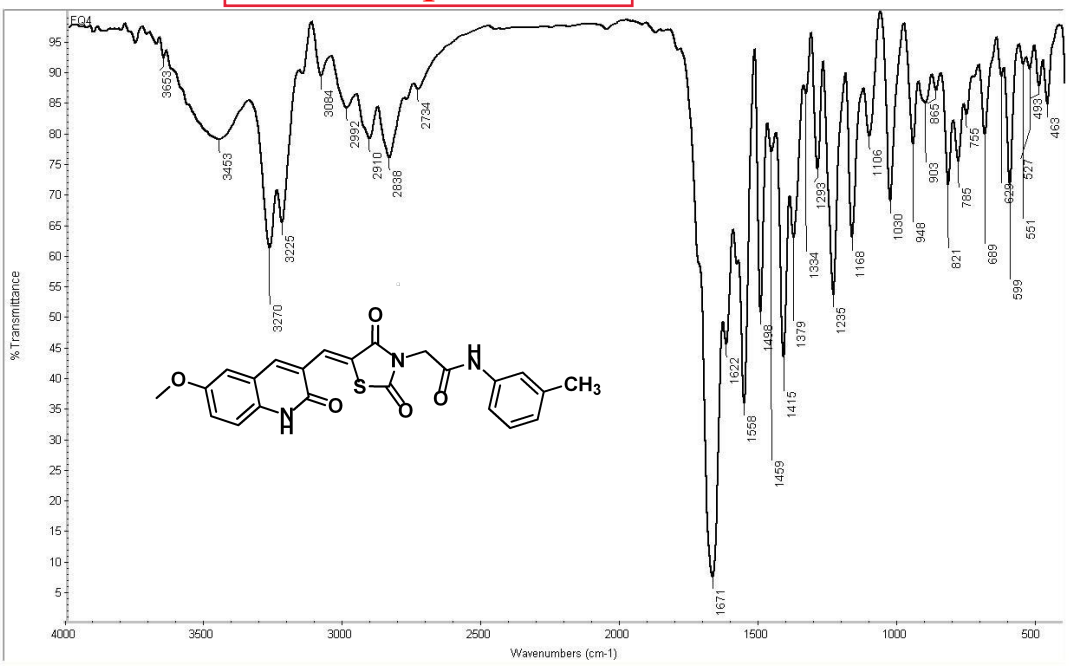

13C NMR 14
170.25
168.71
165.69
164.15
144.62
138.76
138.60
133.55
129.38
129.20
128.42
128.18
124.92
122.65
120.24
120.22
116.88
111.15
44.13
21.62
O
O
O
NH
N
S
O
NH
CH3
f1 (ppm)


13C NMR 14
44.13
21.62
O
O
O
NH
CH3
N
NH
S
O
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
-500
65
60
55
50
45
40
35
30
25
20
15
10
f1 (ppm)

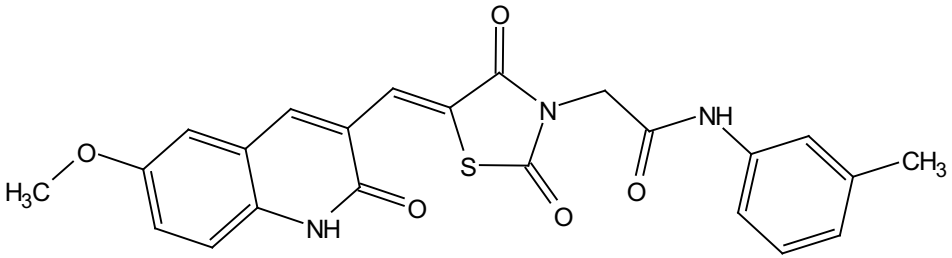

13C NMR 14
144.62
138.76
138.60
133.55
129.38
129.20
128.42
128.18
124.92
122.65
120.24
120.22
116.88
111.15
O
O
O
NH
NH
CH3
N
S
O
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
-500
154 152 150 148 146 144 142 140 138 136 134 132 130 128 126 124 122 120 118 116 114 112 110 108 106 104 102
f1 (ppm)


13C NMR 14
170.25
168.71
165.69
164.15
144.62
138.76
138.60
133.55
O
O
O
NH
N
NH
S
O
CH3
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
-500
175
170
165
160
155
150
145
140
135
f1 (ppm)

IR of compound 9

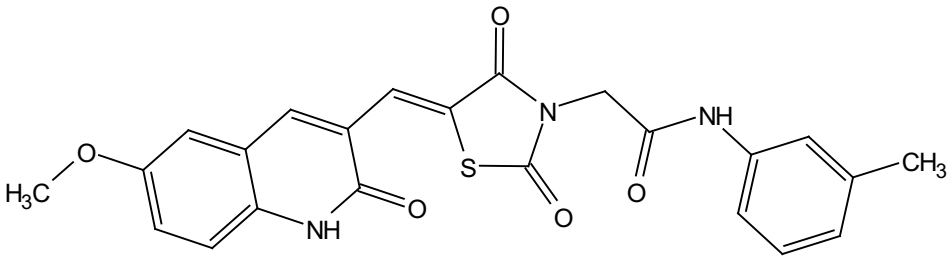


1H NMR of 8


O
N
S
NH
O
O
O
NH
O
H_3C
CH_3

10.40
10.25
8.45
8.20
8.07
8.02
7.98
7.94
7.92
7.47
7.43
7.40
7.40
7.36
7.34
7.32
7.29
4.51
3.82
2.29
1.00
1.38
1.16
1.04
2.15
2.11
3.61
2.18
3.16
3.17
4000
3500
3000
2500
2000
1500
1000
500
0
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 (ppm)

13C NMR of 8



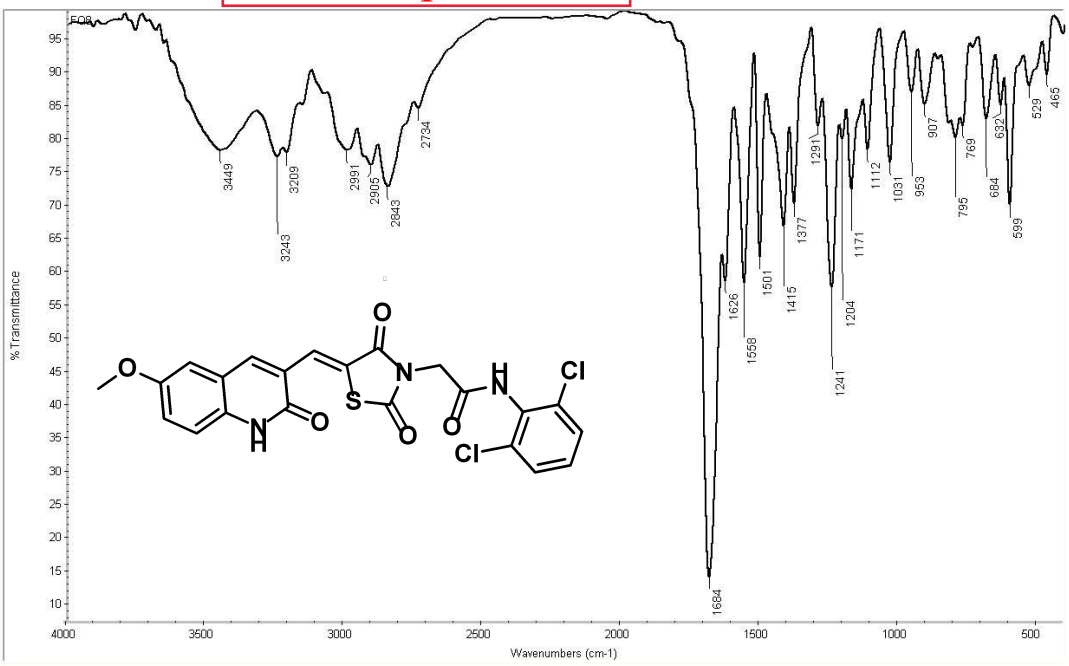
171.60
161.56
157.02
155.00
142.27
138.62
136.50
132.64
129.00
128.83
127.68
126.20
124.34
124.18
119.18
117.30
111.58
56.04
46.47
40.58 DMSO
40.38 DMSO
40.17 DMSO
39.96 DMSO
39.75 DMSO
39.54 DMSO
39.33 DMSO
21.30

5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0

180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0

f1 (ppm)

IR of compound 9



6- Molecular docking of compound 9, 12, 13, and 14

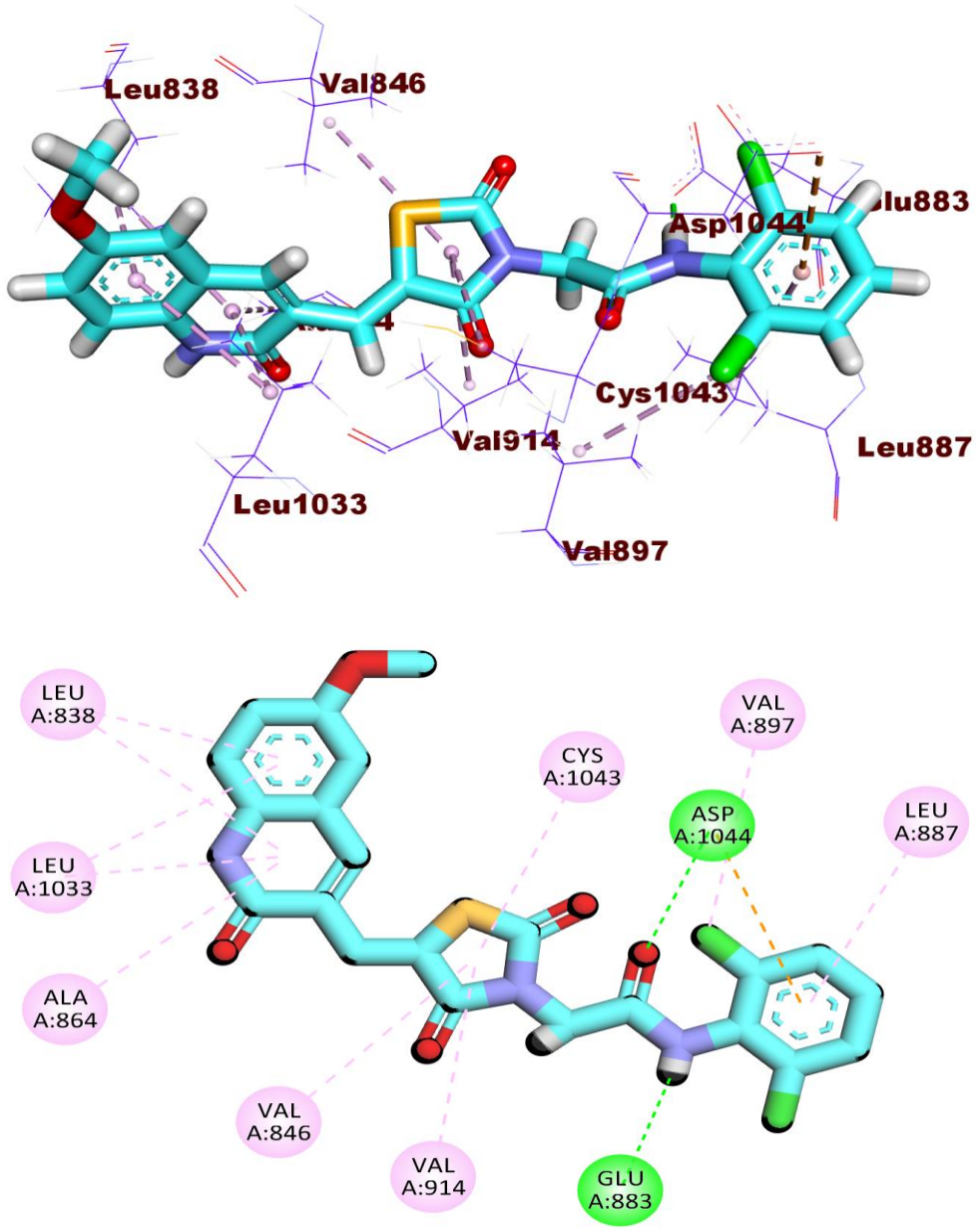


Fig. 1: Binding mode of compound **9** into VEGFR-2 active site

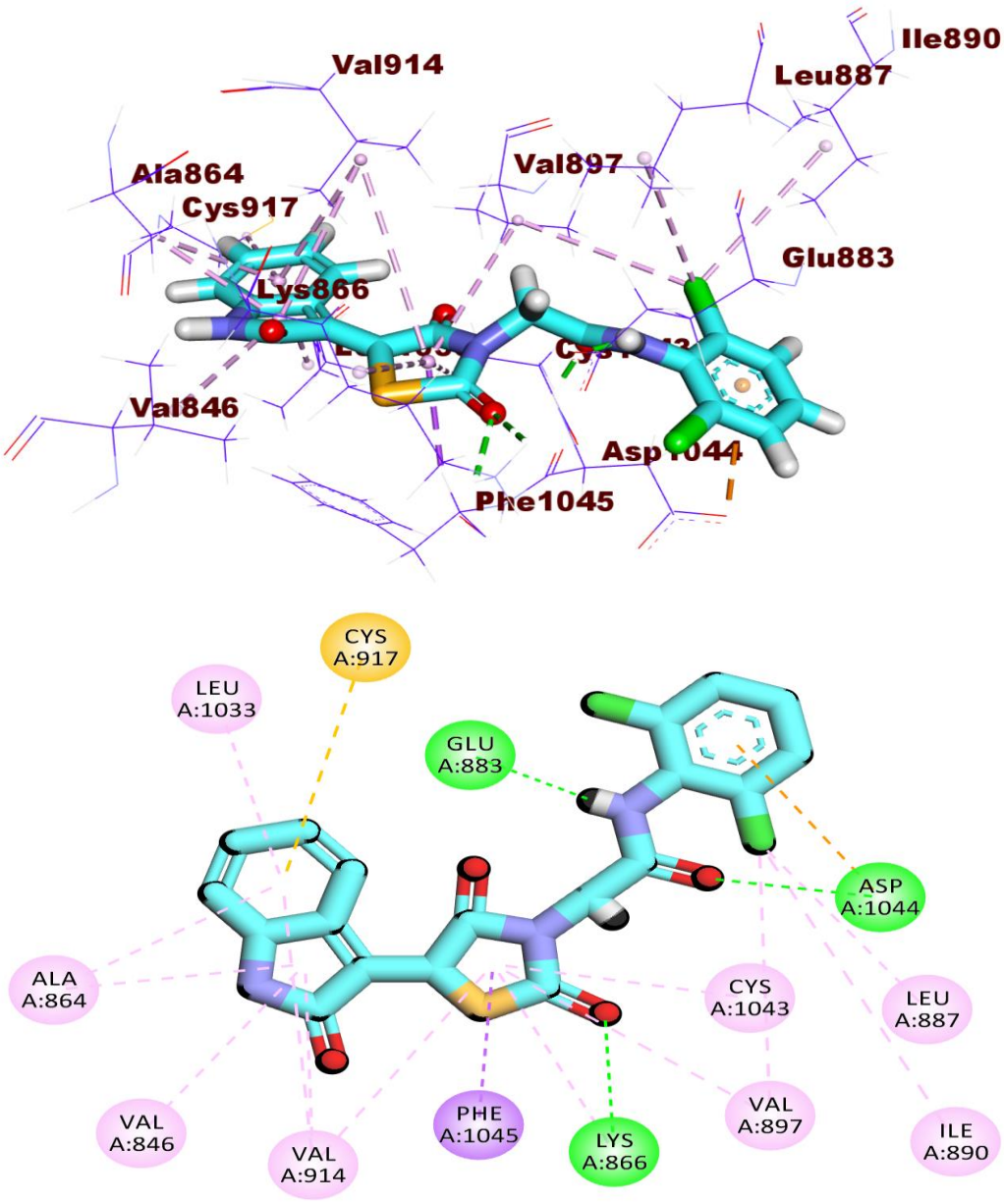
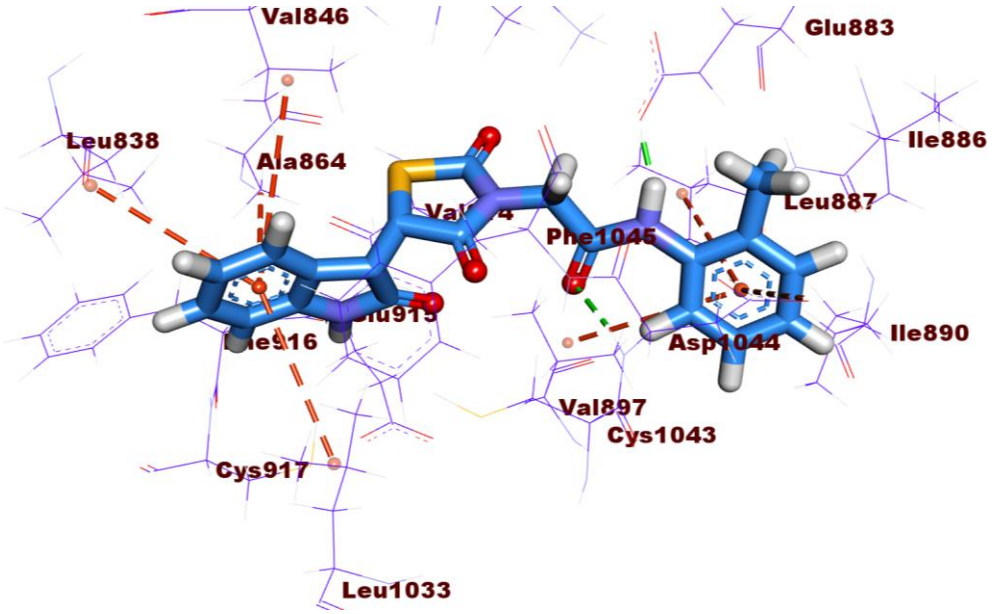
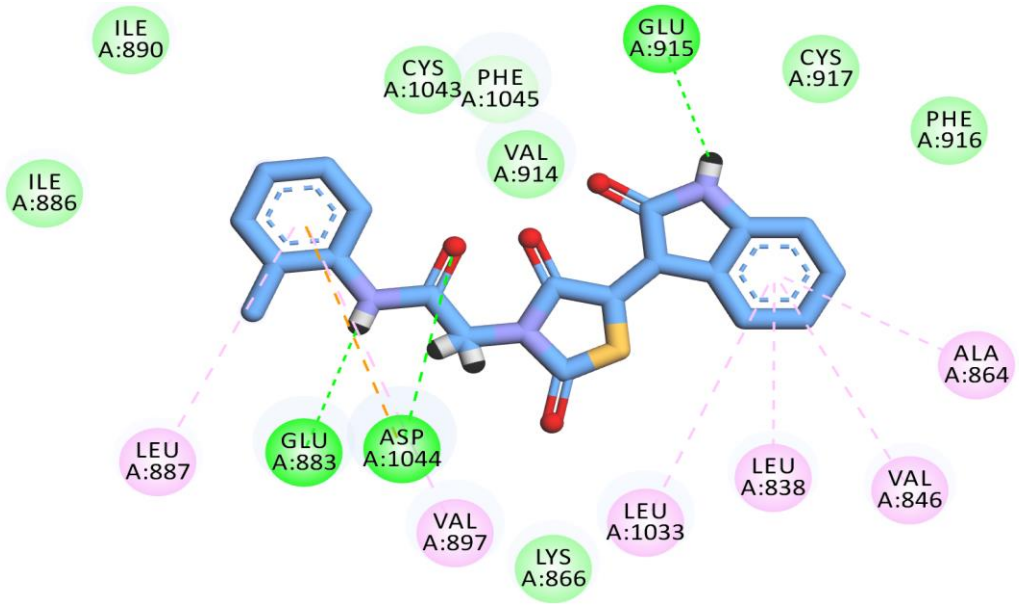


Fig. 2: Binding mode of compound **12** into VEGFR-2 active site


Val846
Glu883
Leu838
Ala864
Ile886
Leu887
Phe1045
Ile890
Asp1044
Val897
Cys1043
Cys917
Leu1033


ILE A:890
GLU A:915
CYS A:1043
PHE A:1045
CYS A:917
PHE A:916
VAL A:914
ILE A:886
ALA A:864
LEU A:887
GLU A:883
ASP A:1044
LEU A:838
VAL A:846
LEU A:1033
VAL A:897
LYS A:866

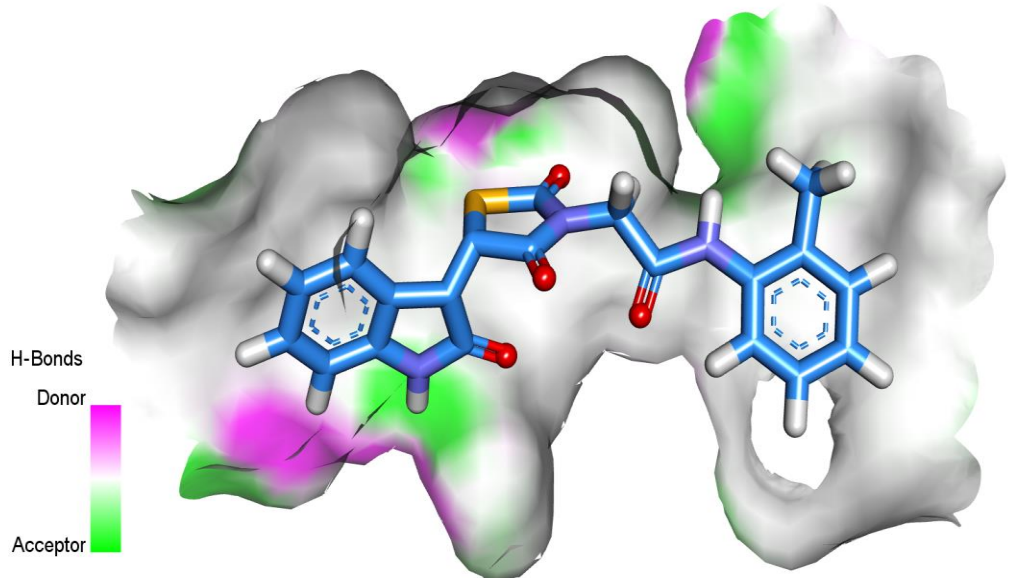
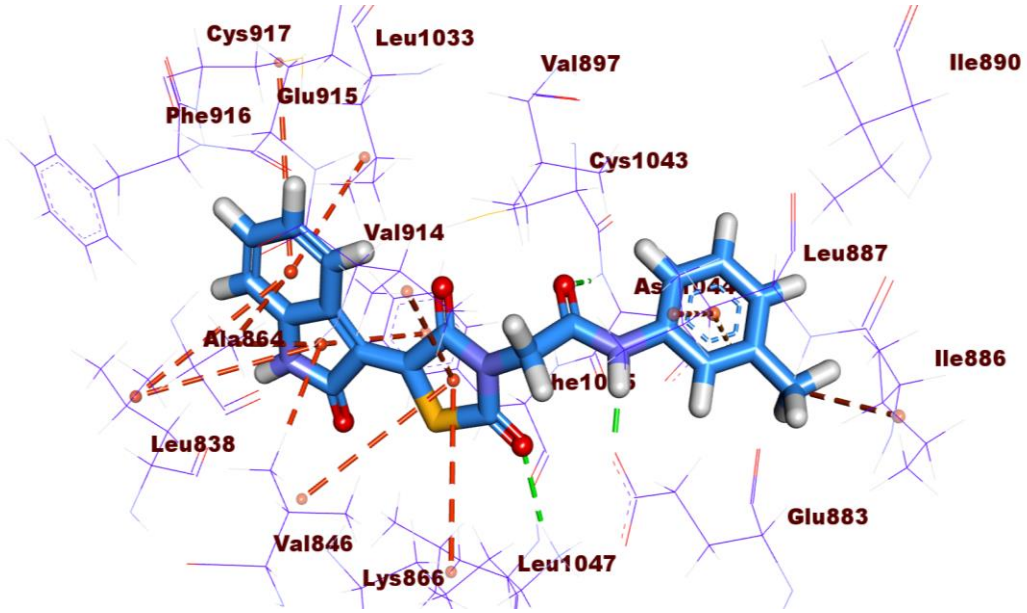


Fig. 3: Binding mode of compound **13** into VEGFR-2 active site



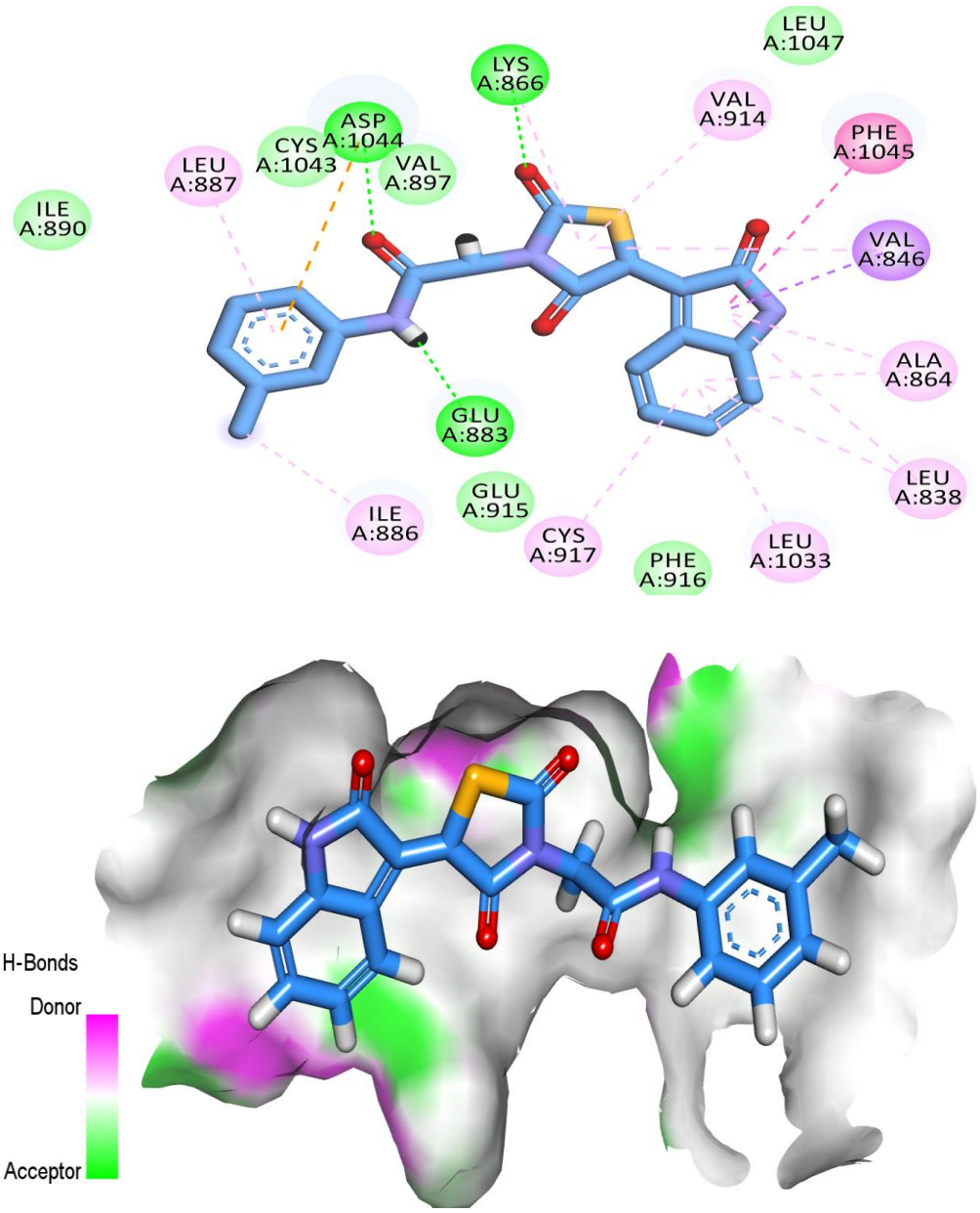
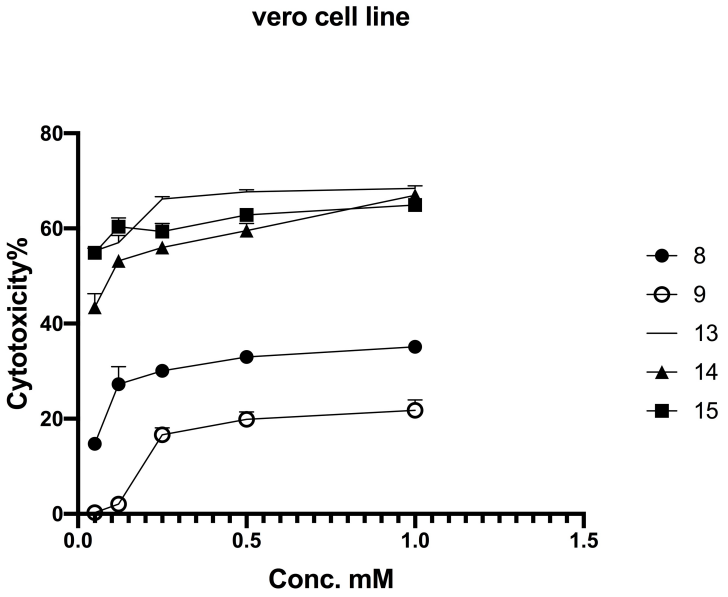
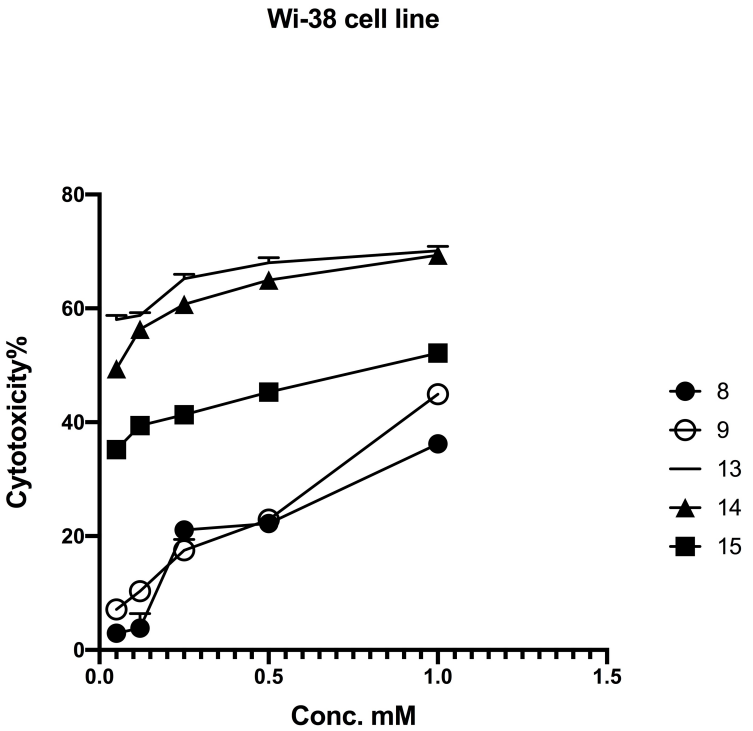


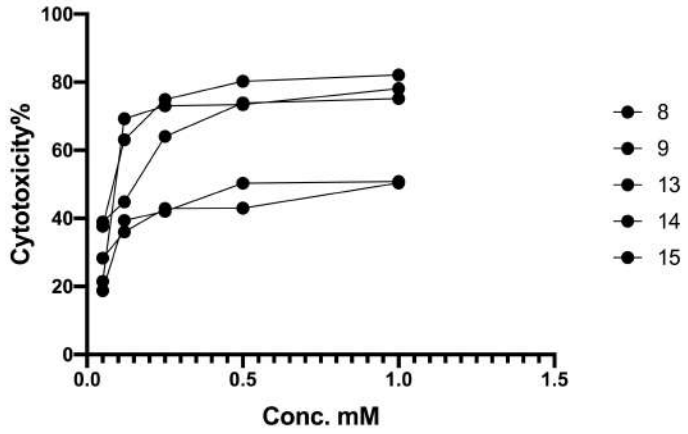
Fig. 4: Binding mode of compound **14** into VEGFR-2 active site

RAW DATA OF CYTOTOXICITY

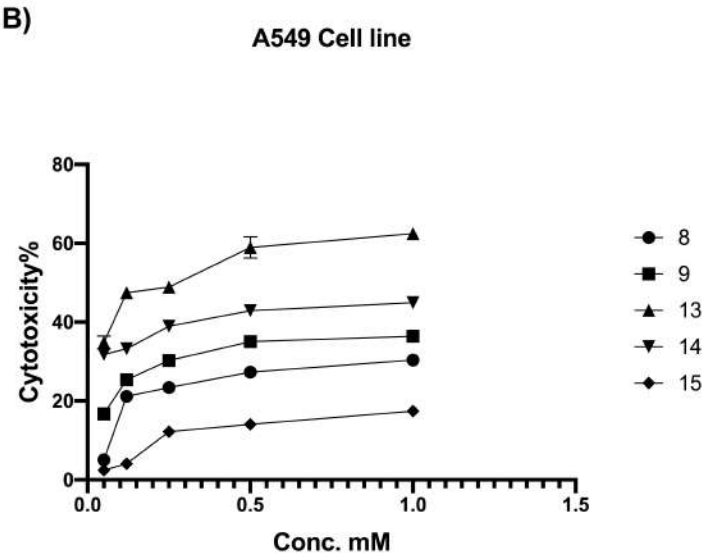

vero cell line
Cytotoxicity%
Conc. mM
0
20
40
60
80
0.0
0.5
1.0
1.5
8
9
13
14
15


Wi-38 cell line
Cytotoxicity%
Conc. mM
0
20
40
60
80
0.0
0.5
1.0
1.5
8
9
13
14
15

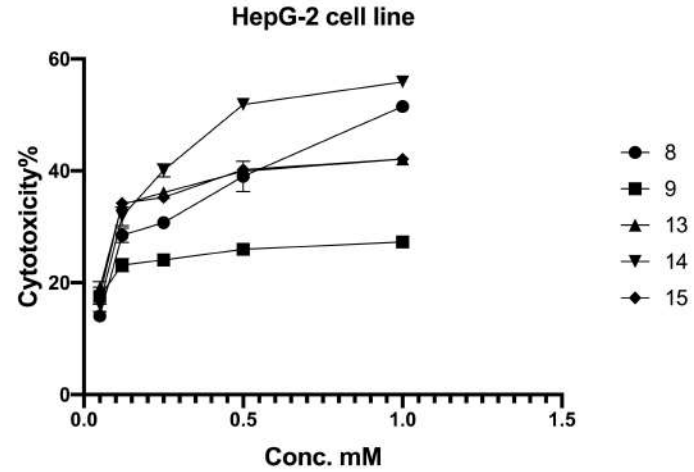
A)


CaCo-2 cell line
Cytotoxicity%
Conc. mM
8
9
13
14
15

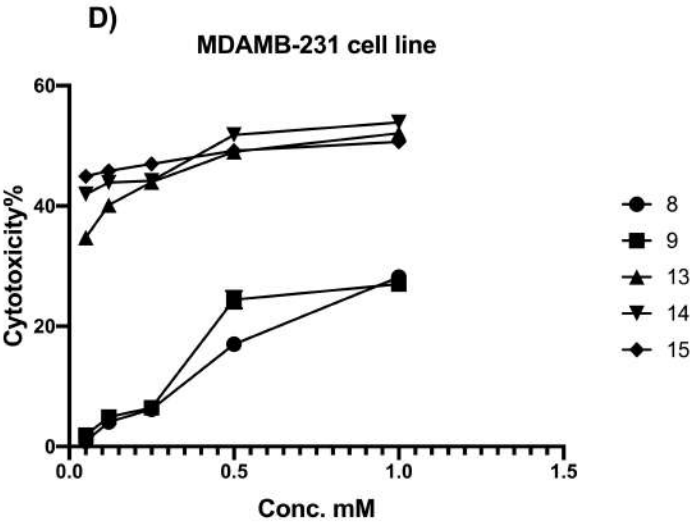
B)


A549 Cell line
Cytotoxicity%
Conc. mM
8
9
13
14
15

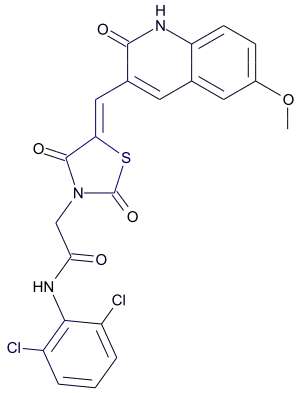
C)


HepG-2 cell line
Cytotoxicity%
Conc. mM
8
9
13
14
15

D)


MDAMB-231 cell line
Cytotoxicity%
Conc. mM
8
9
13
14
15

Compound 8



$C_{22}H_{15}Cl_2N_3O_5S$
Molecular Weight: 504.34259
ALogP: 3.479
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

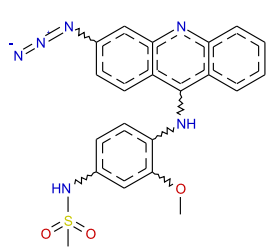
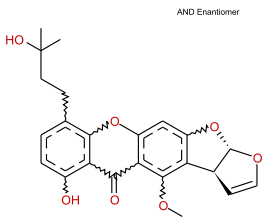
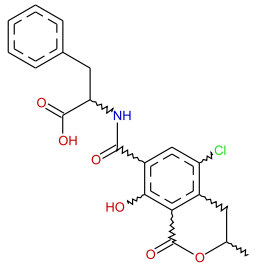
Model Prediction

Prediction: Non-Mutagen
Probability: 0.524
Enrichment: 0.939
Bayesian Score: -6.95
Mahalanobis Distance: 12.3
Mahalanobis Distance p-value: 0.000231

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

TOPKAT_Ames_Mutagenicity

Structural Similar Compounds

Name	80266-02-0	55256-55-8	Ochratoxin A
Structure			
Actual Endpoint	Mutagen	Mutagen	Non-Mutagen
Predicted Endpoint	Mutagen	Mutagen	Non-Mutagen
Distance	0.633	0.633	0.639
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	EMIC

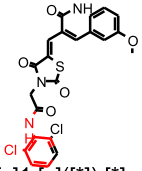
Model Applicability

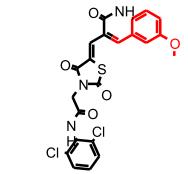
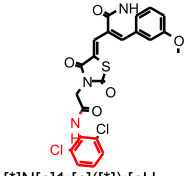
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

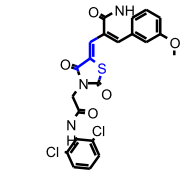
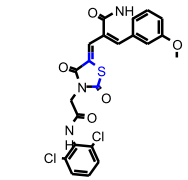
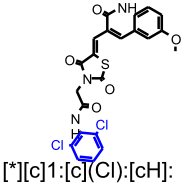
Feature Contribution

Top features for positive contribution

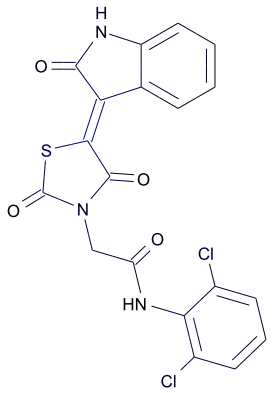
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1562412908	 [*]N[c]1:[c]([*]):[*] :[cH]:[cH]:[c]:1Cl	0.442	5 out of 5

SCFP_12	-577289847	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[c]:1C=[*]	0.399	9 out of 10
SCFP_12	55818998	 [*]N[c]1:[c]([*]):[cH]:[cH]:[cH]:[c]:1Cl	0.241	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-316886873	 [*]\C=C\1/S[*][*]C1=[*]	-0.998	0 out of 3
SCFP_12	-1630519606	 [*]=C1[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	1062412764	 [*][c]1:[c](Cl):[cH]:[cH]:[cH]:[c]:1Cl	-0.675	3 out of 12

Compound 9



$C_{19}H_{11}Cl_2N_3O_4S$
Molecular Weight: 448.27933
ALogP: 3.028
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

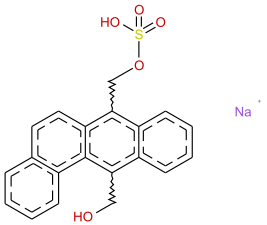
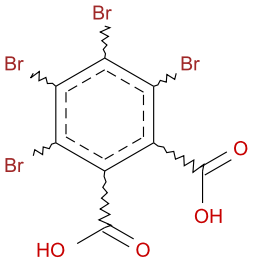
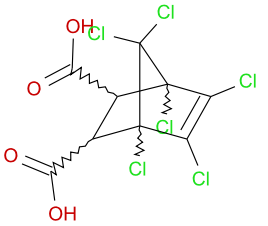
Model Prediction

Prediction: Non-Mutagen
Probability: 0.533
Enrichment: 0.955
Bayesian Score: -6.74
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 0.0049

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

Structural Similar Compounds

Name	98644-23-6	TETRABROMOPHTHALIC ACID	Chlorendic acid
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.614	0.616	0.619
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	EMIC	Environ. Mol. Mut. 16(Suppl 18):1-14;1990

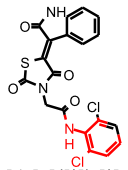
Model Applicability

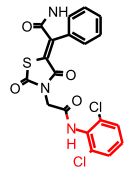
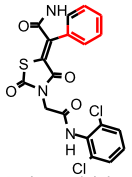
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

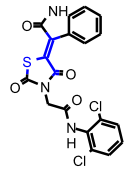
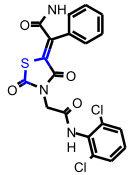
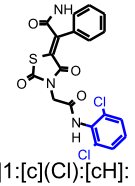
Feature Contribution

Top features for positive contribution

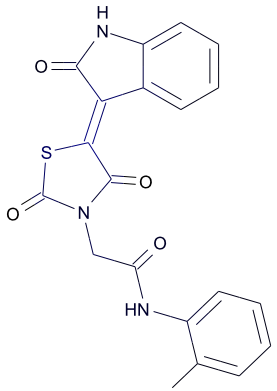
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1562412908	 [*]N[c]1:[c]([*]):[*]:[cH]:[cH]:[c]:1Cl	0.442	5 out of 5

SCFP_12	55818998	 [*]N[c]1:[c]([*]):[cH]:[cH]:[cH]:[c]:1Cl	0.241	1 out of 1
SCFP_12	-1379591900	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	0.108	1480 out of 2326

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-316886873	 [*]\C=C\1/S[*][*]C1=[*]	-0.998	0 out of 3
SCFP_12	-1630519606	 [*]=C1[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	1062412764	 [*][c]1:[c](Cl):[cH]: [cH]:[cH]:[c]:1Cl	-0.675	3 out of 12

Compound 12



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

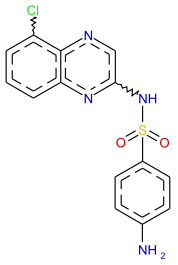
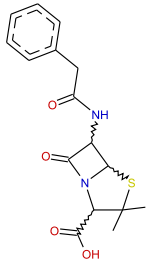
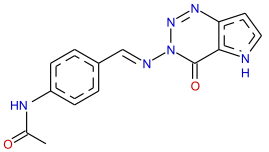
Model Prediction

Prediction: Non-Mutagen
Probability: 0.582
Enrichment: 1.04
Bayesian Score: -5.51
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.0249

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

Structural Similar Compounds

Name	97919-22-7	PENICILLIN G POTASSIUM	135086-96-3
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Distance	0.568	0.590	0.591
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	EMIC	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

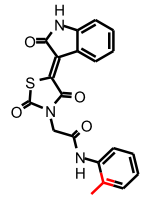
Model Applicability

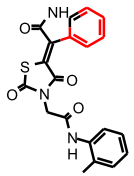
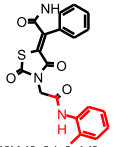
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

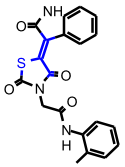
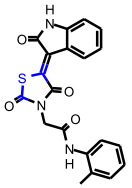
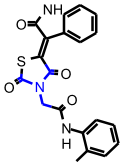
Feature Contribution

Top features for positive contribution

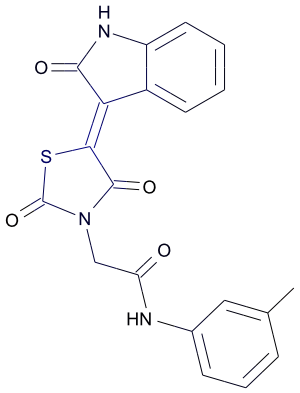
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	136686699	 [*]:[c](:[*])C	0.129	446 out of 686

SCFP_12	-1379591900	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	0.108	1480 out of 2326
SCFP_12	548903629	 [*]C(=[*])N[c]1:[cH]: [cH]:[cH]:[cH]:[c]:1 C	0.1	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-316886873	 [*]\C=C\1/S[*][*]C1=[*]	-0.998	0 out of 3
SCFP_12	-1630519606	 [*]=C1[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	399659969	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.55	21 out of 65

Compound 13



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

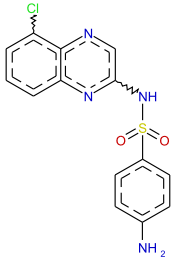
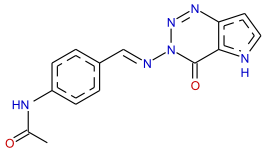
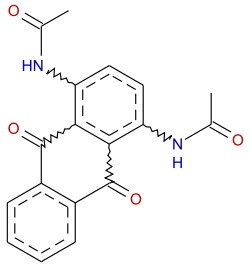
Model Prediction

Prediction: Non-Mutagen
Probability: 0.591
Enrichment: 1.06
Bayesian Score: -5.28
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.0249

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

Structural Similar Compounds

Name	97919-22-7	135086-96-3	1;4-DIACETYLAMINOANTHRAQUINONE
Structure			
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen
Distance	0.568	0.588	0.594
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	EMIC

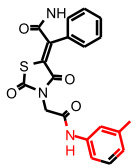
Model Applicability

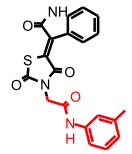
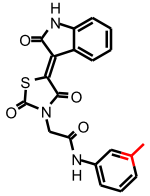
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

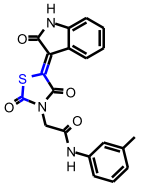
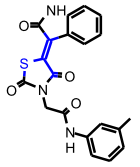
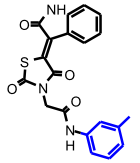
Feature Contribution

Top features for positive contribution

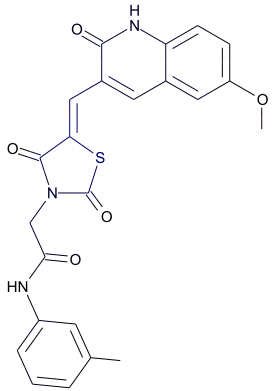
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-347281112	 [*]N[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	0.337	18 out of 22

SCFP_12	-236487363	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.26	4 out of 5
SCFP_12	136686699	 [*]:[c](:[*])C	0.129	446 out of 686

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]=C1[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-316886873	 [*]\C=C\1/S[*][*]C1=[*]	-0.998	0 out of 3
SCFP_12	2052999617	 [*][c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	-0.743	6 out of 24

Compound 14



$C_{23}H_{19}N_3O_5S$
Molecular Weight: 449.47905
ALogP: 2.637
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

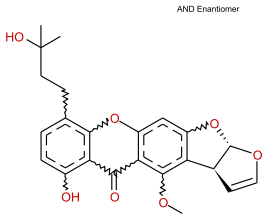
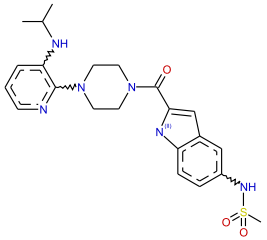
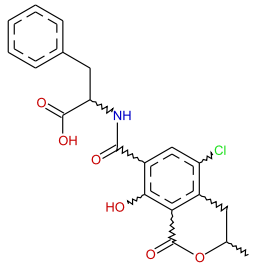
Model Prediction

Prediction: Non-Mutagen
Probability: 0.582
Enrichment: 1.04
Bayesian Score: -5.53
Mahalanobis Distance: 11.8
Mahalanobis Distance p-value: 0.00163

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

Structural Similar Compounds

Name	55256-55-8	Delavirdine	Ochratoxin A
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.594	0.595	0.605
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Contrera, J.F., Matthews, E.J., Kruhlak, N.L., and Benz, R.D., Regulatory Toxicology and Pharmacology 2005, 313-323.	EMIC

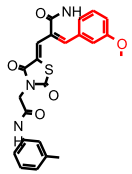
Model Applicability

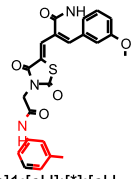
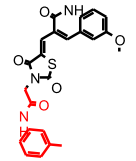
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

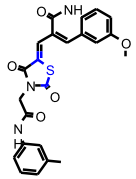
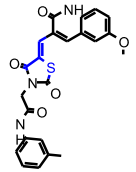
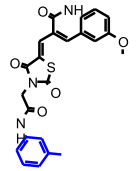
Feature Contribution

Top features for positive contribution

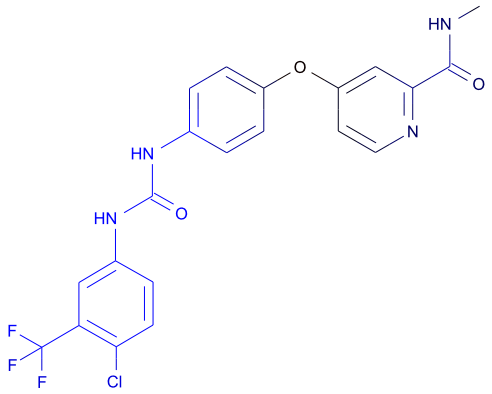
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-577289847	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[c]:1C=[*]	0.399	9 out of 10

SCFP_12	-347281112	 [*]N[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	0.337	18 out of 22
SCFP_12	-236487363	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.26	4 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]=C1[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-316886873	 [*]\C=C\1/S[*][*]C1=[*]	-0.998	0 out of 3
SCFP_12	2052999617	 [*][c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	-0.743	6 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

Model Prediction

Prediction: Non-Mutagen

Probability: 0.0531

Enrichment: 0.0951

Bayesian Score: -19.7

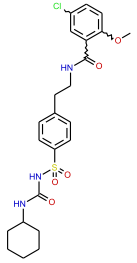
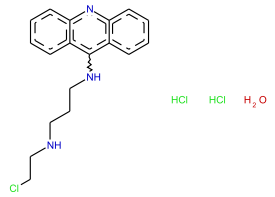
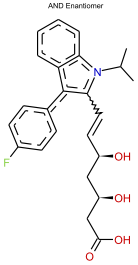
Mahalanobis Distance: 13.1

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.

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

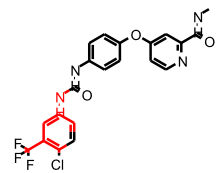
Model Applicability

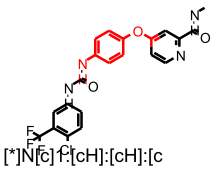
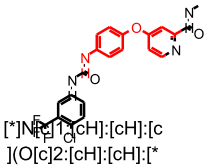
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

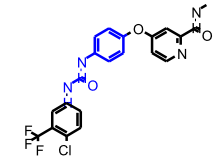
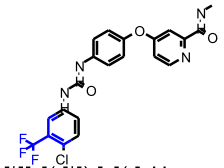
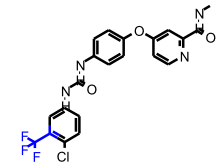
Feature Contribution

Top features for positive contribution

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

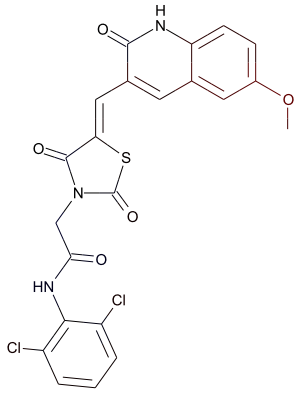
SCFP_12	1208843554		0.337	6 out of 7
SCFP_12	-1943080297		0.304	5 out of 6

Top Features for negative contribution

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

Compound 8

TOPKAT_Developmental_Toxicity_Potential



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Toxic

Probability: 0.644

Enrichment: 1.22

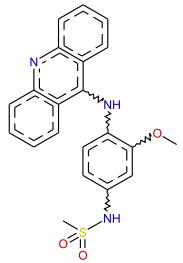
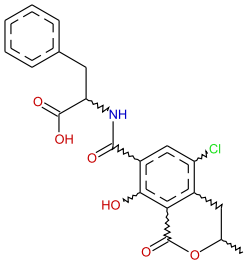
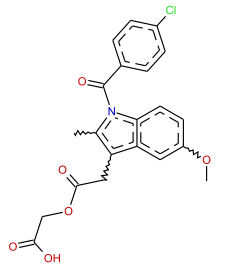
Bayesian Score: 2.47

Mahalanobis Distance: 12.6

Mahalanobis Distance p-value: 2.71e-006

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

Structural Similar Compounds

Name	Amsacrine	Ochratoxin a	Acemetacin
Structure			
Actual Endpoint	Toxic	Toxic	Non-Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.620	0.627	0.651
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	Oyo Yakuri 22(6):777-786; 1981

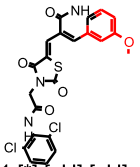
Model Applicability

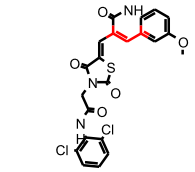
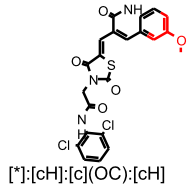
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

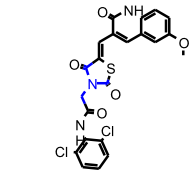
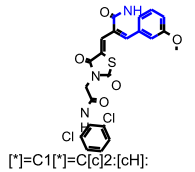
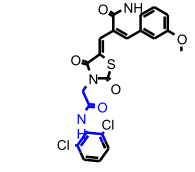
Feature Contribution

Top features for positive contribution

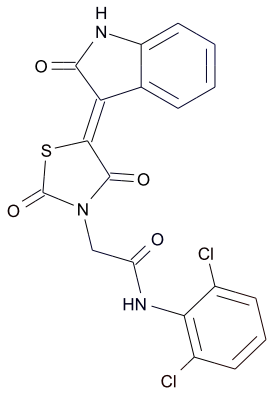
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1237755852	 [*][c]1:[*]:[cH]:[cH]:[c](OC):[cH]:1	0.453	8 out of 9

SCFP_6	-1971137145	 [*]C(=C[c](:[*]):[*]) [*]	0.431	7 out of 8
SCFP_6	591469355	 [*]:[cH]:[c](OC):[cH] :[*]	0.411	10 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.526	3 out of 11
SCFP_6	1420330831	 [*]=C1[*]=C[c]2:[cH]: [*]:[cH]:[cH]:[c]:2N 1	-0.422	0 out of 1
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	-0.422	0 out of 1

Compound 9



$C_{19}H_{11}Cl_2N_3O_4S$
Molecular Weight: 448.27933
ALogP: 3.028
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

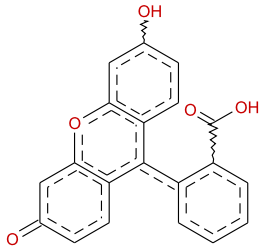
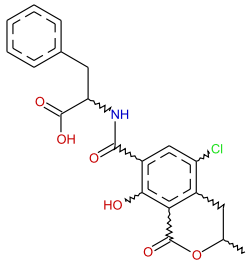
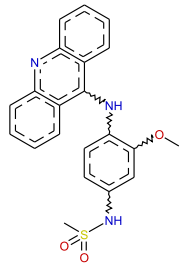
Model Prediction

Prediction: Non-Toxic
Probability: 0.503
Enrichment: 0.956
Bayesian Score: -1.22
Mahalanobis Distance: 8.41
Mahalanobis Distance p-value: 0.407

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

Structural Similar Compounds

Name	D&C Yellow 8	Ochratoxin a	Amsacrine
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.626	0.643	0.648
Reference	Food Chem Toxicol 24:819-823; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 7(2):214-20; 1986

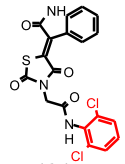
Model Applicability

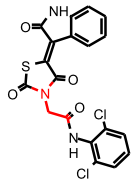
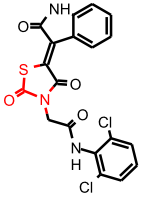
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

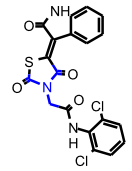
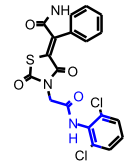
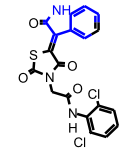
Feature Contribution

Top features for positive contribution

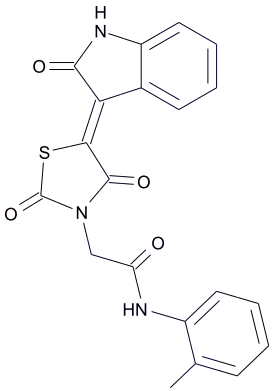
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1062412764	 [*][c]1:[c](Cl):[cH]:[cH]:[cH]:[c]:1Cl	0.381	2 out of 2

SCFP_6	-587539325	 [*]N([*])CC(=[*])[*]	0.271	1 out of 1
SCFP_6	199205675	 [*]N1[*][*]SC1=O	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1=[*]	-0.526	3 out of 11
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	-0.422	0 out of 1
SCFP_6	1420330831	 [*]=C1[*]=C[c]2:[cH]:[*]:[cH]:[cH]:[c]:2N1	-0.422	0 out of 1

Compound 12



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

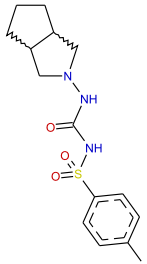
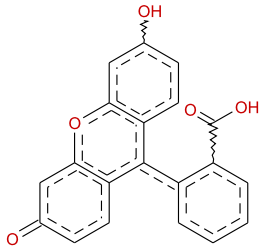
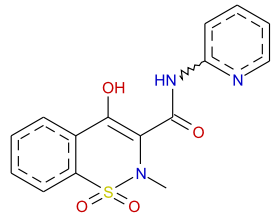
Model Prediction

Prediction: Non-Toxic
Probability: 0.49
Enrichment: 0.931
Bayesian Score: -1.58
Mahalanobis Distance: 9.59
Mahalanobis Distance p-value: 0.0562

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

Structural Similar Compounds

Name	Sulfonylurea Gliclazide	D&C Yellow 8	Piroxicam
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.579	0.602	0.623
Reference	Yakuri to Chiryo 9:3551-3571; 1981	Food Chem Toxicol 24:819-823; 1986	Yakuri to Chiryo 8:4655-4671; 1980

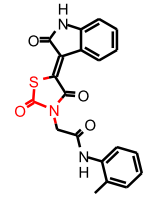
Model Applicability

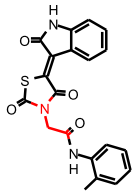
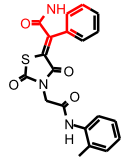
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

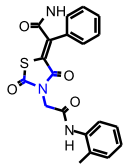
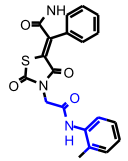
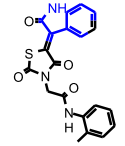
Feature Contribution

Top features for positive contribution

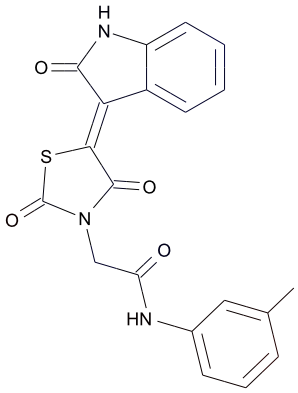
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	199205675	 [*]N1[*][*]SC1=O	0.271	1 out of 1

SCFP_6	-587539325	 [*]N([*])CC(=[*])[*]	0.271	1 out of 1
SCFP_6	2102703671	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.526	3 out of 11
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	-0.422	0 out of 1
SCFP_6	1420330831	 [*]=C1[*]=C[c]2:[cH]: [*]:[cH]:[cH]:[c]:2N 1	-0.422	0 out of 1

Compound 13



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

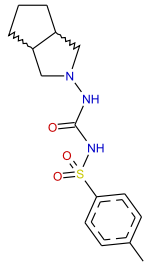
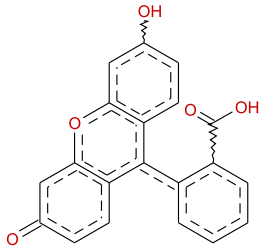
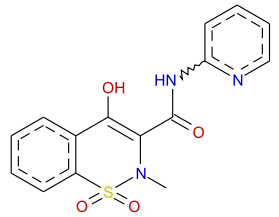
Model Prediction

Prediction: Non-Toxic
Probability: 0.517
Enrichment: 0.982
Bayesian Score: -0.832
Mahalanobis Distance: 8.48
Mahalanobis Distance p-value: 0.375

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

TOPKAT_Developmental_Toxicity_Potential

Structural Similar Compounds

Name	Sulfonylurea Gliclazide	D&C Yellow 8	Piroxicam
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.575	0.603	0.624
Reference	Yakuri to Chiryo 9:3551-3571; 1981	Food Chem Toxicol 24:819-823; 1986	Yakuri to Chiryo 8:4655-4671; 1980

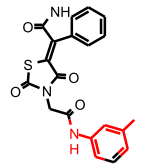
Model Applicability

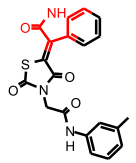
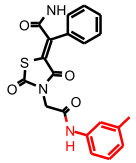
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: 4.1216. Training min, max, SD, explained variance: -3.5766, 3.955, 1.214, 0.0216.

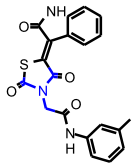
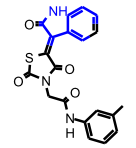
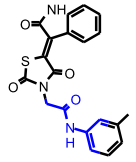
Feature Contribution

Top features for positive contribution

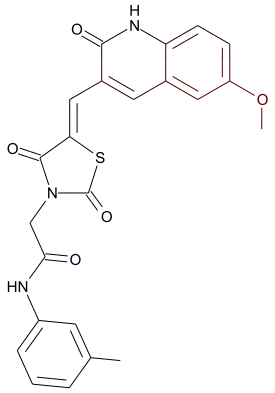
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-347281112	 [*]N[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	0.381	2 out of 2

SCFP_6	2102703671	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.271	1 out of 1
SCFP_6	795669118	 [*]N[c]1:[cH]:[cH]:[c H]:[c](C):[cH]:1	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.526	3 out of 11
SCFP_6	1420330831	 [*]=C1[*]=C[c]2:[cH]: [*]:[cH]:[cH]:[c]:2N 1	-0.422	0 out of 1
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	-0.422	0 out of 1

Compound 14



$C_{23}H_{19}N_3O_5S$
Molecular Weight: 449.47905
ALogP: 2.637
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

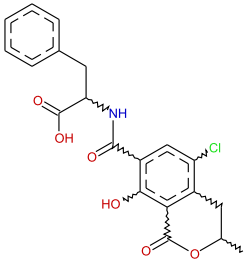
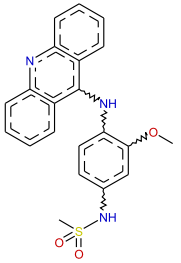
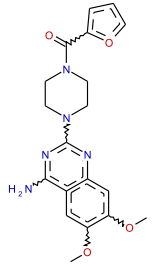
Model Prediction

Prediction: Toxic
Probability: 0.632
Enrichment: 1.2
Bayesian Score: 2.17
Mahalanobis Distance: 12.9
Mahalanobis Distance p-value: 5.95e-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.

TOPKAT_Developmental_Toxicity_Potential

Structural Similar Compounds

Name	Ochratoxin a	Amsacrine	Prazosin .HCl (Free base form)
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.601	0.614	0.643
Reference	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 7(2):214-20; 1986	Oyo Yakuri 17:57-62; 1979

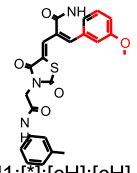
Model Applicability

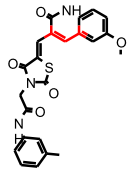
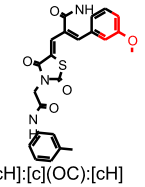
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

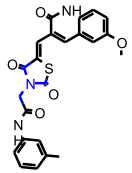
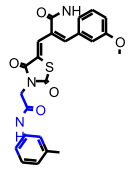
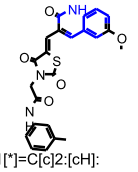
Feature Contribution

Top features for positive contribution

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

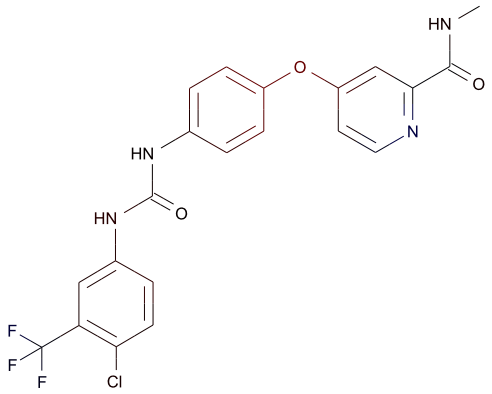
SCFP_6	-1971137145	 [*]C(=C[c](:[*]):[*]) [*]	0.431	7 out of 8
SCFP_6	591469355	 [*]:[cH]:[c](OC):[cH] :[*]	0.411	10 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.526	3 out of 11
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	-0.422	0 out of 1
SCFP_6	1420330831	 [*]=C1[*]=C[c]2:[cH]: [*]:[cH]:[cH]:[c]:2N 1	-0.422	0 out of 1

Sorafenib

TOPKAT_Developmental_Toxicity_Potential



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Toxic

Probability: 0.592

Enrichment: 1.13

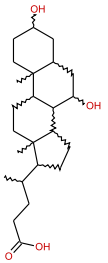
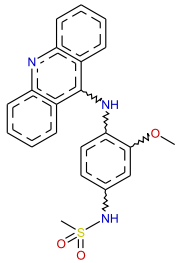
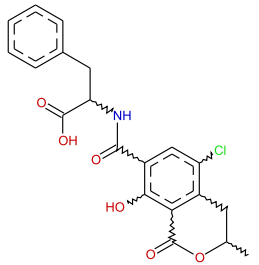
Bayesian Score: 1.15

Mahalanobis Distance: 12.6

Mahalanobis Distance p-value: 2.07e-006

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

Structural Similar Compounds

Name	Chenodiol	Amsacrine	Ochratoxin a
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.631	0.637	0.644
Reference	Arch Int Pharm 246:149-158; 1980	Fundam Appl Toxicol 7(2):214-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976

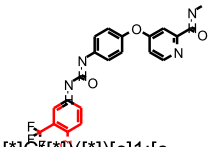
Model Applicability

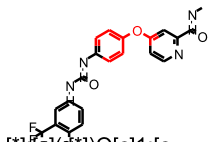
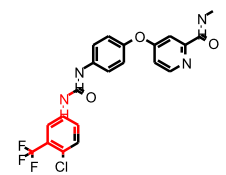
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

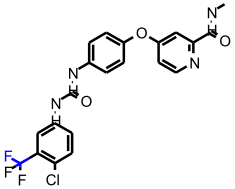
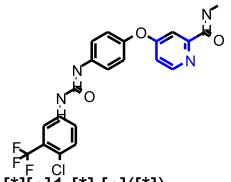
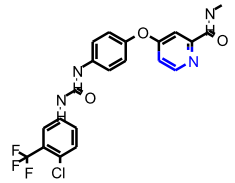
Feature Contribution

Top features for positive contribution

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

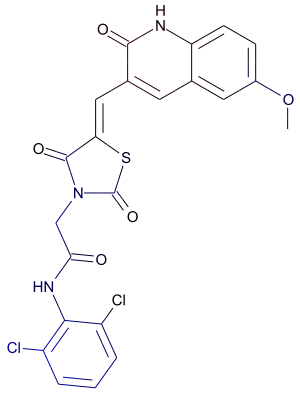
SCFP_6	-488587948	 [*]:[c](:[*])O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.381	2 out of 2
SCFP_6	-347281112	 [*]N[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	0.381	2 out of 2

Top Features for negative contribution

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

Compound 8

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.211

Enrichment: 0.658

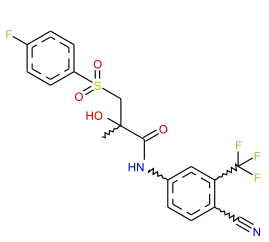
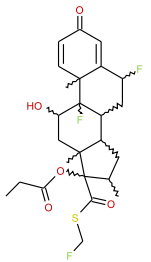
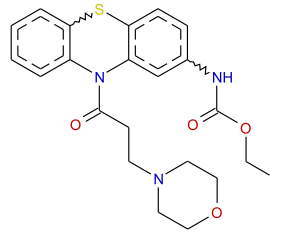
Bayesian Score: -8.43

Mahalanobis Distance: 13.4

Mahalanobis Distance p-value: 0.000146

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Fluticasone	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.633	0.656	0.666
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

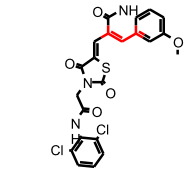
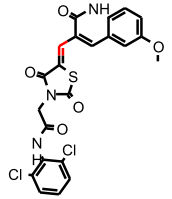
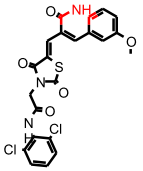
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC29 out of range. Value: -2.9535. Training min, max, SD, explained variance: -2.8294, 4.0152, 1.011, 0.0104.
2. Unknown ECFP_2 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
3. Unknown ECFP_2 feature: 1182722866: [*]C(=CC(=[*])[*])[*]
4. Unknown ECFP_2 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
5. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

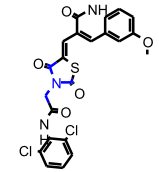
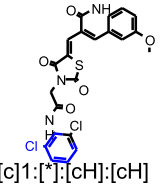
Feature Contribution

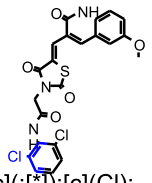
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_6	464808839	 [*]C(=C[c](:[*]):[*]) [*]	0.524	8 out of 14
ECFP_6	-1925046727	 [*]C=[*]	0.391	11 out of 23
ECFP_6	-1699286547	 [*]C(=[*])N[c](:[*]): [*]	0.297	12 out of 28

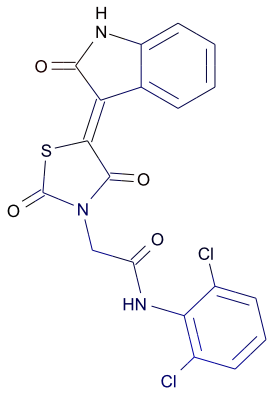
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-661097313	 [*]CN1C(=[*])[*][*]C1 =[*]	-1.55	0 out of 12
ECFP_6	1641317964	 [*][c]1:[*]:[cH]:[cH] :[cH]:[c]:1Cl	-0.789	1 out of 11

ECFP_6	1335691903	 [*][c](:[*]):[c](Cl): [cH]:[*]	-0.669	3 out of 22

Compound 9

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.214

Enrichment: 0.667

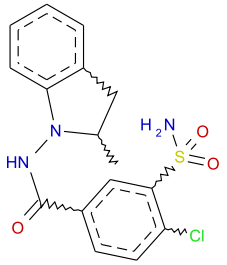
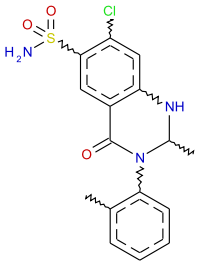
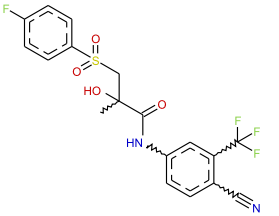
Bayesian Score: -8.92

Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 0.188

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

Structural Similar Compounds

Name	Indapamide	Metolazone	Bicalutamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.616	0.620	0.654
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

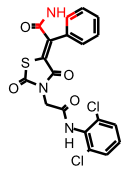
Model Applicability

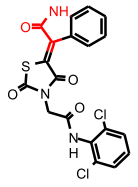
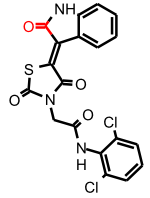
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

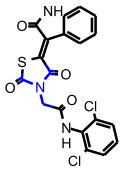
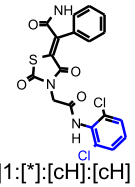
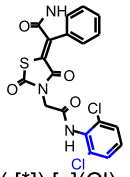
Feature Contribution

Top features for positive contribution

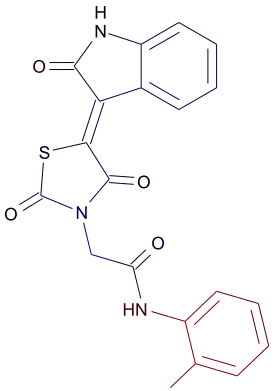
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1699286547	 [*]C(=[*])N[c](:[*]):[*]	0.297	12 out of 28

ECFP_6	1298725959	 [*]NC(=O)C(=[*])[*]	0.279	4 out of 9
ECFP_6	2106656448	 [*]C(=O)[*]	0.254	31 out of 77

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-661097313	 [*]CN1C(=[*])[*][*]C1 =[*]	-1.55	0 out of 12
ECFP_6	1641317964	 [*][c]1:[*]:[cH]:[cH] :[cH]:[c]:1Cl	-0.789	1 out of 11
ECFP_6	1335691903	 [*][c](:[*]):[c](Cl): [cH]:[*]	-0.669	3 out of 22

Compound 12



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

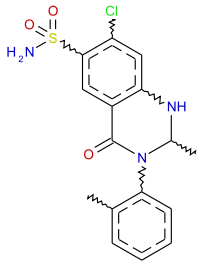
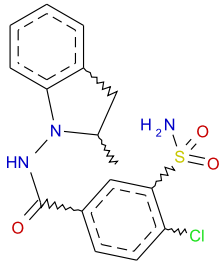
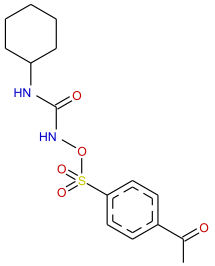
Model Prediction

Prediction: Non-Carcinogen
Probability: 0.223
Enrichment: 0.694
Bayesian Score: -2.72
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 0.308

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

Structural Similar Compounds

Name	Metolazone	Indapamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.551	0.568	0.618
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

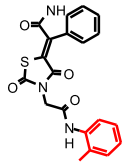
Model Applicability

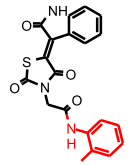
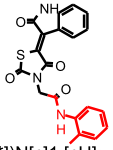
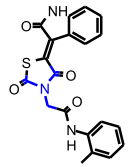
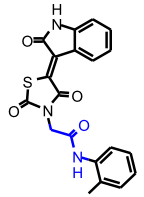
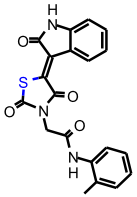
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

Feature Contribution

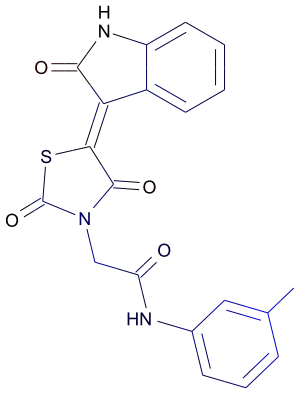
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-907895376	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.451	3 out of 5

ECFP_6	-52177950	 [*]N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.424	1 out of 1
ECFP_6	1360781590	 [*]C(=[*])N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.424	1 out of 1
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_6	-661097313	 [*]CN1C(=[*])[*][*]C1=[*]	-1.55	0 out of 12
ECFP_6	1731843802	 [*]CC(=O)N[*]	-0.657	0 out of 3
ECFP_6	912478223	 [*]S[*]	-0.638	1 out of 9

Compound 13

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{20}H_{15}N_3O_4S$

Molecular Weight: 393.4158

ALogP: 2.185

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.215

Enrichment: 0.672

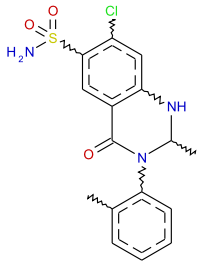
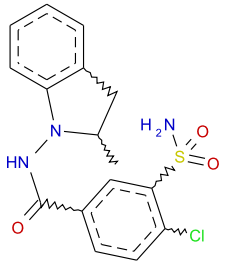
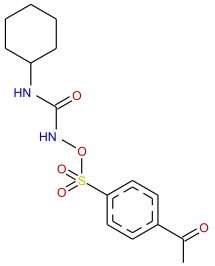
Bayesian Score: -9.11

Mahalanobis Distance: 11

Mahalanobis Distance p-value: 0.108

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

Structural Similar Compounds

Name	Metolazone	Indapamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.562	0.564	0.619
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

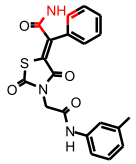
Model Applicability

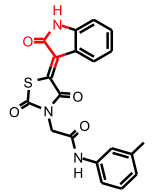
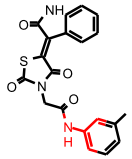
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

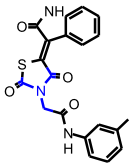
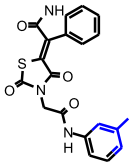
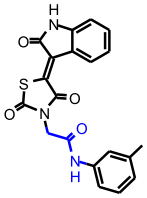
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1699286547	 [*]C(=[*])N[c](:[*]):[*]	0.297	12 out of 28

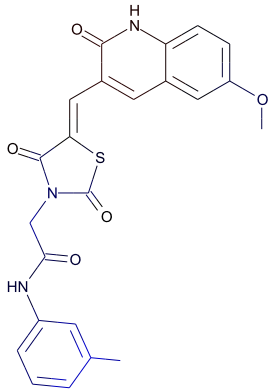
ECFP_6	1298725959	 [*]NC(=O)C(=[*])[*]	0.279	4 out of 9
ECFP_6	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-661097313	 [*]CN1C(=[*])[*][*]C1=[*]	-1.55	0 out of 12
ECFP_6	-179515162	 [*]:[cH]:[c](C):[cH]:[*]	-1.41	0 out of 10
ECFP_6	1731843802	 [*]CC(=O)N[*]	-0.657	0 out of 3

Compound 14

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.21

Enrichment: 0.656

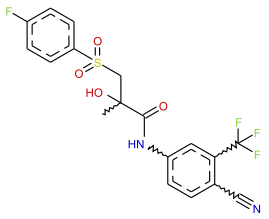
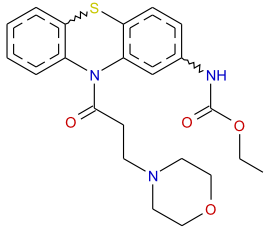
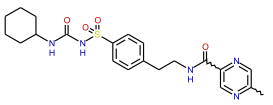
Bayesian Score: -8.36

Mahalanobis Distance: 13.3

Mahalanobis Distance p-value: 0.000213

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Moricizine	Glipizide
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.568	0.628	0.666
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

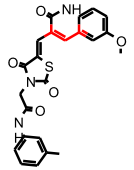
Model Applicability

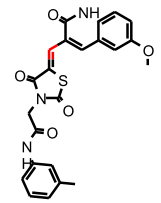
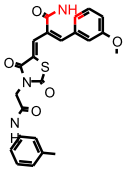
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
3. Unknown ECFP_2 feature: 1182722866: [*]C(=CC(=[*])[*])[*]
4. Unknown ECFP_2 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
5. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

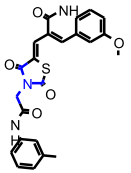
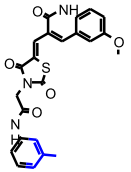
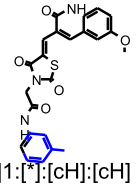
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	464808839	 [*]C(=C[c](:[*]):[*])[*]	0.524	8 out of 14

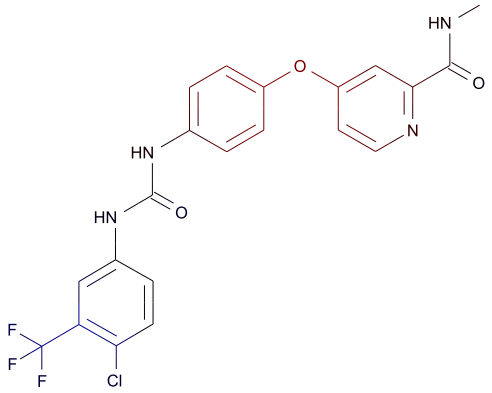
ECFP_6	-1925046727	 [*]C=[*]	0.391	11 out of 23
ECFP_6	-1699286547	 [*]C(=[*])N[c](:[*]): [*]	0.297	12 out of 28

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-661097313	 [*]CN1C(=[*])[*][*]C1 =[*]	-1.55	0 out of 12
ECFP_6	-179515162	 [*]:[cH]:[c](C):[cH]: [*]	-1.41	0 out of 10
ECFP_6	-317125107	 [*][c]1:[*]:[cH]:[cH] :[c](C):[cH]:1	-0.657	0 out of 3

Sorafenib

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.257

Enrichment: 0.801

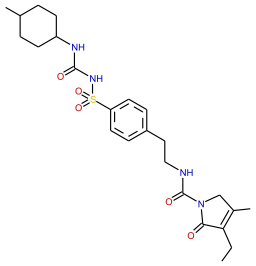
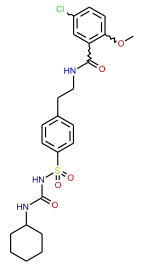
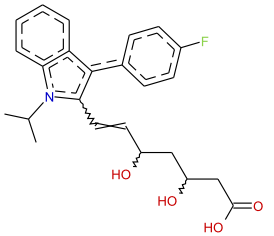
Bayesian Score: -0.321

Mahalanobis Distance: 14.9

Mahalanobis Distance p-value: 4.21e-007

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

Structural Similar Compounds

Name	Glimepride	Glyburide	Fluvastatin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.605	0.615	0.625
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

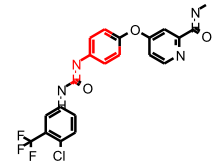
Model Applicability

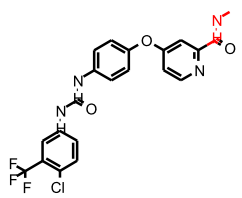
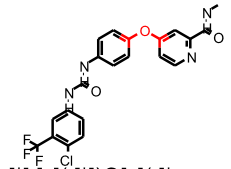
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

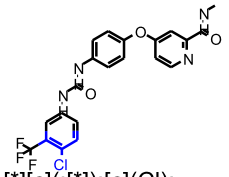
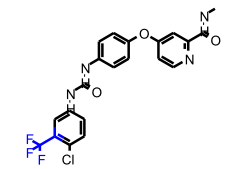
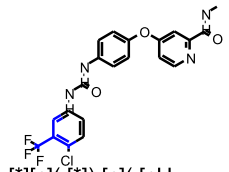
Feature Contribution

Top features for positive contribution

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

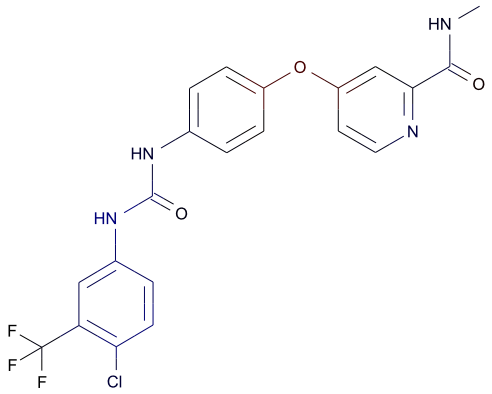
ECFP_6	1338334141	 [*]C(=[*])NC	0.442	2 out of 3
ECFP_6	1305253718	 [*]:[c](:[*])O[c](:[*]):[*]	0.424	1 out of 1

Top Features for negative contribution

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

Sorafenib

TOPKAT_Mouse_Female_FDA_Single_vs_Multiple



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.283

Enrichment: 0.691

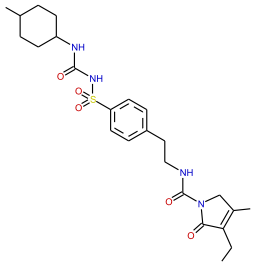
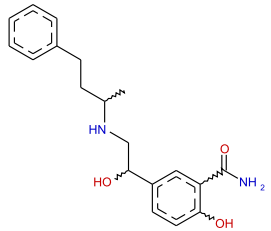
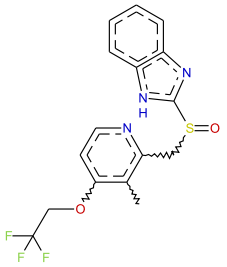
Bayesian Score: -3.89

Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.00221

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

Structural Similar Compounds

Name	Glimepride	Labetalol	Lansoprazole
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.599	0.808	0.820
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

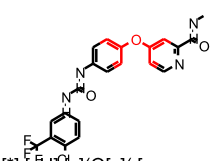
Model Applicability

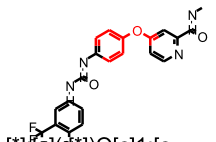
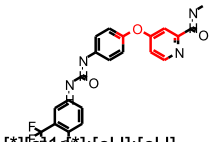
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

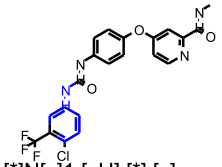
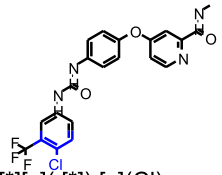
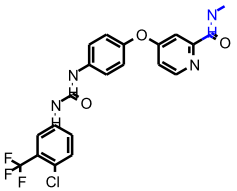
Feature Contribution

Top features for positive contribution

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

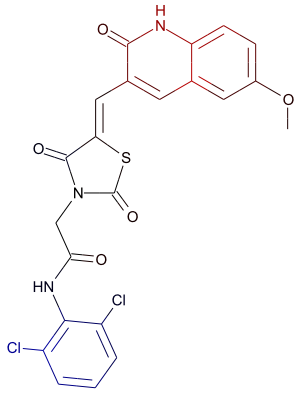
ECFP_4	1407472008	 [*]:[c](:[*])O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.351	1 out of 1
ECFP_4	143734695	 [*][c]1:[*]:[cH]:[cH]:[c](O[c](:[*]):[*]):[cH]:1	0.351	1 out of 1

Top Features for negative contribution

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

Compound 8

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.312

Enrichment: 1.06

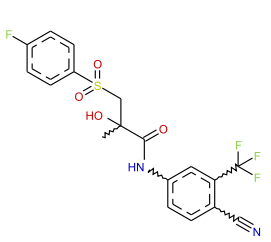
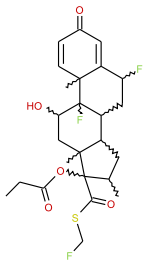
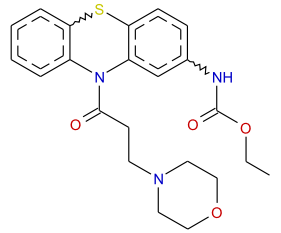
Bayesian Score: 0.208

Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 8.9e-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	Bicalutamide	Fluticasone	Moricizine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.599	0.652	0.657
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

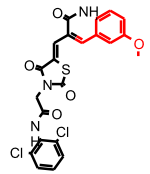
Model Applicability

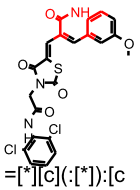
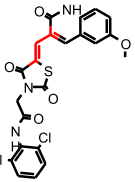
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

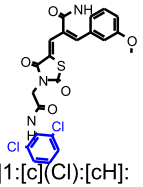
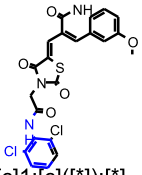
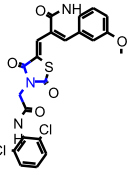
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1757681964	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[c]:1C=[*]	0.676	2 out of 2

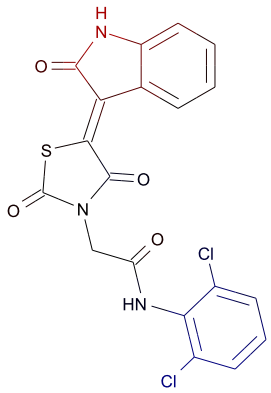
FCFP_6	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.655	7 out of 12
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.479	21 out of 48

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1161767339	 [*][c]1:[c](Cl):[cH]: [cH]:[cH]:[c]:1Cl	-0.719	0 out of 4
FCFP_6	1783756416	 [*]N[c]1:[c]([*]):[*] :[cH]:[cH]:[c]:1Cl	-0.719	0 out of 4
FCFP_6	-1553874037	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.45	5 out of 32

Compound 9

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.286

Enrichment: 0.972

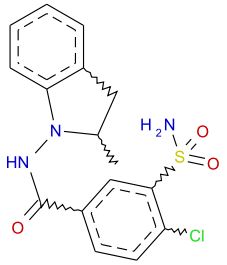
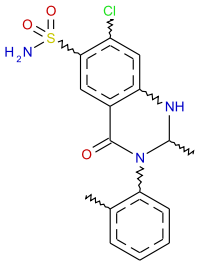
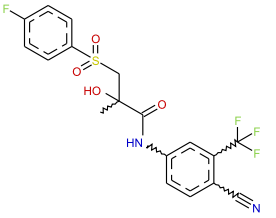
Bayesian Score: -0.754

Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.0239

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

Structural Similar Compounds

Name	Indapamide	Metolazone	Bicalutamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.596	0.599	0.601
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

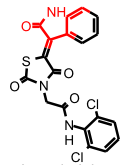
Model Applicability

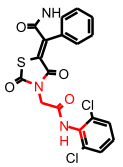
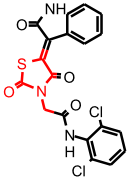
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

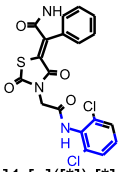
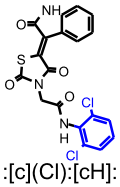
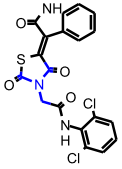
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.655	7 out of 12

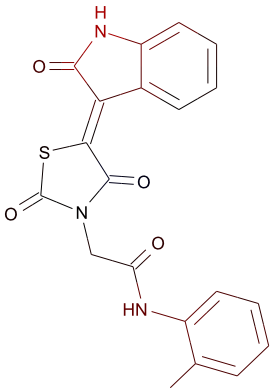
FCFP_6	-1947166985	 [*]N([*])CC(=O)N[c](: [*]):[*]	0.46	1 out of 1
FCFP_6	2036120522	 [*]CN1C(=O)SC(=[*])C1 =[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1783756416	 [*]N[c]1:[c]([*]):[*] :[cH]:[cH]:[c]:1Cl	-0.719	0 out of 4
FCFP_6	1161767339	 [*][c]1:[c](Cl):[cH]: [cH]:[cH]:[c]:1Cl	-0.719	0 out of 4
FCFP_6	-1553874037	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.45	5 out of 32

Compound 12

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



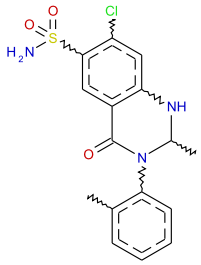
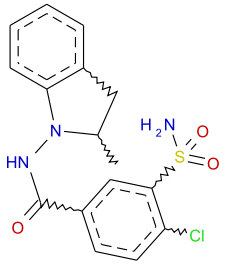
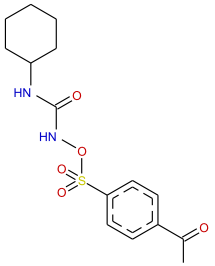
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.442
Enrichment: 1.5
Bayesian Score: 4.16
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.219

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

Structural Similar Compounds

Name	Metolazone	Indapamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.547	0.559	0.606
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

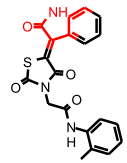
Model Applicability

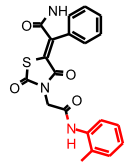
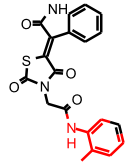
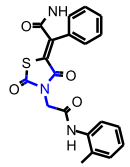
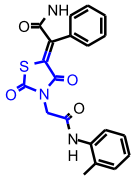
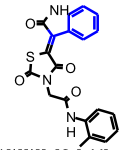
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

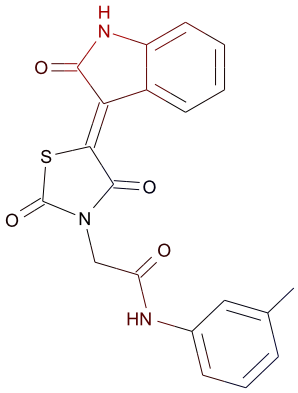
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.655	7 out of 12

FCFP_6	1396506317	 [*]N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.517	2 out of 3
FCFP_6	755520106	 [*]N[c]1:[cH]:[*]:[cH]:[cH]:[c]:1C	0.517	2 out of 3
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
FCFP_6	-1553874037	 [*]CN1C(=[*])[*][*]C1=[*]	-0.45	5 out of 32
FCFP_6	1764344789	 [*]C(=[*])CN1C(=O)SC(=[*])C1=O	-0.233	0 out of 1
FCFP_6	-1698724694	 [*]=C1[*][*][c]2:[cH]:[cH]:[cH]:[cH]:[c]1:2	-0.22	15 out of 72

Compound 13

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



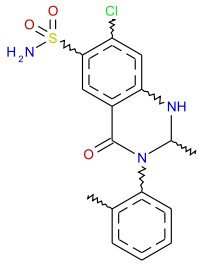
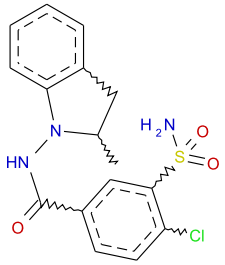
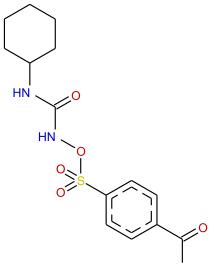
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.378
Enrichment: 1.28
Bayesian Score: 2.38
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.219

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

Structural Similar Compounds

Name	Metolazone	Indapamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.551	0.562	0.609
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

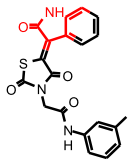
Model Applicability

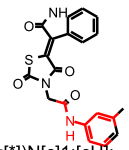
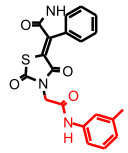
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

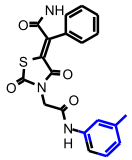
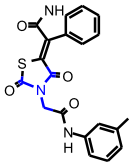
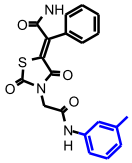
Feature Contribution

Top features for positive contribution

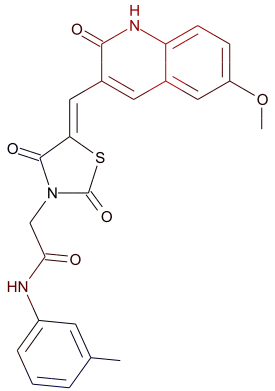
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.655	7 out of 12

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.565	4 out of 7
FCFP_6	-453277354	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.517	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	 [*][c]1:[*]:[cH]:[cH]:[c](C):[cH]:1	-1.29	0 out of 10
FCFP_6	-1553874037	 [*]CN1C(=[*])[*][*]C1=[*]	-0.45	5 out of 32
FCFP_6	630418361	 [*][c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	-0.233	0 out of 1

Compound 14



$C_{23}H_{19}N_3O_5S$
Molecular Weight: 449.47905
ALogP: 2.637
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

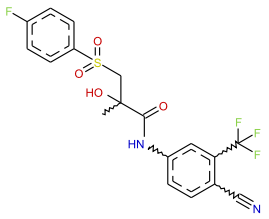
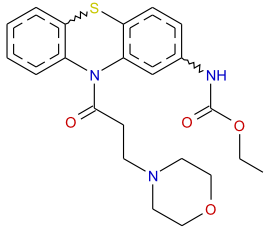
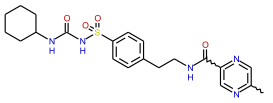
Model Prediction

Prediction: Carcinogen
Probability: 0.393
Enrichment: 1.33
Bayesian Score: 2.81
Mahalanobis Distance: 13.9
Mahalanobis Distance p-value: 1.01e-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.

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen

Structural Similar Compounds

Name	Bicalutamide	Moricizine	Glipizide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.541	0.621	0.662
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

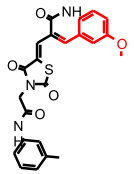
Model Applicability

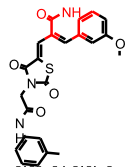
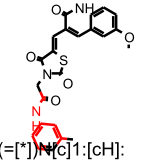
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

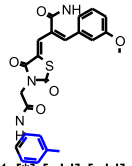
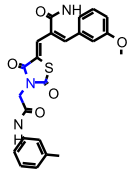
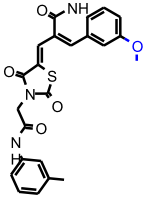
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1757681964	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[c]:1C=[*]	0.676	2 out of 2

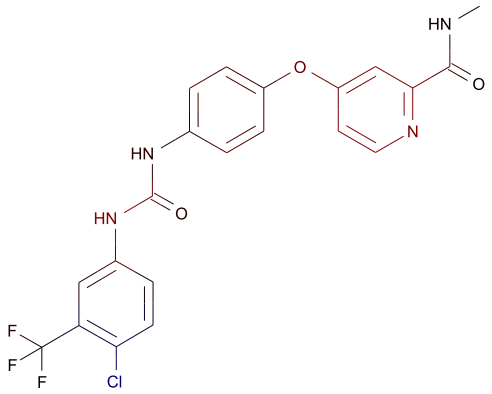
FCFP_6	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.655	7 out of 12
FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[c]([*]):[c H]:1	0.565	4 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	 [*][c]1:[*]:[cH]:[cH] :[c](C):[cH]:1	-1.29	0 out of 10
FCFP_6	-1553874037	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.45	5 out of 32
FCFP_6	136627117	 [*]OC	-0.252	10 out of 50

Sorafenib

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.444

Enrichment: 1.51

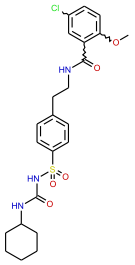
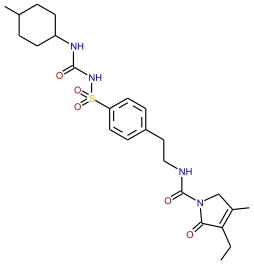
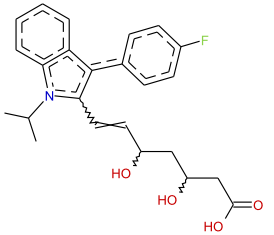
Bayesian Score: 4.21

Mahalanobis Distance: 20.3

Mahalanobis Distance p-value: 1.28e-019

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

Structural Similar Compounds

Name	Glyburide	Glimepride	Fluvastatin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.594	0.599	0.603
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

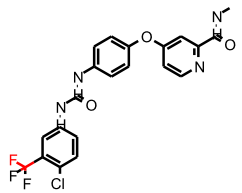
Model Applicability

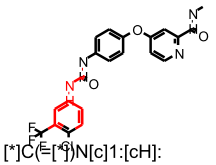
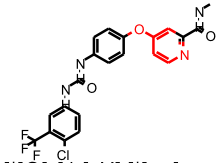
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

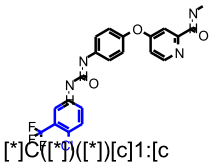
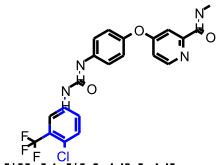
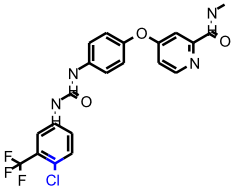
Feature Contribution

Top features for positive contribution

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

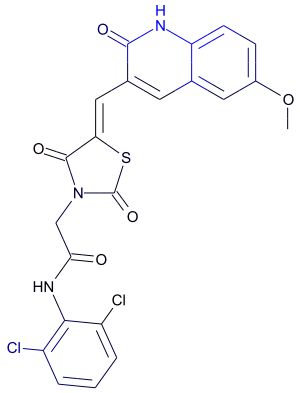
FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.565	4 out of 7
FCFP_6	735850272	 [*]O[c]1:[cH]:[*]:n:[cH]:[cH]:1	0.46	1 out of 1

Top Features for negative contribution

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

Compound 8

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.145

Enrichment: 0.48

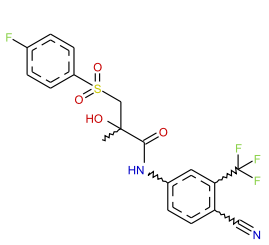
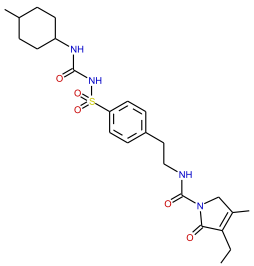
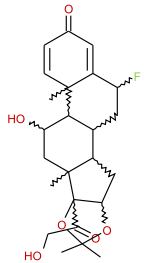
Bayesian Score: -14.1

Mahalanobis Distance: 15.7

Mahalanobis Distance p-value: 7.58e-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	Bicalutamide	Glimepride	Flunisolide
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.641	0.728	0.812
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

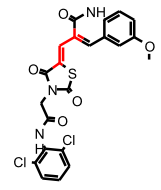
Model Applicability

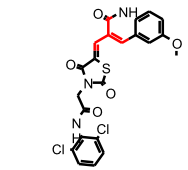
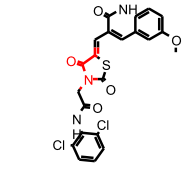
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 4.8042. Training min, max, SD, explained variance: -3.5268, 3.8048, 1.733, 0.0560.

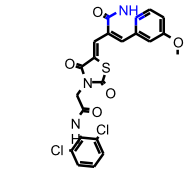
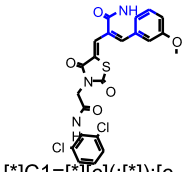
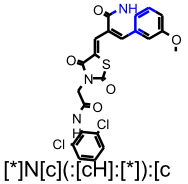
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	451847724	 [*]C(=CC(=[*])[*])[*]	0.3	10 out of 21

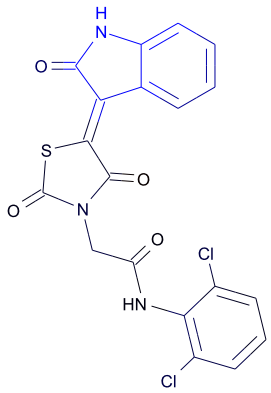
FCFP_12	436886043	 [*]\C=C(\C=[*])/C(=[*])[*]	0.27	7 out of 15
FCFP_12	565998553	 [*]N1[*][*]C(=[*])C1= O	0.194	6 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-1.63	0 out of 12
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.22	0 out of 7
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.998	1 out of 13

Compound 9

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.125

Enrichment: 0.416

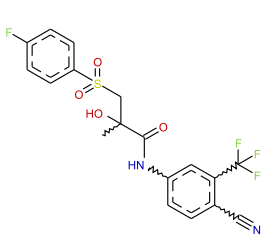
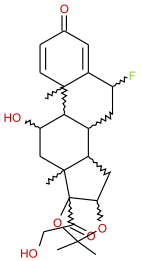
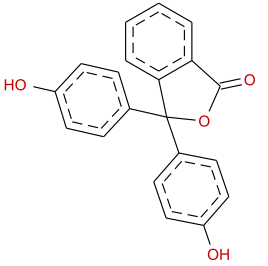
Bayesian Score: -16.3

Mahalanobis Distance: 13.2

Mahalanobis Distance p-value: 7.44e-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	Flunisolide	Phenolphthalein
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.647	0.730	0.752
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

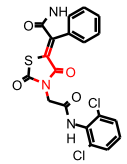
Model Applicability

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

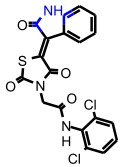
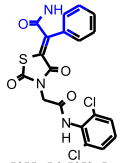
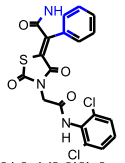
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

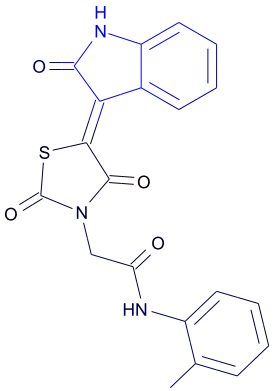
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	565998553	 [*]N1[*][*]C(=[*])C1= O	0.194	6 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-1.63	0 out of 12
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.22	0 out of 7
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.998	1 out of 13

Compound 12

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{20}H_{15}N_3O_4S$

Molecular Weight: 393.4158

ALogP: 2.185

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.136

Enrichment: 0.453

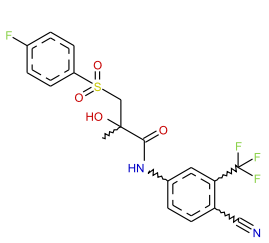
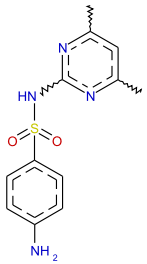
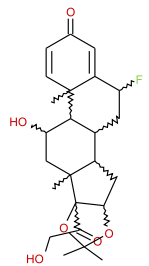
Bayesian Score: -15

Mahalanobis Distance: 15.6

Mahalanobis Distance p-value: 7.9e-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	Bicalutamide	Sulfamethazine	Flunisolide
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.681	0.689	0.710
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

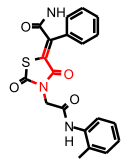
Model Applicability

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

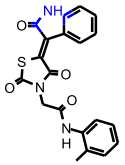
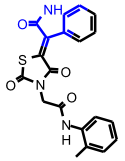
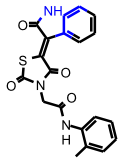
1. OPS PC12 out of range. Value: 3.6573. Training min, max, SD, explained variance: -3.4599, 2.3291, 1.246, 0.0290.

Feature Contribution

Top features for positive contribution

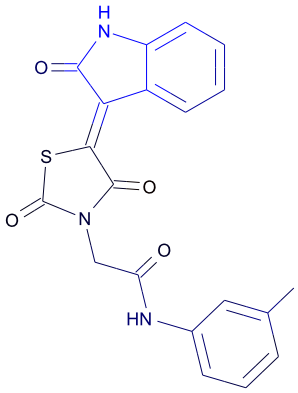
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	565998553	 [*]N1[*][*]C(=[*])C1= O	0.194	6 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-1.63	0 out of 12
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.22	0 out of 7
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.998	1 out of 13

Compound 13

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



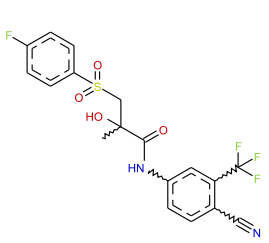
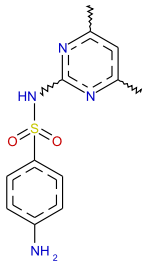
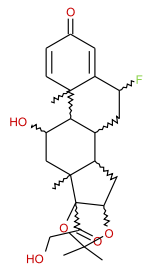
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.132
Enrichment: 0.437
Bayesian Score: -15.6
Mahalanobis Distance: 15.6
Mahalanobis Distance p-value: 7.9e-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	Bicalutamide	Sulfamethazine	Flunisolide
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.671	0.691	0.710
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

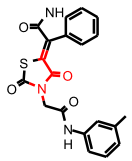
Model Applicability

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

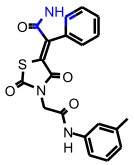
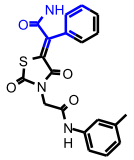
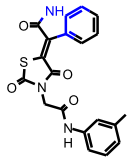
1. OPS PC12 out of range. Value: 3.6573. Training min, max, SD, explained variance: -3.4599, 2.3291, 1.246, 0.0290.

Feature Contribution

Top features for positive contribution

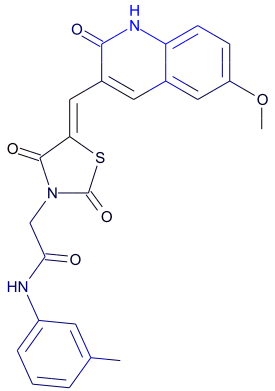
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	565998553	 [*]N1[*][*]C(=[*])C1=O	0.194	6 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]):[*]	-1.63	0 out of 12
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.22	0 out of 7
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.998	1 out of 13

Compound 14

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.143

Enrichment: 0.476

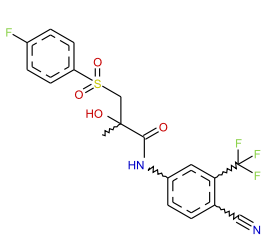
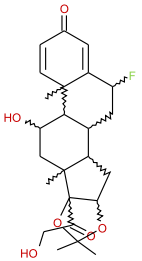
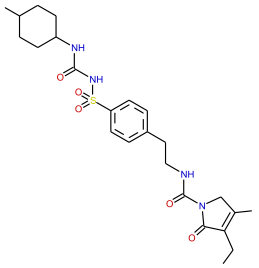
Bayesian Score: -14.2

Mahalanobis Distance: 13.3

Mahalanobis Distance p-value: 6.15e-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	Flunisolide	Glimepride
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.585	0.751	0.753
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

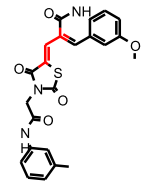
Model Applicability

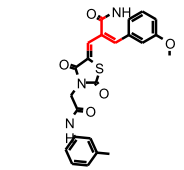
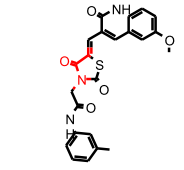
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

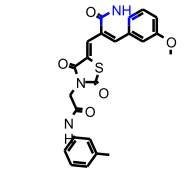
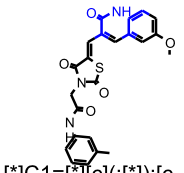
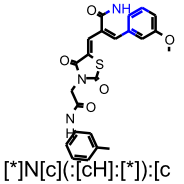
Feature Contribution

Top features for positive contribution

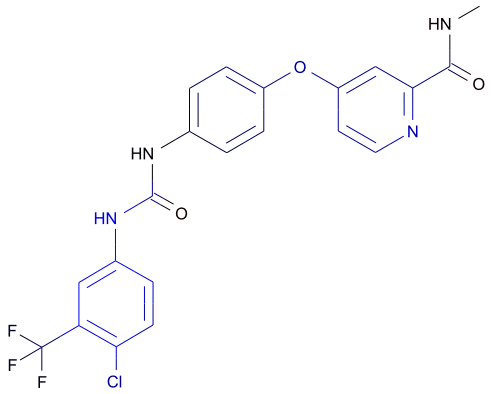
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	451847724	 [*]C(=CC(=[*])[*])[*]	0.3	10 out of 21

FCFP_12	436886043	 [*]\C=C(\C=[*])/C(=[*])[*]	0.27	7 out of 15
FCFP_12	565998553	 [*]N1[*][*]C(=[*])C1= O	0.194	6 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-1.63	0 out of 12
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.22	0 out of 7
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.998	1 out of 13

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.139

Enrichment: 0.461

Bayesian Score: -14.7

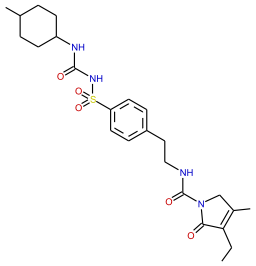
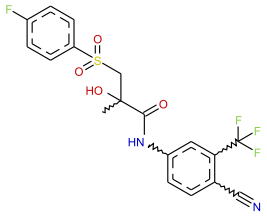
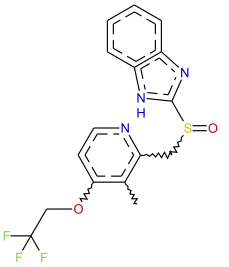
Mahalanobis Distance: 21.3

Mahalanobis Distance p-value: 4.93e-011

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

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

Structural Similar Compounds

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

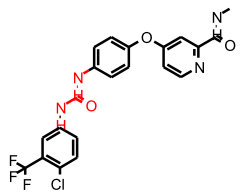
Model Applicability

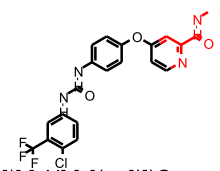
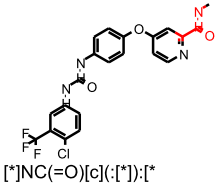
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

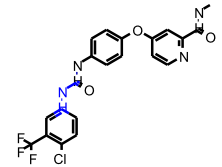
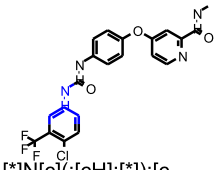
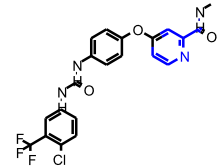
Feature Contribution

Top features for positive contribution

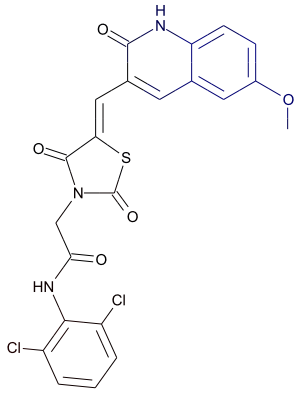
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1499521844	 [*]NC(=O)N[*]	0.39	5 out of 9

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

Top Features for negative contribution

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

Compound 8



$C_{22}H_{15}Cl_2N_3O_5S$
Molecular Weight: 504.34259
ALogP: 3.479
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

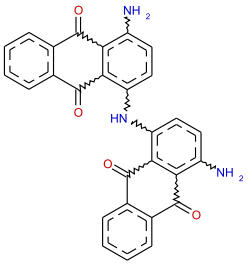
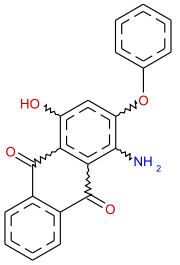
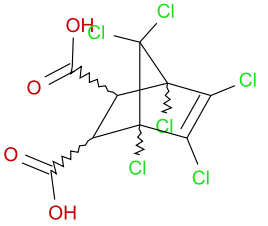
Model Prediction

Prediction: Mild
Probability: 0.722
Enrichment: 1.05
Bayesian Score: -3.1
Mahalanobis Distance: 12.1
Mahalanobis Distance p-value: 4.02e-005

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

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe

Structural Similar Compounds

Name	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.742	0.790	0.790
Reference	28ZPAK-;125;72	28ZPAK 239;72	28ZPAK-;92;72

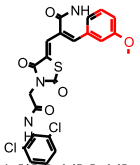
Model Applicability

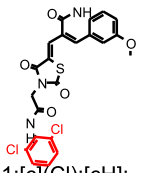
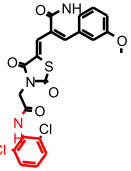
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

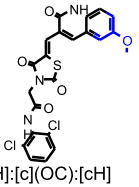
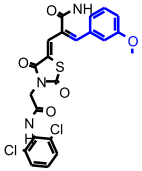
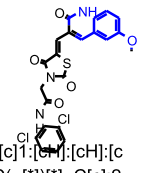
Feature Contribution

Top features for positive contribution

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

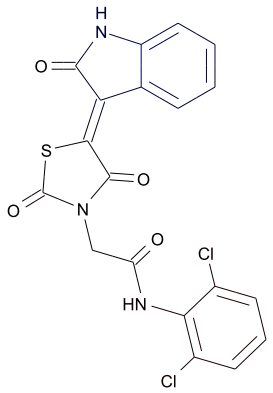
FCFP_10	1161767339	 [*][c]1:[c](Cl):[cH]:[cH]:[cH]:[c]:1Cl	0.186	1 out of 1
FCFP_10	-545052888	 [*]N[c]1:[c]([*]):[cH]:[cH]:[cH]:[c]:1Cl	0.186	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1977641857	 [*]:[cH]:[c](OC):[cH]:[*]	-0.78	4 out of 15
FCFP_10	-1757681964	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[c]:1C=[*]	-0.507	0 out of 1
FCFP_10	723745966	 [*]O[c]1:[cH]:[cH]:[c]2NC(=[*])[*]=C[c]:2:[cH]:1	-0.507	0 out of 1

Compound 9

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.787

Enrichment: 1.14

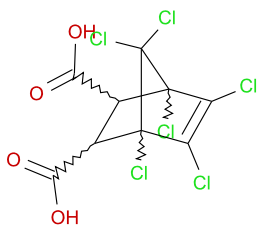
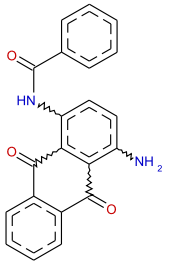
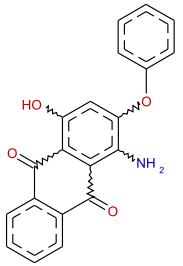
Bayesian Score: -1.46

Mahalanobis Distance: 11.3

Mahalanobis Distance p-value: 0.00136

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

Structural Similar Compounds

Name	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.651	0.656	0.668
Reference	28ZPAK-;92;72	28ZPAK-;124;72	28ZPAK 239;72

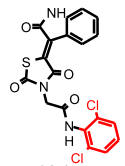
Model Applicability

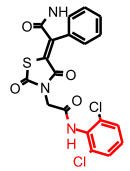
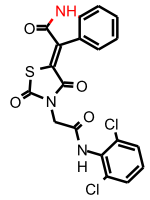
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

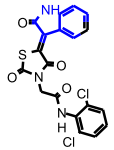
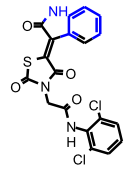
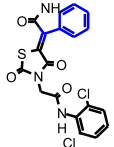
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1161767339	 [*][c]1:[c](Cl):[cH]:[cH]:[cH]:[c]:1Cl	0.186	1 out of 1

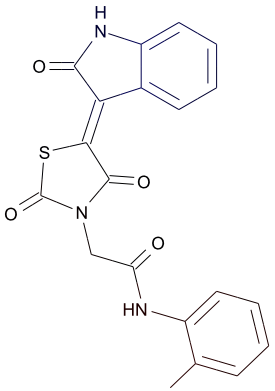
FCFP_10	-545052888	 [*]N[c]1:[c]([*]):[cH]:[cH]:[cH]:[c]:1Cl	0.186	1 out of 1
FCFP_10	3	 [*]N[*]	0.165	383 out of 491

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1011367537	 [*]=C1[*]=C[c]2:[cH]: [*]:[cH]:[cH]:[c]:2N 1	-0.329	4 out of 9
FCFP_10	-773983804	 [*]N[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-0.294	50 out of 102
FCFP_10	-1698724694	 [*]=C1[*][*][c]2:[cH] :[cH]:[cH]:[cH]:[c]1 :2	-0.284	53 out of 107

Compound 12

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



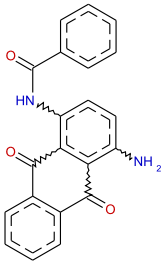
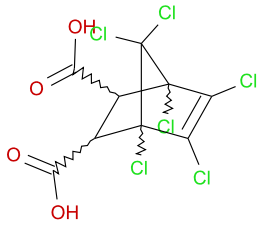
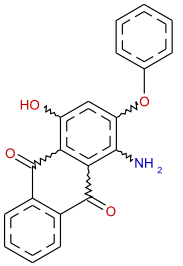
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild
Probability: 0.791
Enrichment: 1.15
Bayesian Score: -1.32
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.0373

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

Structural Similar Compounds

Name	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.604	0.629	0.629
Reference	28ZPAK-;124;72	28ZPAK-;92;72	28ZPAK 239;72

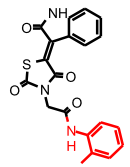
Model Applicability

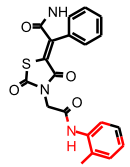
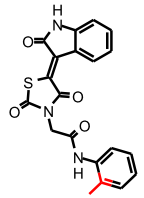
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

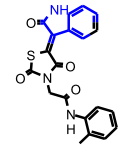
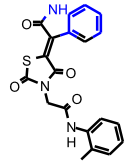
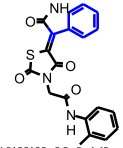
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1396506317	 [*]N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.317	4 out of 4

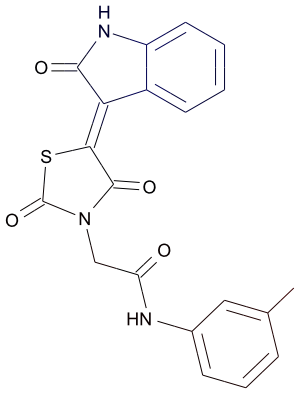
FCFP_10	755520106	 [*]N[c]1:[cH]:[*]:[cH]:[cH]:[c]:1C	0.273	9 out of 10
FCFP_10	136120670	 [*]:[c](:[*])C	0.206	53 out of 65

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1011367537	 [*]=C1[*]=C[c]2:[cH]:[*]:[cH]:[cH]:[c]:2N1	-0.329	4 out of 9
FCFP_10	-773983804	 [*]N[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-0.294	50 out of 102
FCFP_10	-1698724694	 [*]=C1[*][*][c]2:[cH]:[cH]:[cH]:[cH]:[c]1:2	-0.284	53 out of 107

Compound 13

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{20}H_{15}N_3O_4S$

Molecular Weight: 393.4158

ALogP: 2.185

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.787

Enrichment: 1.14

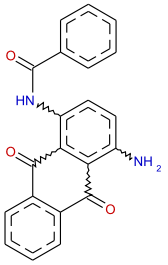
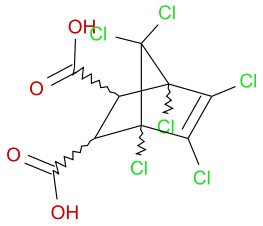
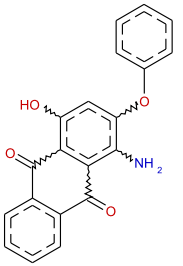
Bayesian Score: -1.48

Mahalanobis Distance: 10.4

Mahalanobis Distance p-value: 0.0373

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

Structural Similar Compounds

Name	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.606	0.630	0.631
Reference	28ZPAK-;124;72	28ZPAK-;92;72	28ZPAK 239;72

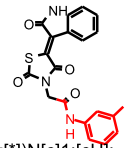
Model Applicability

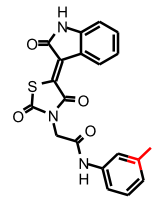
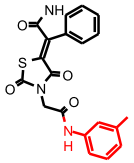
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

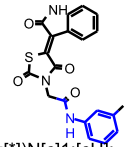
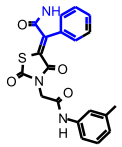
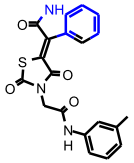
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	393262357	 [*]C(=[*])N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.294	3 out of 3

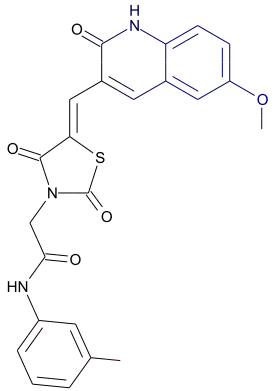
FCFP_10	136120670	 [*]:[c](:[*])C	0.206	53 out of 65
FCFP_10	2104181982	 [*]N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.186	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-792685140	 [*]C(=[*])N[c]1:[cH]:[cH]:[cH]:[c]([*]):[cH]:1	-0.361	2 out of 5
FCFP_10	1011367537	 [*]=C1[*]=C[c]2:[cH]:[*]:[cH]:[cH]:[c]:2N1	-0.329	4 out of 9
FCFP_10	-773983804	 [*]N[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-0.294	50 out of 102

Compound 14

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.72

Enrichment: 1.04

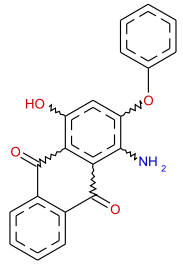
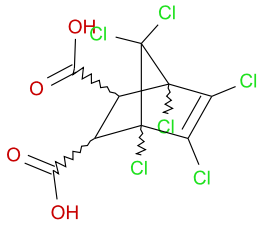
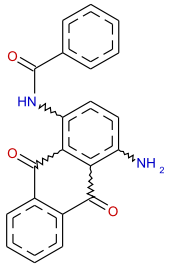
Bayesian Score: -3.15

Mahalanobis Distance: 11

Mahalanobis Distance p-value: 0.00428

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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-AMINO-4-HYDROXY-2-PHENOXY-	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE
Structure			
Actual Endpoint	Mild	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.713	0.727	0.734
Reference	28ZPAK 239;72	28ZPAK-;92;72	28ZPAK-;124;72

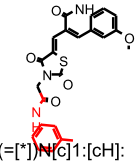
Model Applicability

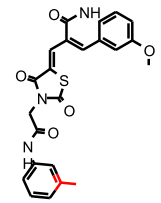
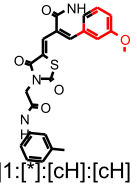
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

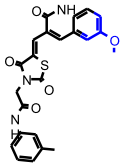
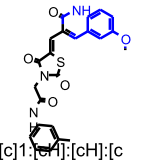
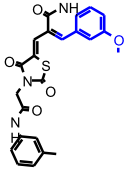
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	393262357	 [*]C(=[*])N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.294	3 out of 3

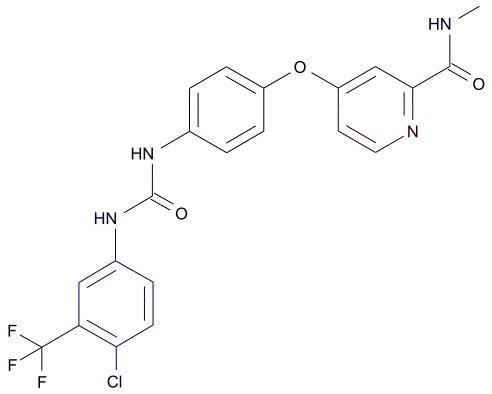
FCFP_10	136120670	 [*]:[c](:[*])C	0.206	53 out of 65
FCFP_10	346218766	 [*][c]1:[*]:[cH]:[cH] :[c](OC):[cH]:1	0.197	30 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1977641857	 [*]:[cH]:[c](OC):[cH] :[*]	-0.78	4 out of 15
FCFP_10	723745966	 [*]O[c]1:[cH]:[cH]:[c]2NC(=[*])[*]=C[c]:2 :[cH]:1	-0.507	0 out of 1
FCFP_10	-1757681964	 [*][c]1:[cH]:[cH]:[c] (OC):[cH]:[c]:1C=[*]	-0.507	0 out of 1

Sorafenib

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.776

Enrichment: 1.13

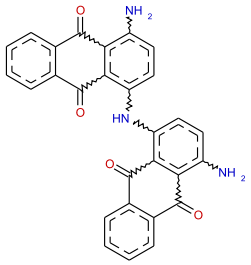
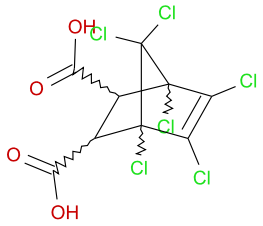
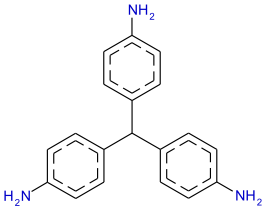
Bayesian Score: -1.8

Mahalanobis Distance: 8.95

Mahalanobis Distance p-value: 0.537

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

Structural Similar Compounds

Name	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	METHANE;TRIS(4-AMINOPHENYL)-
Structure			
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe
Distance	0.799	0.816	0.827
Reference	28ZPAK-;125;72	28ZPAK-;92;72	28ZPAK-;73;72

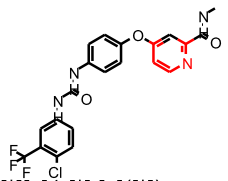
Model Applicability

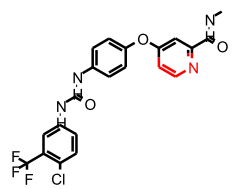
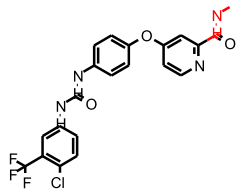
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

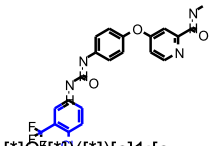
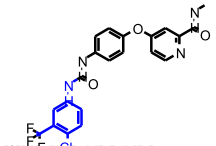
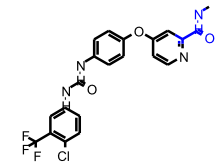
Feature Contribution

Top features for positive contribution

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

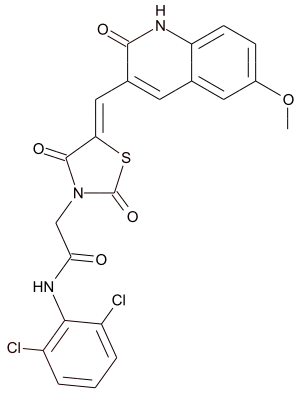
FCFP_10	-124655670	 [*]:[cH]:[cH]:n:[*]	0.259	14 out of 16
FCFP_10	-885550502	 [*]C(=[*])NC	0.239	54 out of 64

Top Features for negative contribution

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

Compound 8

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

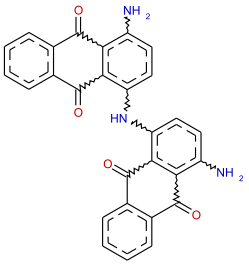
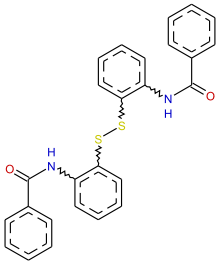
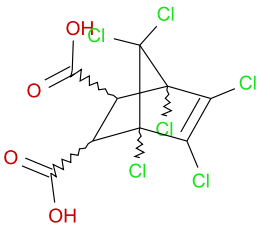
Bayesian Score: 1.73

Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.000107

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	BENZANILIDE;2';2"'-DITHIOBIS-	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.741	0.771	0.774
Reference	28ZPAK-;125;72	28ZPAK-;173;72	28ZPAK-;92;72

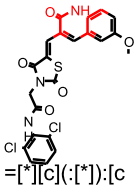
Model Applicability

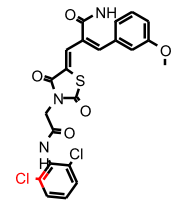
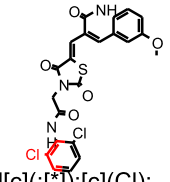
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

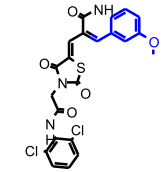
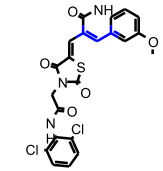
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.198	14 out of 14

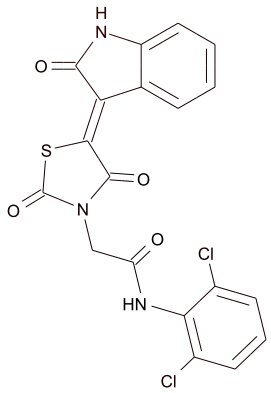
FCFP_12	71476542	 [*]:[c](:[*])Cl	0.175	81 out of 84
FCFP_12	367998008	 [*][c](:[*]):[c](Cl):[cH]:[*]	0.172	75 out of 78

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1060187936	 [*]N[c]1:[cH]:[cH]:[c](OC):[cH]:[c]:1[*]	-0.344	2 out of 4
FCFP_12	-1757681964	 [*][c]1:[cH]:[cH]:[c](OC):[cH]:[c]:1C=[*]	-0.268	1 out of 2
FCFP_12	451371068	 [*]C(=C[c](:[*]):[*])[*]	-0.167	6 out of 9

Compound 9

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

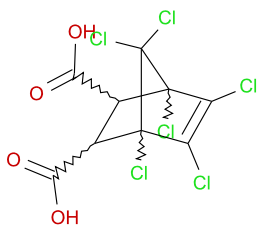
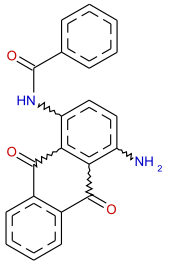
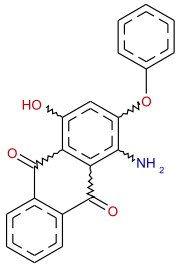
Bayesian Score: 2.14

Mahalanobis Distance: 8.64

Mahalanobis Distance p-value: 0.699

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

Structural Similar Compounds

Name	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.645	0.651	0.659
Reference	28ZPAK-;92;72	28ZPAK-;124;72	28ZPAK 239;72

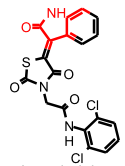
Model Applicability

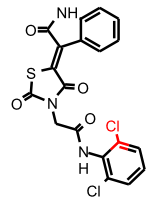
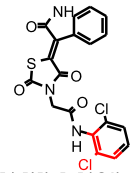
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

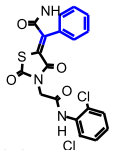
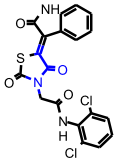
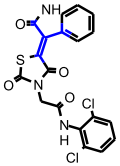
Feature Contribution

Top features for positive contribution

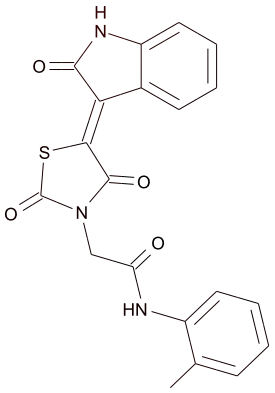
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.198	14 out of 14

FCFP_12	71476542	 [*]:[c](:[*])Cl	0.175	81 out of 84
FCFP_12	367998008	 [*][c](:[*]):[c](Cl):[cH]:[*]	0.172	75 out of 78

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]=C1[*][*][c]2:[cH]:[cH]:[cH]:[cH]:[c]1:2	-0.0964	107 out of 146
FCFP_12	565998553	 [*]N1[*][*]C(=[*])C1=O	-0.0662	198 out of 262
FCFP_12	-1678275541	 [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]	-0.0561	3 out of 4

Compound 12



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

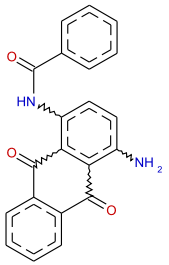
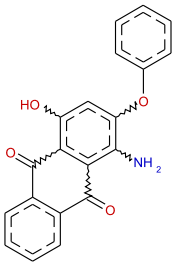
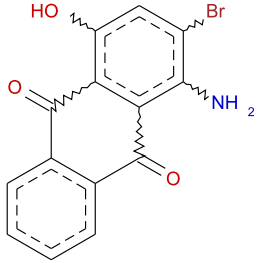
Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 1.88
Mahalanobis Distance: 8.23
Mahalanobis Distance p-value: 0.865

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

Structural Similar Compounds

Name	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	Anthraquinone; 1-amino-2-bromo-4-hydroxy-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.601	0.618	0.627
Reference	28ZPAK-;124;72	28ZPAK 239;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 535;86

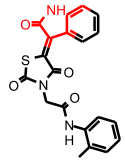
Model Applicability

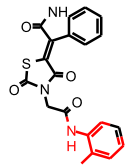
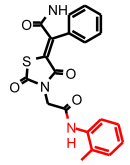
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

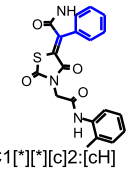
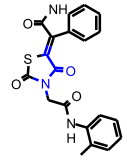
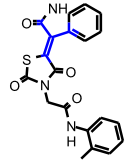
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.198	14 out of 14

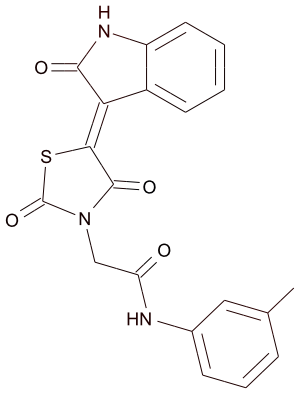
FCFP_12	755520106	 [*]N[c]1:[cH]:[*]:[cH]:[cH]:[c]:1C	0.192	10 out of 10
FCFP_12	1396506317	 [*]N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.167	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]=C1[*][*]c2:[cH]:[cH]:[cH]:[cH]:c1:2	-0.0964	107 out of 146
FCFP_12	565998553	 [*]N1[*][*]C(=[*])C1=O	-0.0662	198 out of 262
FCFP_12	-1678275541	 [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]	-0.0561	3 out of 4

Compound 13

TOPKAT_Ocular_Irritancy_None_vs_Irritant



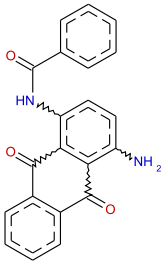
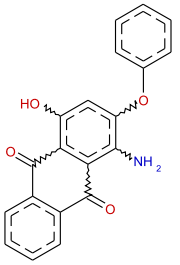
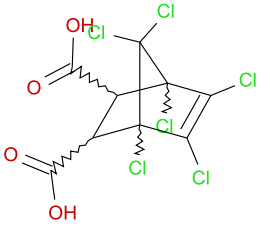
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 2.18
Mahalanobis Distance: 8.23
Mahalanobis Distance p-value: 0.865

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

Structural Similar Compounds

Name	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.603	0.619	0.628
Reference	28ZPAK-;124;72	28ZPAK 239;72	28ZPAK-;92;72

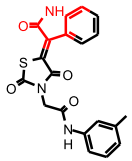
Model Applicability

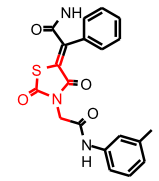
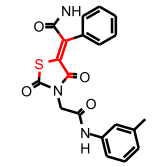
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

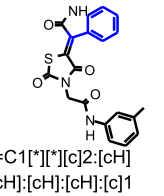
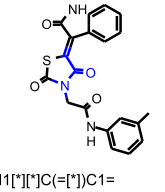
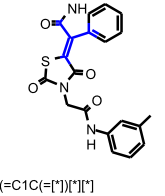
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.198	14 out of 14

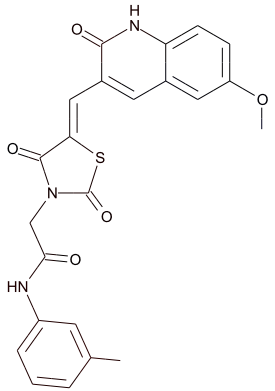
FCFP_12	2036120522	 [*]CN1C(=O)SC(=[*])C1 =[*]	0.167	4 out of 4
FCFP_12	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.167	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]=C1[*][*][c]2:[cH] :[cH]:[cH]:[cH]:[c]1 :2	-0.0964	107 out of 146
FCFP_12	565998553	 [*]N1[*][*]C(=[*])C1= O	-0.0662	198 out of 262
FCFP_12	-1678275541	 [*]C(=C1C(=[*])[*][*] :[c]1:[*])[*]	-0.0561	3 out of 4

Compound 14

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

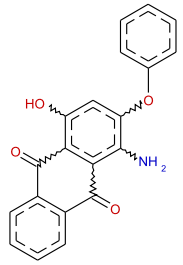
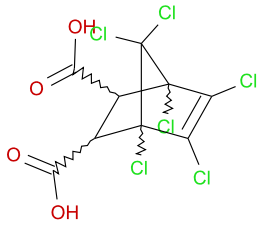
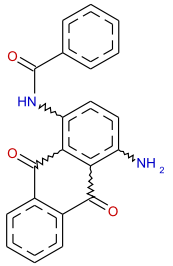
Bayesian Score: 1.76

Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.000646

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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-AMINO-4-HYDROXY-2-PHENOXY-	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.701	0.717	0.725
Reference	28ZPAK 239;72	28ZPAK-;92;72	28ZPAK-;124;72

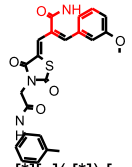
Model Applicability

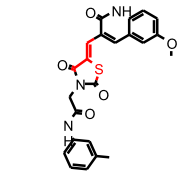
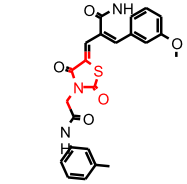
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

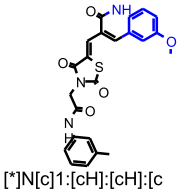
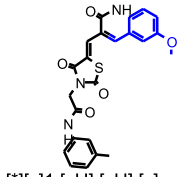
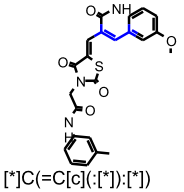
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	0.198	14 out of 14

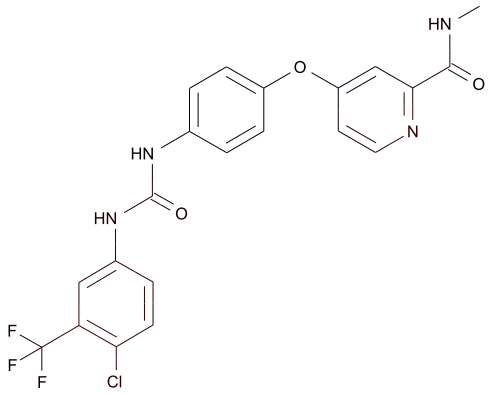
FCFP_12	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.167	4 out of 4
FCFP_12	2036120522	 [*]CN1C(=O)SC(=[*])C1 =[*]	0.167	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1060187936	 [*]N[c]1:[cH]:[cH]:[c](OC):[cH]:[c]:1[*]	-0.344	2 out of 4
FCFP_12	-1757681964	 [*][c]1:[cH]:[cH]:[c] (OC):[cH]:[c]:1C=[*]	-0.268	1 out of 2
FCFP_12	451371068	 [*]C(=C[c](:[*]):[*]) [*]	-0.167	6 out of 9

Sorafenib

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

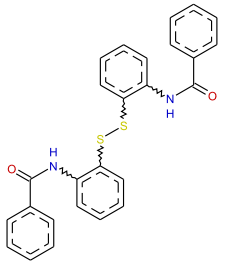
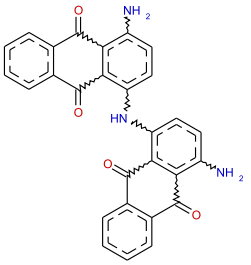
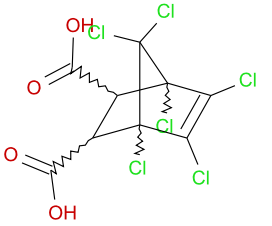
Bayesian Score: 3.04

Mahalanobis Distance: 6.28

Mahalanobis Distance p-value: 1

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

Structural Similar Compounds

Name	BENZANILIDE;2';2'''-DITHIOBIS-	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Non-Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.743	0.791	0.801
Reference	28ZPAK-;173;72	28ZPAK-;125;72	28ZPAK-;92;72

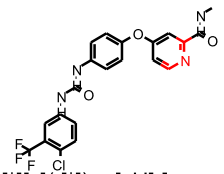
Model Applicability

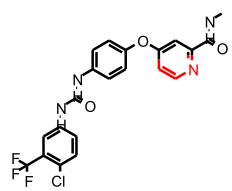
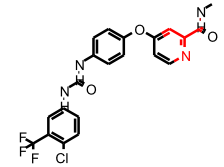
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

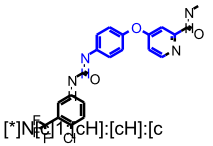
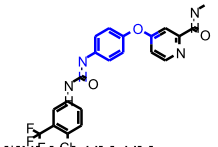
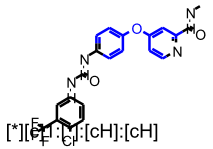
Feature Contribution

Top features for positive contribution

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

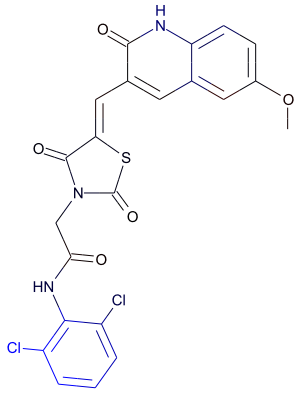
FCFP_12	-124655670	 [*]:[cH]:[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1539132615	 [*]C(=[*])[c](:[cH]:[*]):n:[*]	0.197	13 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	702861189	 [*]N[c]1:[cH]:[cH]:[c](O[c]2:[cH]:[cH]:[*]:[c]([*]):[cH]:2):[cH]:[cH]:1	-0.268	1 out of 2
FCFP_12	-747629521	 [*]N[c]1:[cH]:[cH]:[c](O[c](:[*]):[*]):[cH]:[cH]:1	-0.268	1 out of 2
FCFP_12	-215363676	 [*]N[c]1:[cH]:[cH]:[c](O[c]2:[cH]:[cH]:[*]:[cH]:[cH]:2):[cH]:[cH]:1	0	4 out of 5

Compound 8

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.19

Enrichment: 0.589

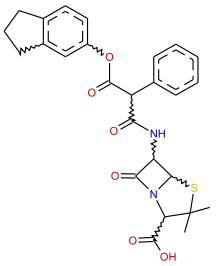
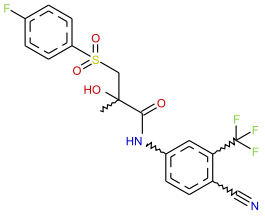
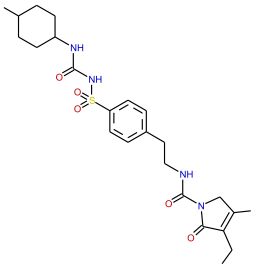
Bayesian Score: -8.6

Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.00281

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

Structural Similar Compounds

Name	Carbenicillin	Bicalutamide	Glimepride
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.618	0.650	0.687
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

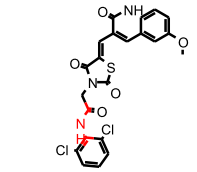
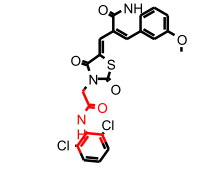
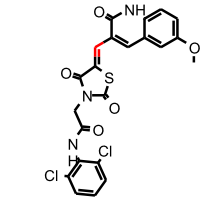
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC30 out of range. Value: 3.2851. Training min, max, SD, explained variance: -2.9582, 2.682, 0.9684, 0.0098.
2. Unknown ECFP_2 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
3. Unknown ECFP_2 feature: 1182722866: [*]C(=CC(=[*])[*])[*]
4. Unknown ECFP_2 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
5. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

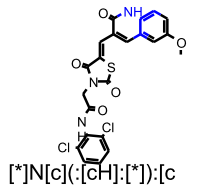
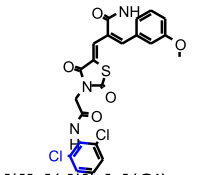
Feature Contribution

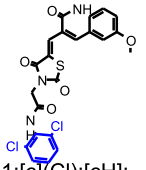
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

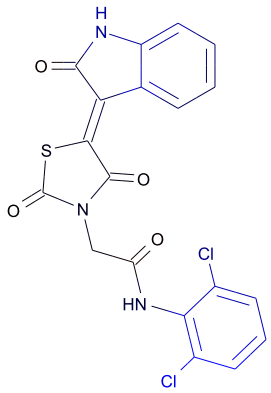
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17
ECFP_12	574399351	 [*]CC(=O)N[c](:[c]([*])):[*]):[c]([*]):[*]	0.421	1 out of 1
ECFP_12	-1925046727	 [*]C=[*]	0.407	16 out of 33

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335833675	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-1.25	0 out of 8
ECFP_12	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-1.11	2 out of 26

ECFP_12	-1354065290	 [*][c]1:[c](Cl):[cH]:[cH]:[cH]:[c]:1Cl	-0.941	0 out of 5

Compound 9



$C_{19}H_{11}Cl_2N_3O_4S$
Molecular Weight: 448.27933
ALogP: 3.028
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

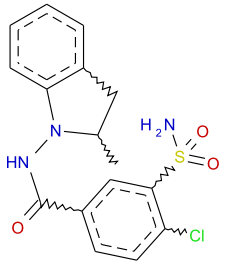
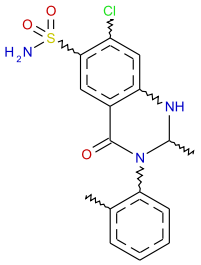
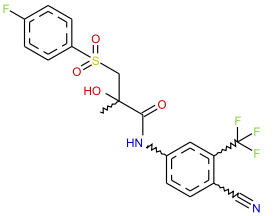
Model Prediction

Prediction: Non-Carcinogen
Probability: 0.175
Enrichment: 0.545
Bayesian Score: -11.5
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 0.205

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Indapamide	Metolazone	Bicalutamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.634	0.643	0.664
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

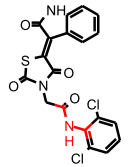
Model Applicability

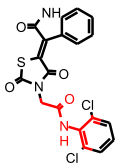
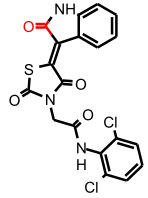
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

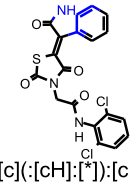
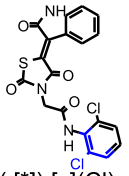
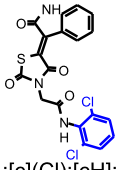
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17

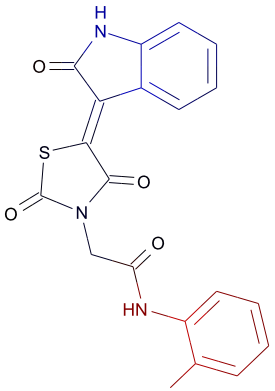
ECFP_12	574399351	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	0.421	1 out of 1
ECFP_12	2106656448	 [*]C(=O)[*]	0.141	30 out of 83

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335833675	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-1.25	0 out of 8
ECFP_12	1335691903	 [*][c](:[*]):[c](Cl): [cH]:[*]	-1.11	2 out of 26
ECFP_12	-1354065290	 [*][c]1:[c](Cl):[cH]: [cH]:[cH]:[c]:1Cl	-0.941	0 out of 5

Compound 12

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{20}H_{15}N_3O_4S$

Molecular Weight: 393.4158

ALogP: 2.185

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.261

Enrichment: 0.809

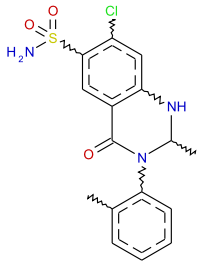
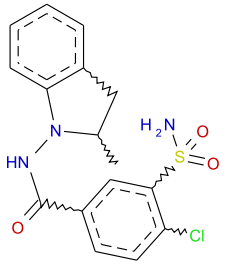
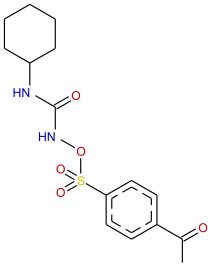
Bayesian Score: -2.1

Mahalanobis Distance: 7.91

Mahalanobis Distance p-value: 0.991

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

Structural Similar Compounds

Name	Metolazone	Indapamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.575	0.583	0.629
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

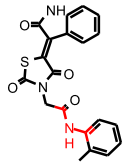
Model Applicability

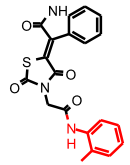
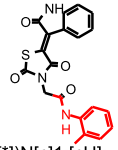
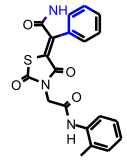
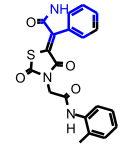
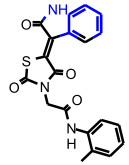
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

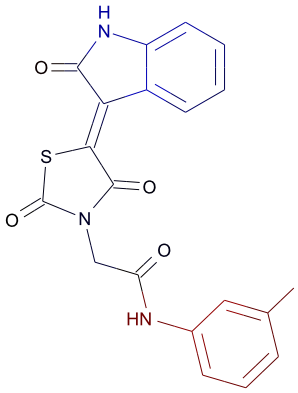
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17

ECFP_12	-52177950	 [*]N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.421	1 out of 1
ECFP_12	1360781590	 [*]C(=[*])N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.421	1 out of 1
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_12	1335833675	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-1.25	0 out of 8
ECFP_12	-319922023	 [*]=C1N[c]2:[cH]:[cH]:[*]:[cH]:[c]:2C1=[*]	-0.661	0 out of 3
ECFP_12	1640720160	 [*]1[*][c]2:[*]:[cH]:[cH]:[cH]:[c]:2N1	-0.485	0 out of 2

Compound 13



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

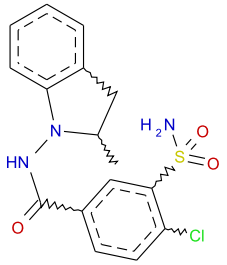
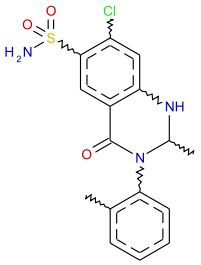
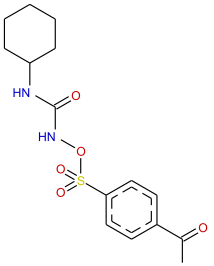
Model Prediction

Prediction: Non-Carcinogen
Probability: 0.248
Enrichment: 0.769
Bayesian Score: -2.95
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 0.266

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	Indapamide	Metolazone	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.580	0.582	0.630
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

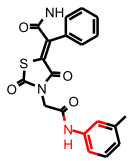
Model Applicability

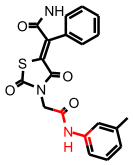
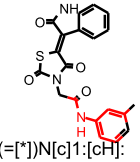
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

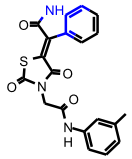
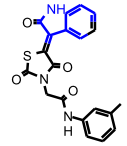
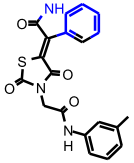
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10

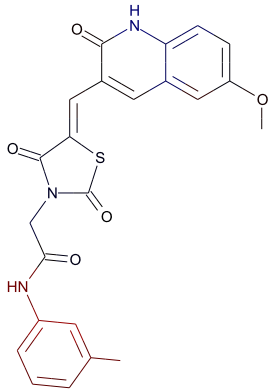
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17
ECFP_12	1435111106	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.445	3 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335833675	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-1.25	0 out of 8
ECFP_12	-319922023	 [*]=C1N[c]2:[cH]:[cH]:[*]:[cH]:[c]:2C1=[*]	-0.661	0 out of 3
ECFP_12	1640720160	 [*]1[*][c]2:[*]:[cH]:[cH]:[cH]:[c]:2N1	-0.485	0 out of 2

Compound 14

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.302

Enrichment: 0.937

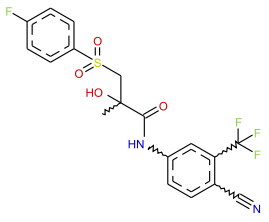
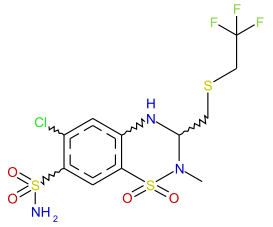
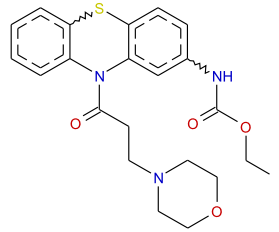
Bayesian Score: 0.16

Mahalanobis Distance: 12.1

Mahalanobis Distance p-value: 0.00345

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Polythiazide	Moricizine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.586	0.632	0.650
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

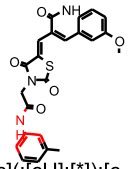
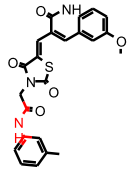
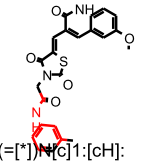
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC30 out of range. Value: 2.8769. Training min, max, SD, explained variance: -2.9582, 2.682, 0.9684, 0.0098.
2. Unknown ECFP_2 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
3. Unknown ECFP_2 feature: 1182722866: [*]C(=CC(=[*])[*])[*]
4. Unknown ECFP_2 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
5. Unknown ECFP_2 feature: 190445529: [*]N1[*][*]SC1=O

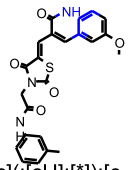
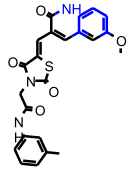
Feature Contribution

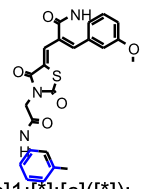
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

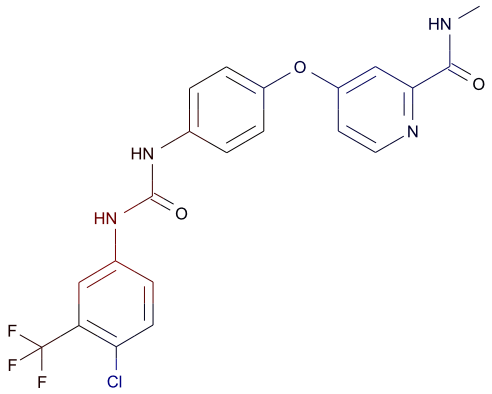
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[c H]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]): [*]	0.46	9 out of 17
ECFP_12	1435111106	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[c]([*]):[c H]:1	0.445	3 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1335833675	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-1.25	0 out of 8
ECFP_12	2083628577	 [*]N[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	-0.811	0 out of 4

ECFP_12	2007300961	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.426	7 out of 36

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.236

Enrichment: 0.734

Bayesian Score: -3.76

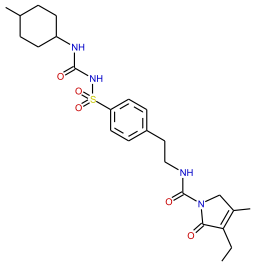
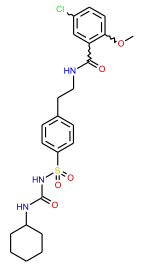
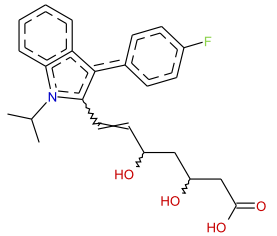
Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.00229

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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

Structural Similar Compounds

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

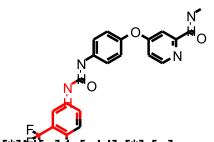
Model Applicability

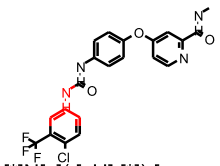
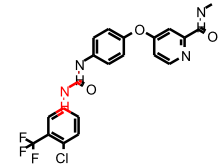
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

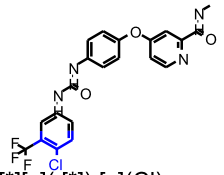
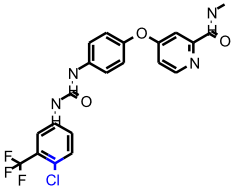
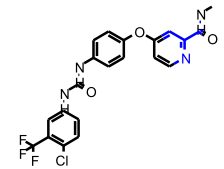
Feature Contribution

Top features for positive contribution

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

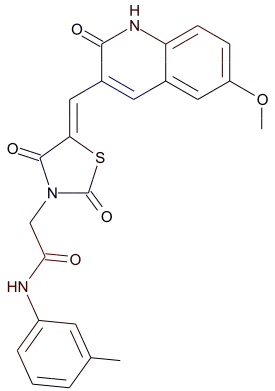
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17

Top Features for negative contribution

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

Compound 14

TOPKAT_Rat_Female_FDA_Single_vs_Multiple



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.569

Enrichment: 1.52

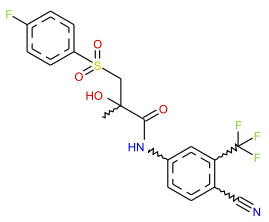
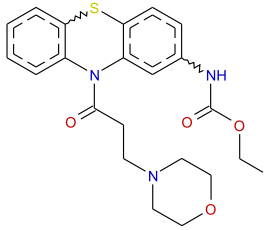
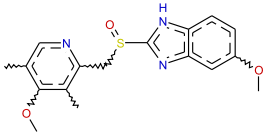
Bayesian Score: 2.55

Mahalanobis Distance: 14.4

Mahalanobis Distance p-value: 1.07e-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	Bicalutamide	Moricizine	Omeprazole
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.529	0.623	0.690
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

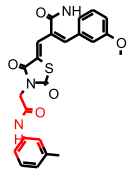
Model Applicability

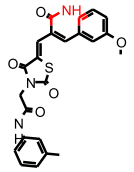
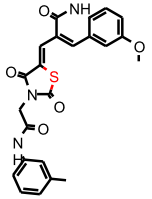
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC16 out of range. Value: 2.7978. Training min, max, SD, explained variance: -2.5301, 2.7699, 1.075, 0.0171.

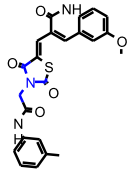
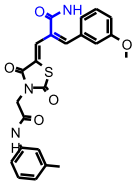
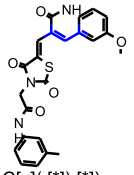
Feature Contribution

Top features for positive contribution

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

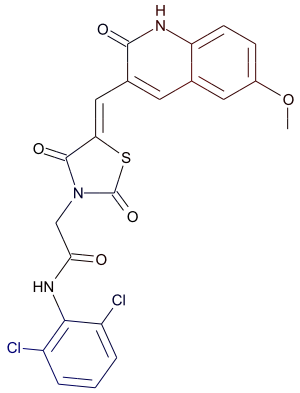
SCFP_4	1631845520	 [*]C(=[*])N[c](:[*]):[*]	0.601	6 out of 9
SCFP_4	17	 [*]S[*]	0.548	10 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	399659969	 [*]CN1C(=[*])[*][*]C1=[*]	-0.666	0 out of 3
SCFP_4	1257024795	 [*]NC(=O)C(=[*])[*]	-0.666	0 out of 3
SCFP_4	-1971137145	 [*]C(=C[c](:[*]):[*])[*]	-0.489	0 out of 2

Compound 8

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.292

Enrichment: 0.875

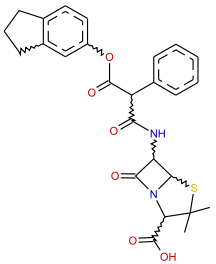
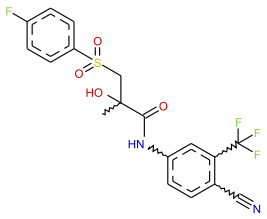
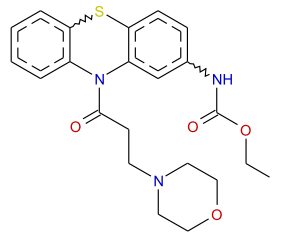
Bayesian Score: -2.44

Mahalanobis Distance: 13.3

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	Carbenicillin	Bicalutamide	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.602	0.624	0.667
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

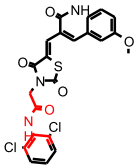
Model Applicability

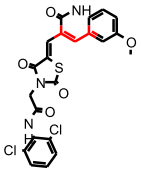
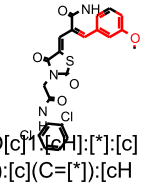
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC16 out of range. Value: -4.5449. Training min, max, SD, explained variance: -4.5272, 4.5629, 1.366, 0.0174.

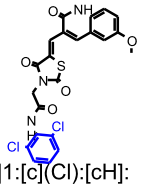
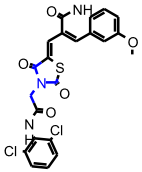
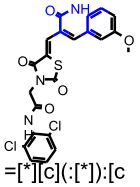
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	0.437	7 out of 13

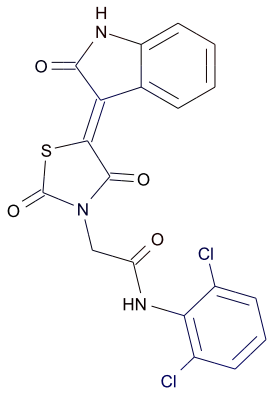
SCFP_6	-1971137145	 [*]C(=C[c](:[*]):[*]) [*]	0.434	5 out of 9
SCFP_6	392579710	 [*]O[c]1:[cH]:[*]:[c] ([*]):[c](C=[*]):[cH]:1	0.425	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1062412764	 [*][c]1:[c](Cl):[cH]: [cH]:[cH]:[c]:1Cl	-0.957	0 out of 5
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.578	1 out of 8
SCFP_6	2102703671	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-0.496	0 out of 2

Compound 9

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.275

Enrichment: 0.824

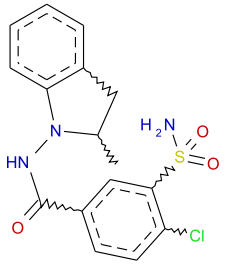
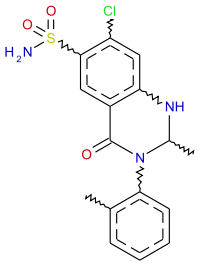
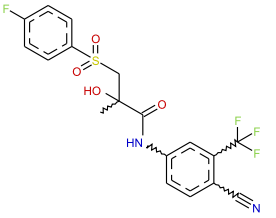
Bayesian Score: -3.18

Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.38

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

Structural Similar Compounds

Name	Indapamide	Metolazone	Bicalutamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.608	0.617	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

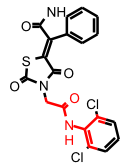
Model Applicability

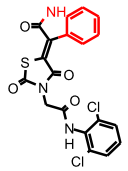
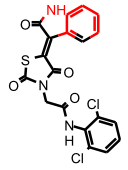
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

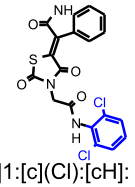
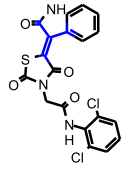
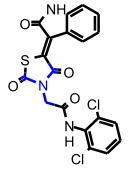
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	0.437	7 out of 13

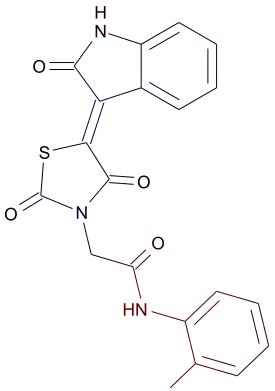
SCFP_6	1655488245	 [*]1[*][c]2:[cH]:[cH] :[cH]:[cH]:[c]:2N1	0.252	4 out of 9
SCFP_6	-1375926917	 [*]N[c]1:[cH]:[cH]:[c]([*]):[*]:[c]:1[*]	0.251	11 out of 26

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1062412764	 [*][c]1:[c](Cl):[cH]: [cH]:[cH]:[c]:1Cl	-0.957	0 out of 5
SCFP_6	1798334293	 [*]C(=C1C(=[*])[*][*] :[c]1:[*])[*]	-0.674	0 out of 3
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1 =[*]	-0.578	1 out of 8

Compound 12

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



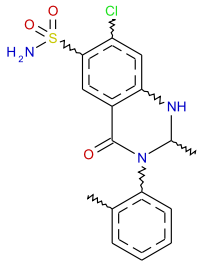
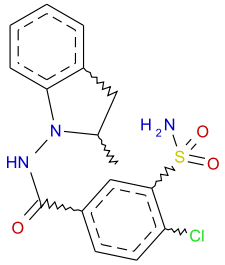
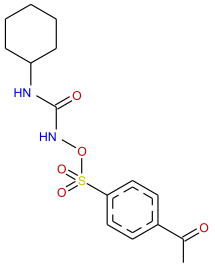
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.364
Enrichment: 1.09
Bayesian Score: 0.261
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 0.276

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

Structural Similar Compounds

Name	Metolazone	Indapamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.537	0.554	0.609
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

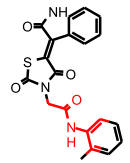
Model Applicability

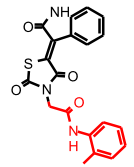
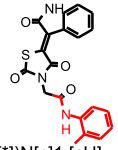
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

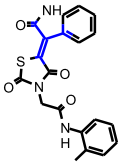
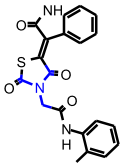
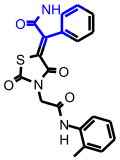
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	0.437	7 out of 13

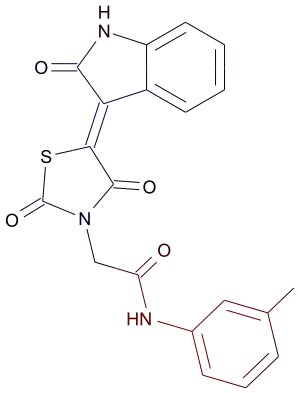
SCFP_6	-900973957	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.415	1 out of 1
SCFP_6	548903629	 [*]C(=[*])N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.415	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1798334293	 [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]	-0.674	0 out of 3
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1=[*]	-0.578	1 out of 8
SCFP_6	2102703671	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-0.496	0 out of 2

Compound 13

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



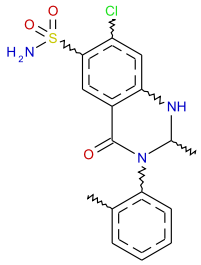
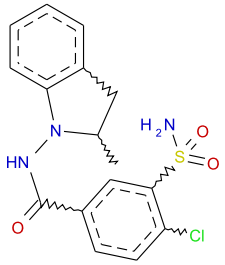
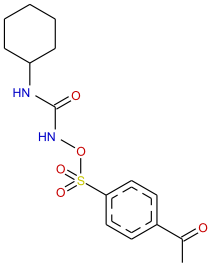
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.38
Enrichment: 1.14
Bayesian Score: 0.812
Mahalanobis Distance: 10
Mahalanobis Distance p-value: 0.647

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

Structural Similar Compounds

Name	Metolazone	Indapamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.555	0.560	0.610
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

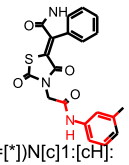
Model Applicability

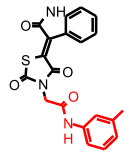
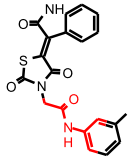
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

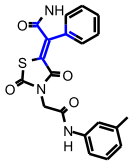
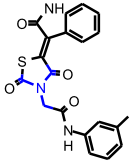
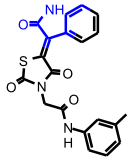
Feature Contribution

Top features for positive contribution

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

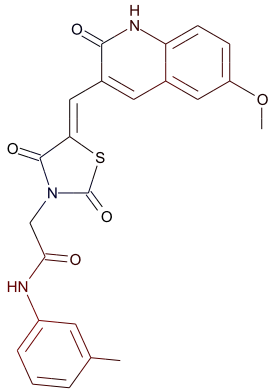
SCFP_6	-236487363	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.603	2 out of 2
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	0.437	7 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1798334293	 [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]	-0.674	0 out of 3
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1=[*]	-0.578	1 out of 8
SCFP_6	2102703671	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-0.496	0 out of 2

Compound 14

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.428

Enrichment: 1.28

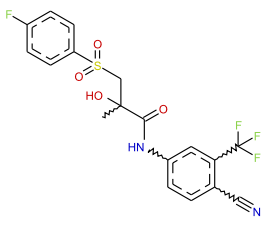
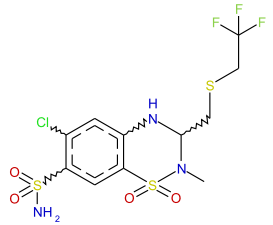
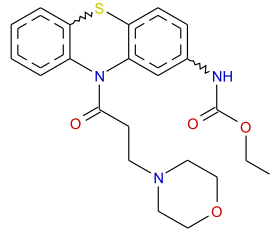
Bayesian Score: 2.3

Mahalanobis Distance: 12.3

Mahalanobis Distance p-value: 0.0121

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Polythiazide	Moricizine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.546	0.629	0.629
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

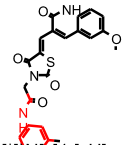
Model Applicability

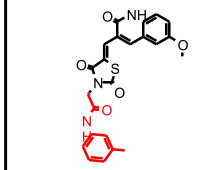
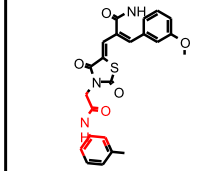
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC16 out of range. Value: -4.5669. Training min, max, SD, explained variance: -4.5272, 4.5629, 1.366, 0.0174.

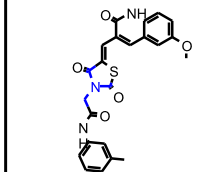
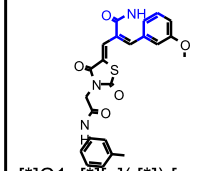
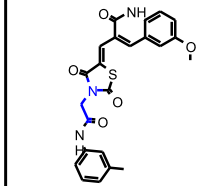
Feature Contribution

Top features for positive contribution

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

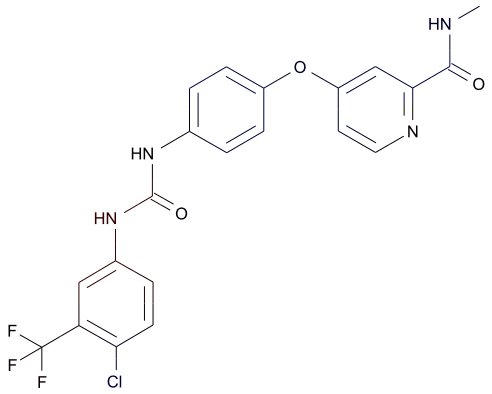
SCFP_6	-236487363	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.603	2 out of 2
SCFP_6	2097618059	 [*]CC(=O)N[c](:[c]([*]):[*]):[c]([*]):[*]	0.437	7 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	399659969	 [*]CN1C(=[*])[*][*]C1=[*]	-0.578	1 out of 8
SCFP_6	2102703671	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-0.496	0 out of 2
SCFP_6	-587539325	 [*]N([*])CC(=[*])[*]	-0.264	1 out of 5

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

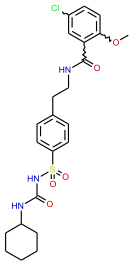
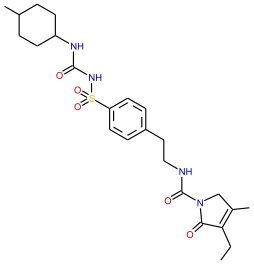
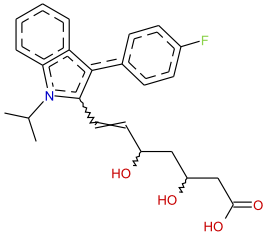
Bayesian Score: -2.4

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

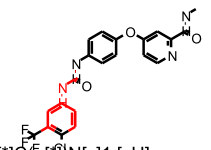
Model Applicability

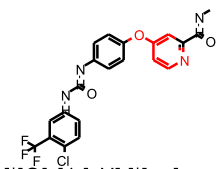
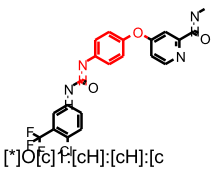
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

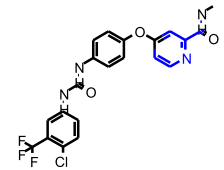
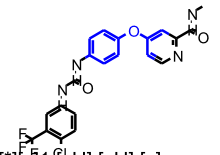
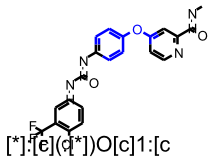
Feature Contribution

Top features for positive contribution

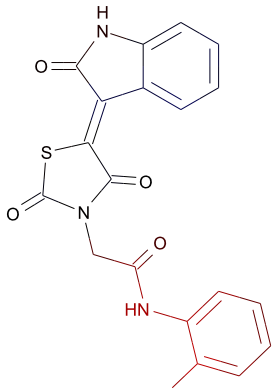
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.615	5 out of 7

SCFP_6	-754059116	 [*]O[c]1:[cH]:[*]:n:[cH]:[cH]:1	0.415	1 out of 1
SCFP_6	-531283893	 [*]O[c]1:[cH]:[cH]:[c](NC(=[*])[*]):[cH]:[cH]:1	0.273	2 out of 4

Top Features for negative contribution

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

Compound 12



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

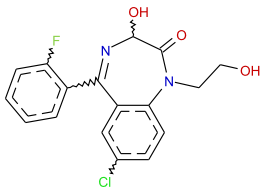
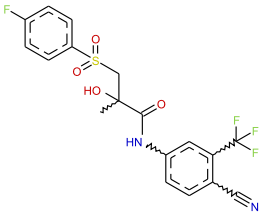
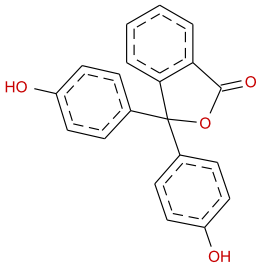
Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.558
Enrichment: 1.35
Bayesian Score: 3.34
Mahalanobis Distance: 12.3
Mahalanobis Distance p-value: 0.00244

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

Structural Similar Compounds

Name	Doxefazepam	Bicalutamide	Phenolphthalein
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.633	0.679	0.723
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

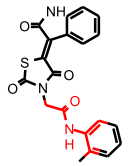
Model Applicability

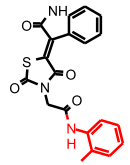
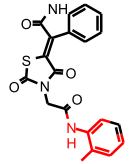
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC18 out of range. Value: -3.3972. Training min, max, SD, explained variance: -2.5092, 3.7383, 1.072, 0.0156.

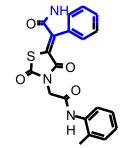
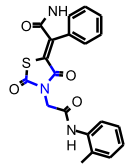
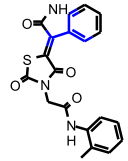
Feature Contribution

Top features for positive contribution

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

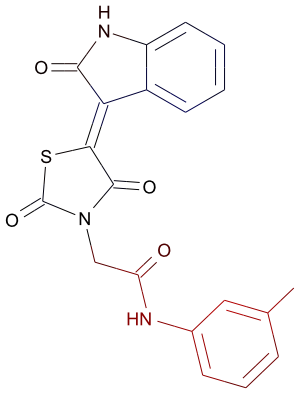
SCFP_8	1269778311	 [*]N[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1C	0.553	2 out of 2
SCFP_8	1792377291	 [*]N[c]1:[cH]:[*]:[cH]:[cH]:[c]:1C	0.553	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	1420330831	 [*]=C1N[c]2:[cH]:[cH]:[*]:[cH]:[c]:2C1=[*]	-0.31	0 out of 1
SCFP_8	399659969	 [*]CN1C(=[*])[*][*]C1=[*]	-0.31	0 out of 1
SCFP_8	-2056718782	 [*]=C1[*][*][c](:[*]):[c]1:[cH]:[*]	-0.29	6 out of 23

Compound 13

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



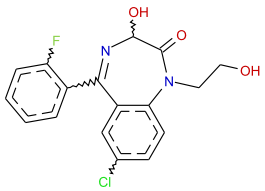
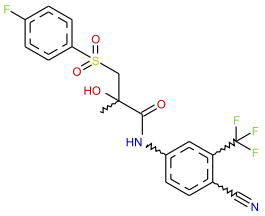
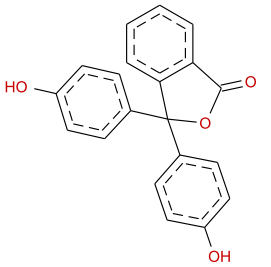
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.557
Enrichment: 1.34
Bayesian Score: 3.43
Mahalanobis Distance: 12.1
Mahalanobis Distance p-value: 0.004

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

Structural Similar Compounds

Name	Doxefazepam	Bicalutamide	Phenolphthalein
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.634	0.667	0.729
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

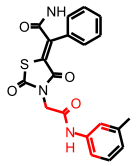
Model Applicability

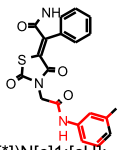
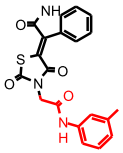
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC18 out of range. Value: -3.4903. Training min, max, SD, explained variance: -2.5092, 3.7383, 1.072, 0.0156.

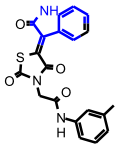
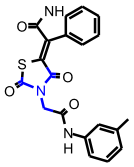
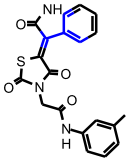
Feature Contribution

Top features for positive contribution

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

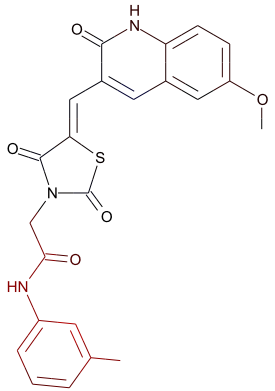
SCFP_8	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.574	4 out of 5
SCFP_8	-236487363	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.553	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	1420330831	 [*]=C1N[c]2:[cH]:[cH]:[*]:[cH]:[c]:2C1=[*]	-0.31	0 out of 1
SCFP_8	399659969	 [*]CN1C(=[*])[*][*]C1=[*]	-0.31	0 out of 1
SCFP_8	-2056718782	 [*]=C1[*][*][c](:[*]):[c]1:[cH]:[*]	-0.29	6 out of 23

Compound 14

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.583

Enrichment: 1.41

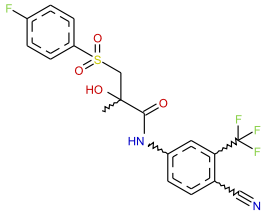
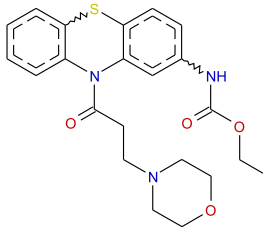
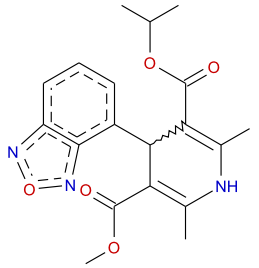
Bayesian Score: 5.03

Mahalanobis Distance: 17.5

Mahalanobis Distance p-value: 5.37e-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	Bicalutamide	Moricizine	Isradipine
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.568	0.676	0.729
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

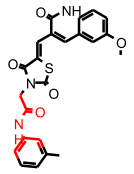
Model Applicability

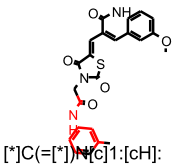
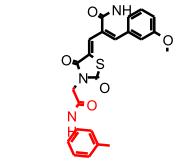
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC4 out of range. Value: 5.3049. Training min, max, SD, explained variance: -6.1092, 5.1042, 2.173, 0.0642.

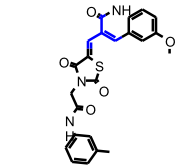
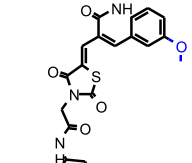
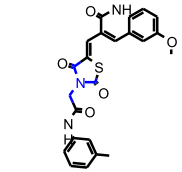
Feature Contribution

Top features for positive contribution

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

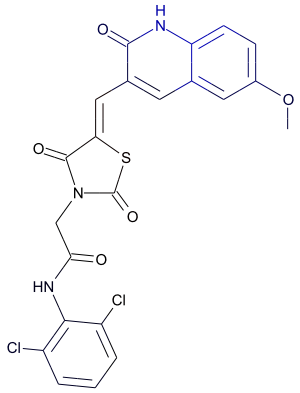
SCFP_8	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	0.574	4 out of 5
SCFP_8	-236487363	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.553	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	1798274711	 [*]\C=C(\C=[*])/C(=[*])[*]	-0.546	0 out of 2
SCFP_8	136239834	 [*]OC	-0.358	3 out of 13
SCFP_8	399659969	 [*]CN1C(=[*])[*]C1=[*]	-0.31	0 out of 1

Compound 8

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.211

Enrichment: 0.229

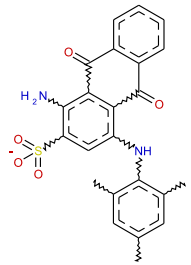
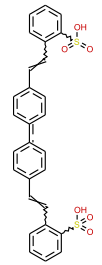
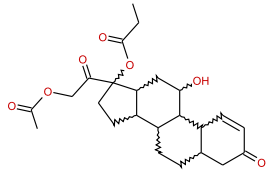
Bayesian Score: -5.37

Mahalanobis Distance: 13.5

Mahalanobis Distance p-value: 4.1e-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	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	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)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.715	0.724	0.820
Reference	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	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2-5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1-1972- Volume(issue) /page/year: 19,3103,1991

Model Applicability

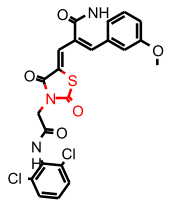
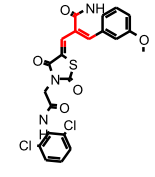
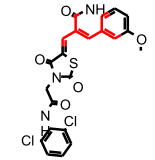
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

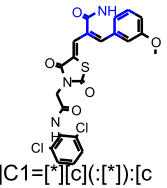
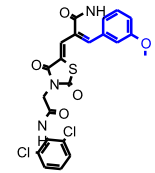
Feature Contribution

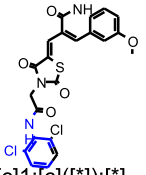
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]N1[*][*]SC1=O	0.0821	13 out of 13
FCFP_12	436886043	 [*]\C=C(\C=[*])/C(=[*])[*]	0.0804	129 out of 130
FCFP_12	1383817444	 [*]=CC1=C[c](:[cH]:[*]):[c](:[*])[*]C1=[*]	0.0772	7 out of 7

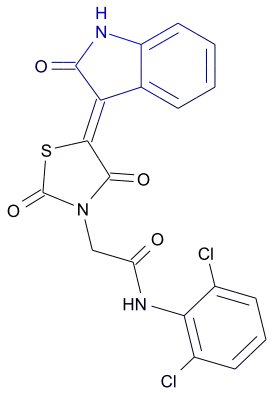
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.02	2 out of 8
FCFP_12	-1757681964	 [*][c]1:[cH]:[cH]:[c] (OC):[cH]:[c]:1C=[*]	-0.627	1 out of 3

FCFP_12	1783756416	 [*]N[c]1:[c]([*]):[*] :[cH]:[cH]:[c]:1Cl	-0.509	4 out of 8

Compound 9

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.214

Enrichment: 0.232

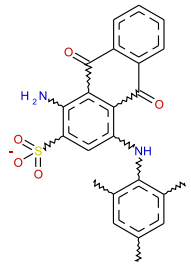
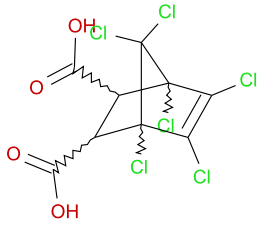
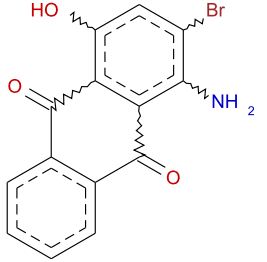
Bayesian Score: -5.36

Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.000313

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

Structural Similar Compounds

Name	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-2-bromo-4-hydroxyanthraquinone
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.695	0.790	0.822
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72

Model Applicability

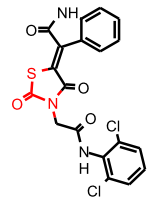
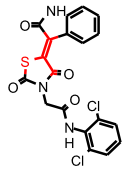
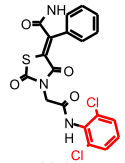
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

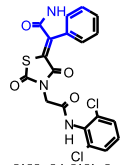
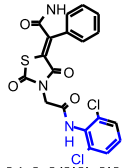
Feature Contribution

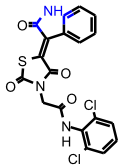
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

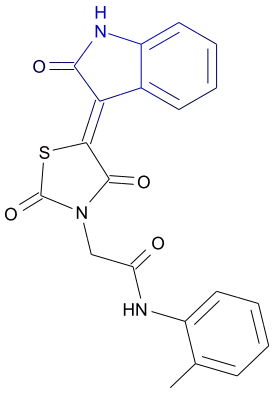
FCFP_12	-1986158408	 [*]N1[*][*]SC1=O	0.0821	13 out of 13
FCFP_12	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.0756	6 out of 6
FCFP_12	1161767339	 [*][c]1:[c](Cl):[cH]: [cH]:[cH]:[c]:1Cl	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.02	2 out of 8
FCFP_12	1783756416	 [*]N[c]1:[c]([*]):[*] :[cH]:[cH]:[c]:1Cl	-0.509	4 out of 8

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

Compound 12



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

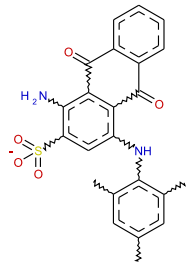
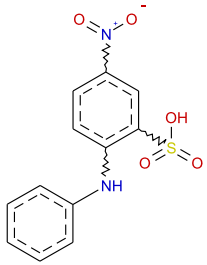
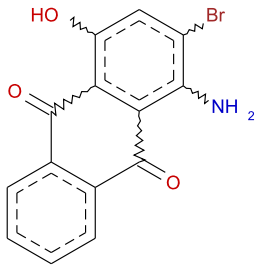
Model Prediction

Prediction: Non-Irritant
Probability: 0.458
Enrichment: 0.498
Bayesian Score: -4.66
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.00128

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

Structural Similar Compounds

Name	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	Benzenesulfonic acid, 2-anilino-5-nitro-	1-Amino-2-bromo-4-hydroxyanthraquinone
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.685	0.729	0.768
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1061,1986	28ZPAK -,83,72

Model Applicability

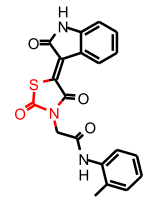
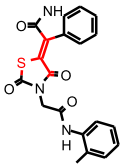
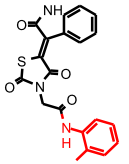
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

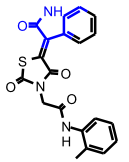
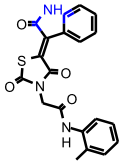
Feature Contribution

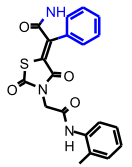
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

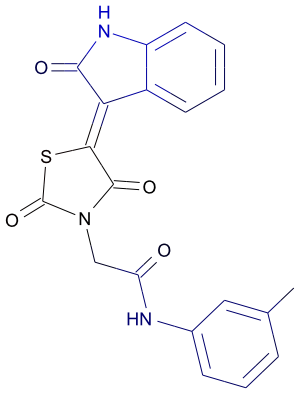
FCFP_12	-1986158408	 [*]N1[*][*]SC1=O	0.0821	13 out of 13
FCFP_12	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.0756	6 out of 6
FCFP_12	1396506317	 [*]N[c]1:[cH]:[cH]:[c H]:[cH]:[c]:1C	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.02	2 out of 8
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-0.486	12 out of 22

FCFP_12	-1724769936	 [*]1[*][c]2:[cH]:[cH] :[cH]:[cH]:[c]:2N1	-0.475	11 out of 20

Compound 13



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

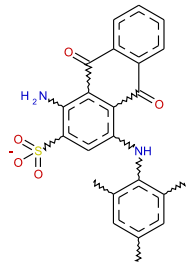
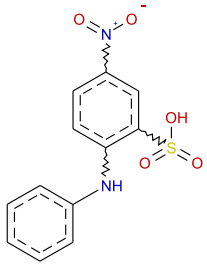
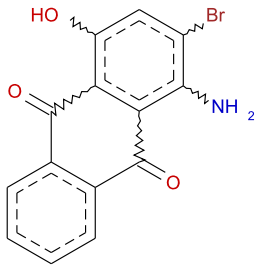
Model Prediction

Prediction: Non-Irritant
Probability: 0.0215
Enrichment: 0.0234
Bayesian Score: -6.7
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.00128

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

Structural Similar Compounds

Name	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	Benzenesulfonic acid, 2-anilino-5-nitro-	1-Amino-2-bromo-4-hydroxyanthraquinone
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.683	0.731	0.770
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1061,1986	28ZPAK -,83,72

Model Applicability

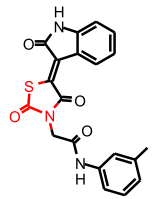
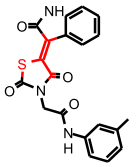
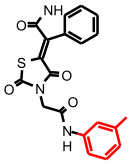
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

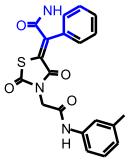
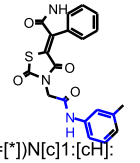
Feature Contribution

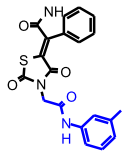
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]N1[*][*]SC1=O	0.0821	13 out of 13
FCFP_12	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.0756	6 out of 6
FCFP_12	630418361	 [*][c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.0734	5 out of 5

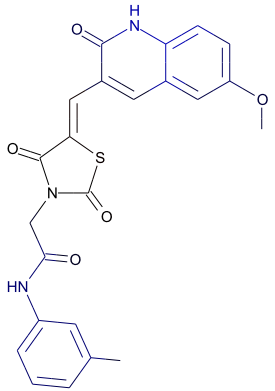
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.02	2 out of 8
FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]: [cH]:[*]:[c]([*]):[c H]:1	-0.692	5 out of 12

FCFP_12	-453277354	 [*]CC(=O)N[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	-0.65	0 out of 1

Compound 14

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.0182

Enrichment: 0.0198

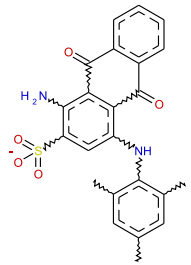
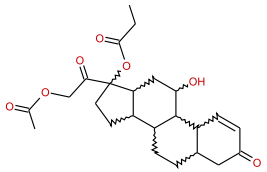
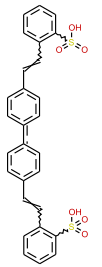
Bayesian Score: -6.78

Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 3.79e-008

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

Structural Similar Compounds

Name	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.615	0.765	0.824
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2-5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1-1972- Volume(issue) /page/year: 19,3103,1991	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973

Model Applicability

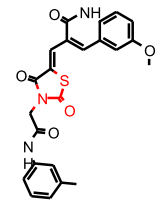
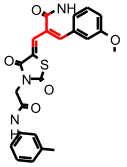
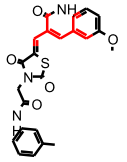
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

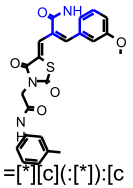
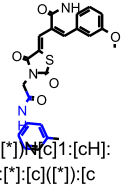
Feature Contribution

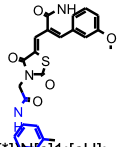
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]N1[*][*]SC1=O	0.0821	13 out of 13
FCFP_12	436886043	 [*]\C=C(\C=[*])/C(=[*])[*]	0.0804	129 out of 130
FCFP_12	1383817444	 [*]=CC1=C[c](:[cH]:[*]):[c](:[*])[*]C1=[*]	0.0772	7 out of 7

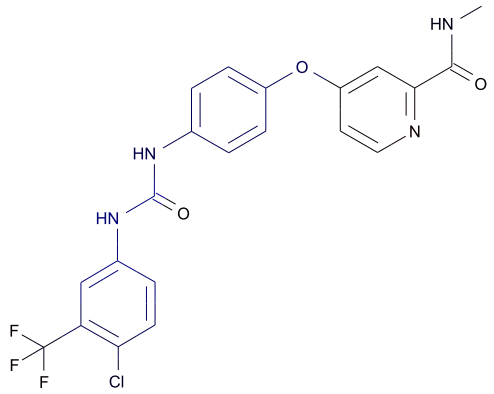
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944	 [*]C1=[*][c](:[*]):[c](NC1=O):[cH]:[*]	-1.02	2 out of 8
FCFP_12	-1838187238	 [*]C(=[*])[*][c]1:[cH]: [cH]:[*]:[c]([*]):[c H]:1	-0.692	5 out of 12

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

Sorafenib

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Irritant

Probability: 0.264

Enrichment: 0.287

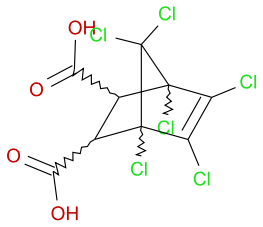
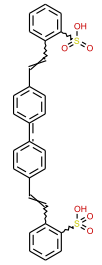
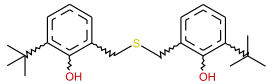
Bayesian Score: -5.23

Mahalanobis Distance: 8.27

Mahalanobis Distance p-value: 0.791

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

Structural Similar Compounds

Name	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	Sulfide, bis(4-t-butyl-m-cresyl)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.844	0.871	0.884
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	AMIHBC AMA Archives of Industrial Hygiene and Occupational Medicine. (Chicago , IL) V.2-10, 1950-54. For publisher information, see AEHLAU. Volume(issue)/pag e/year: 5,311,1952

Model Applicability

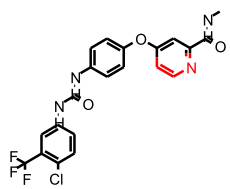
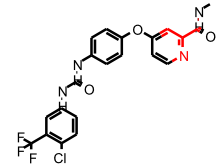
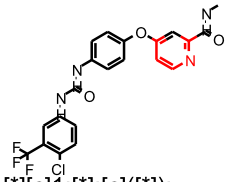
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

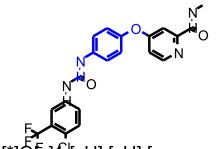
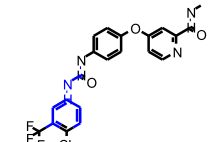
Feature Contribution

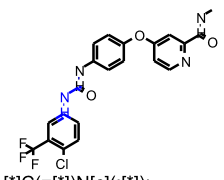
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

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

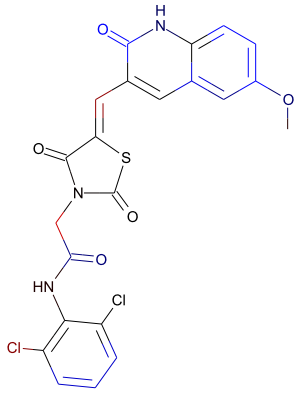
Top Features for negative contribution

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

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

Compound 8

TOPKAT_Carcinogenic_Potency_TD50_Mouse



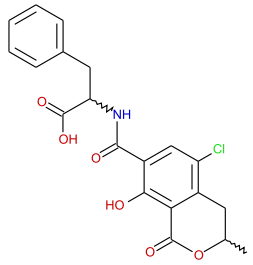
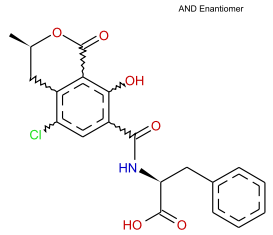
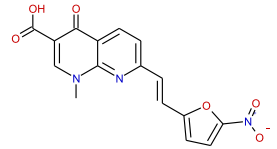
$C_{22}H_{15}Cl_2N_3O_5S$
Molecular Weight: 504.34259
ALogP: 3.479
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 26.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15
Mahalanobis Distance p-value: 4.1e-013

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

Structural Similar Compounds

Name	Ochratoxin A	542	470
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	4.62839
Predicted Endpoint (-log C)	3.6353	3.6353	3.93264
Distance	0.660	0.660	0.747
Reference	CPDB	CPDB	CPDB

Model Applicability

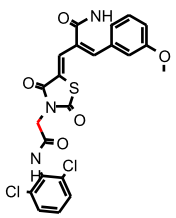
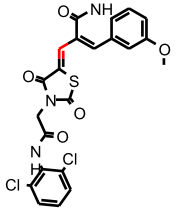
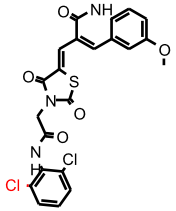
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
3. Unknown ECFP_2 feature: 1182722866: [*]C(=CC(=[*])[*])[*]
4. Unknown ECFP_2 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
5. Unknown ECFP_2 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
6. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]
7. Unknown ECFP_2 feature: -1660205591: [*]N[c](:[c]([*]):[*]):[c]([*]):[*]

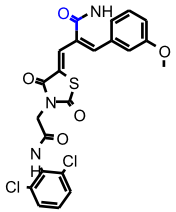
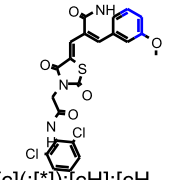
Feature Contribution

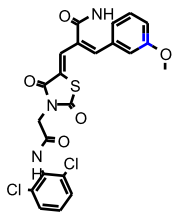
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	1559650422	 [*]C[*]	0.203
ECFP_6	-1925046727	 [*]C=[*]	0.145
ECFP_6	-817402818	 [*]Cl	0.129

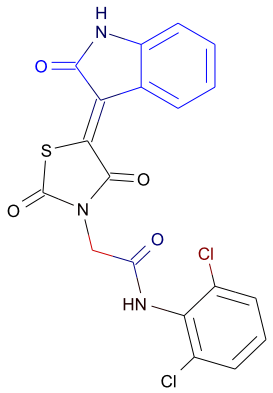
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251

ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

Compound 9

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 36

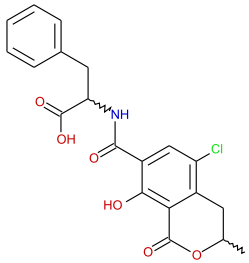
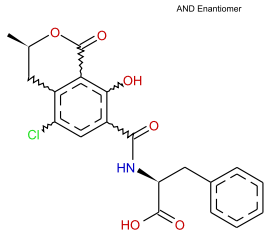
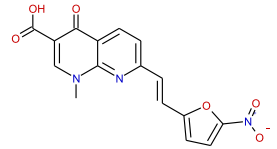
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 12.9

Mahalanobis Distance p-value: 2.06e-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 A	542	470
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	4.62839
Predicted Endpoint (-log C)	3.6353	3.6353	3.93264
Distance	0.659	0.659	0.729
Reference	CPDB	CPDB	CPDB

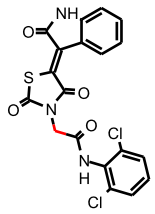
Model Applicability

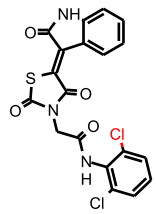
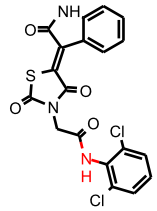
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
4. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]
5. Unknown ECFP_2 feature: -1660205591: [*]N[c](:[c]([*]):[*]):[c]([*]):[*]

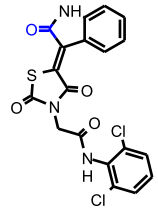
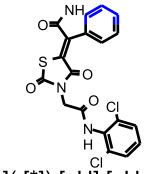
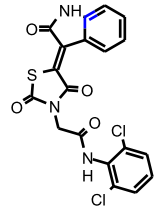
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

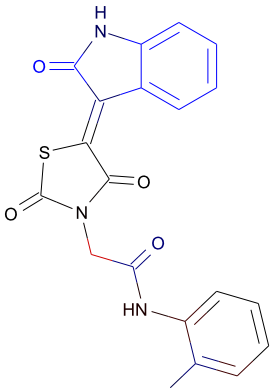
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-1897341097	 [*]N[*]	0.0284

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

Compound 12

TOPKAT_Carcinogenic_Potency_TD50_Mouse



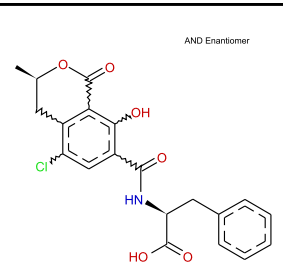
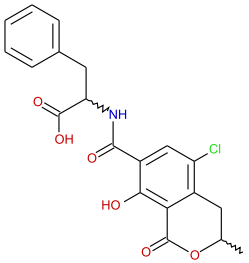
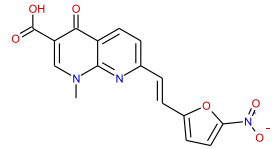
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 79.3
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 1.46e-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	542	Ochratoxin A	470
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	4.62839
Predicted Endpoint (-log C)	3.6353	3.6353	3.93264
Distance	0.674	0.674	0.697
Reference	CPDB	CPDB	CPDB

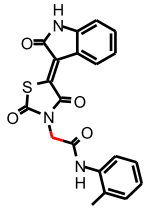
Model Applicability

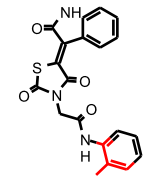
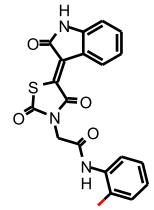
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC22 out of range. Value: 4.0663. Training min, max, SD, explained variance: -3.1587, 3.8589, 1.086, 0.0147.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
4. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]

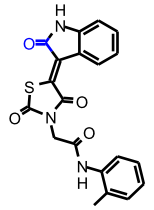
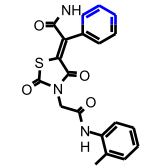
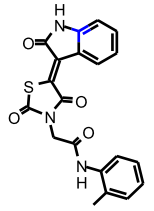
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

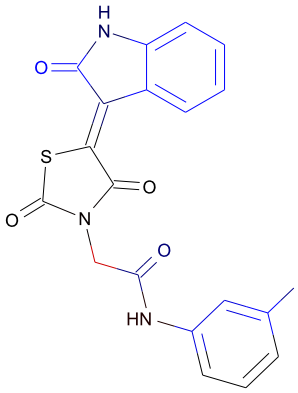
ECFP_6	2147419938	 [*][c](:[*]):[c](C):[cH]:[*]	0.0637
ECFP_6	734603939	 [*]C	0.0424

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

Compound 13

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{20}H_{15}N_3O_4S$

Molecular Weight: 393.4158

ALogP: 2.185

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 151

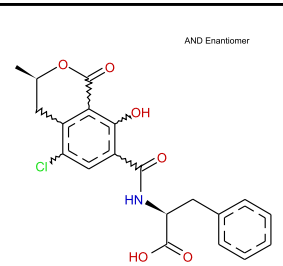
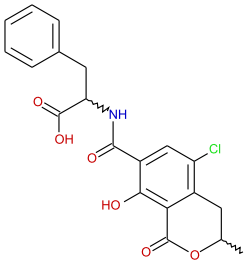
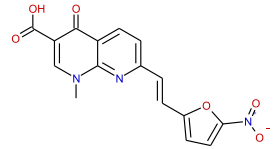
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13

Mahalanobis Distance p-value: 1.12e-007

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

Structural Similar Compounds

Name	542	Ochratoxin A	470
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	4.62839
Predicted Endpoint (-log C)	3.6353	3.6353	3.93264
Distance	0.671	0.671	0.698
Reference	CPDB	CPDB	CPDB

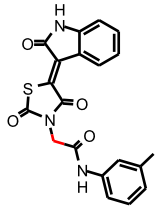
Model Applicability

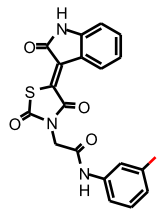
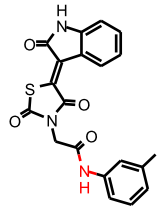
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC22 out of range. Value: 3.938. Training min, max, SD, explained variance: -3.1587, 3.8589, 1.086, 0.0147.
2. Unknown ECFP_2 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
3. Unknown ECFP_2 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
4. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]

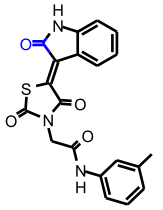
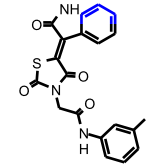
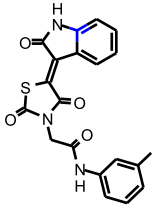
Feature Contribution

Top features for positive contribution

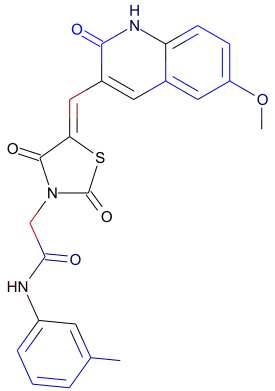
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

ECFP_6	734603939	 [*]C	0.0424
ECFP_6	-1897341097	 [*]N[*]	0.0284

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

Compound 14



$C_{23}H_{19}N_3O_5S$
Molecular Weight: 449.47905
ALogP: 2.637
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

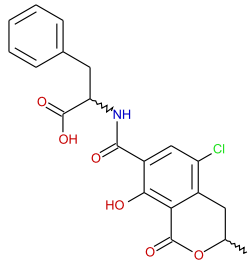
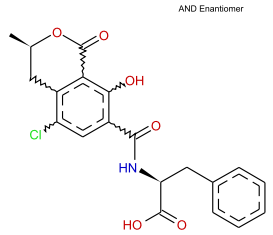
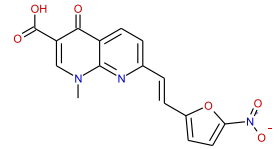
Model Prediction

Prediction: 75.9
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15.2
Mahalanobis Distance p-value: 8.19e-014

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly 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	470
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	4.62839
Predicted Endpoint (-log C)	3.6353	3.6353	3.93264
Distance	0.648	0.648	0.695
Reference	CPDB	CPDB	CPDB

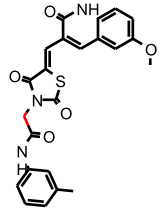
Model Applicability

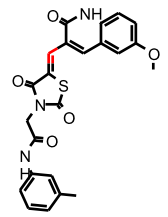
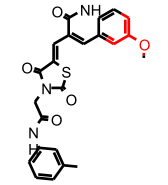
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
3. Unknown ECFP_2 feature: 1182722866: [*]C(=CC(=[*])[*])[*]
4. Unknown ECFP_2 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
5. Unknown ECFP_2 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
6. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]

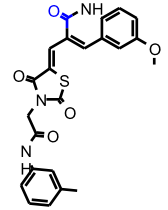
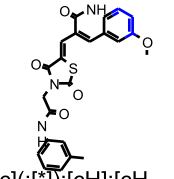
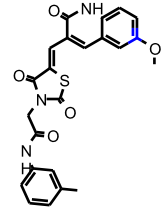
Feature Contribution

Top features for positive contribution

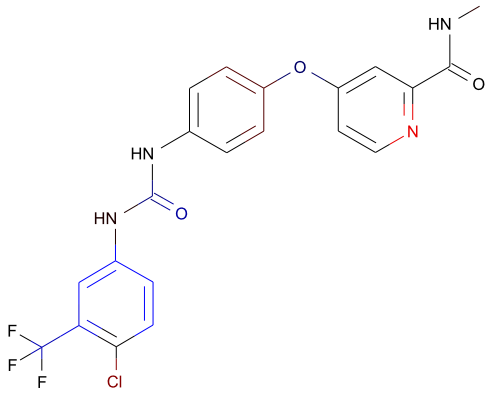
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.203

ECFP_6	-1925046727	 [*]C=[*]	0.145
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.0818

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.275
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 19.2

Unit: mg/kg_body_weight/day

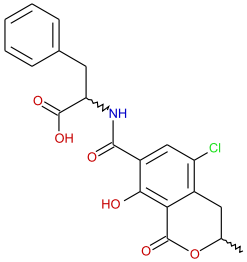
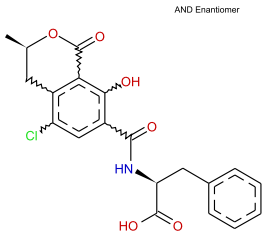
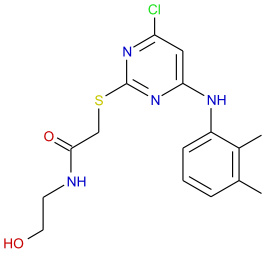
Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 2.94e-006

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse

Structural Similar Compounds

Name	Ochratoxin A	542	4-Chloro-6-(2,3-xylidino)-2-pyri-mi-dinylthio(N-b-hydroxy-ethyl) acetamide
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	3.91517
Predicted Endpoint (-log C)	3.6353	3.6353	3.92186
Distance	0.718	0.718	0.738
Reference	CPDB	CPDB	CPDB

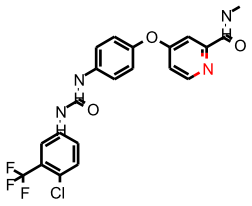
Model Applicability

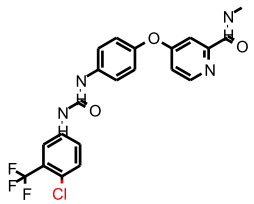
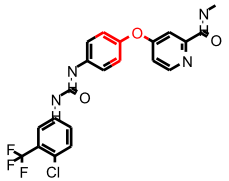
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1338334141: [*]C(=[*])NC
3. Unknown ECFP_2 feature: 1413420509: [*]C(=[*])[c](:n:[*]):c:[*]

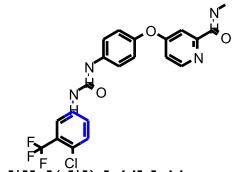
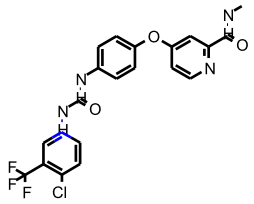
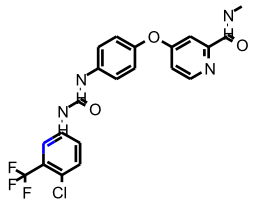
Feature Contribution

Top features for positive contribution

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

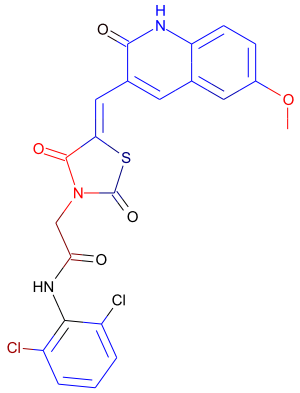
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.0818

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1996767644	 [*][c](:[*]):[cH]:[cH]:[*]	-0.251
ECFP_6	642810091	 [*][c](:[*]):[*]	-0.247
ECFP_6	-182236392	 [*]:[cH]:[*]	-0.232

Compound 8

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 35.9

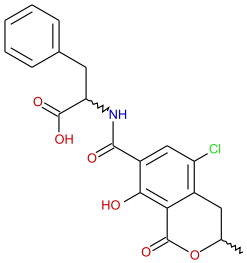
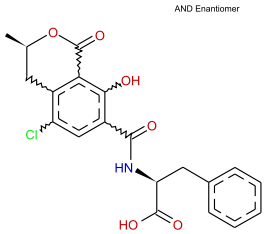
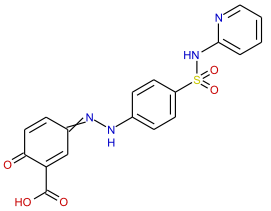
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 16

Mahalanobis Distance p-value: 2.86e-014

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

Structural Similar Compounds

Name	Ochratoxin A	542	Salicylazosulfapyridine
Structure			
Actual Endpoint (-log C)	6.47264	6.59334	2.39891
Predicted Endpoint (-log C)	5.06501	5.06501	3.17598
Distance	0.660	0.660	0.713
Reference	CPDB	CPDB	CPDB

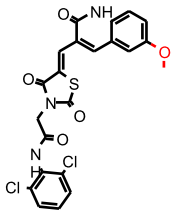
Model Applicability

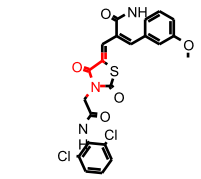
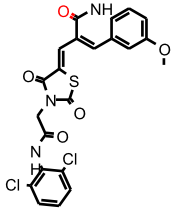
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC7 out of range. Value: -6.0099. Training min, max, SD, explained variance: -5.0422, 6.1749, 1.868, 0.0335.
2. OPS PC16 out of range. Value: -3.6579. Training min, max, SD, explained variance: -3.6111, 5.345, 1.406, 0.0190.

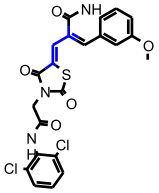
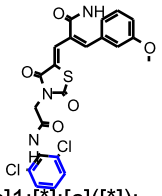
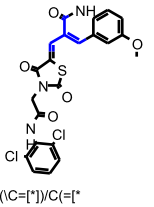
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

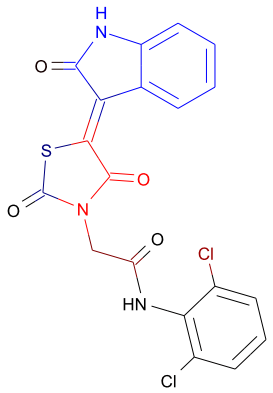
FCFP_6	565998553	 [*]N1[*][*]C(=[*])C1= O	0.357
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	-0.436
FCFP_6	991735244	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.422
FCFP_6	436886043	 [*]\C=C(\C=[*])/C(=[*])[*]	-0.383

Compound 9

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 13

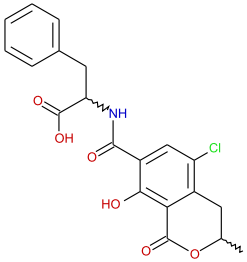
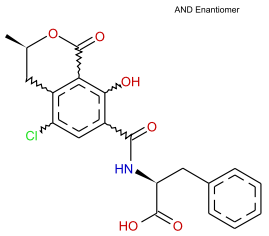
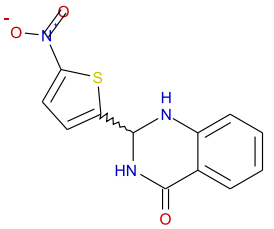
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.00327

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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 A	542	1,2-Dihydro-2-(5-nitro-2-thi-enyl) quinazolin-4(3H)-one
Structure			
Actual Endpoint (-log C)	6.47264	6.59334	5.25509
Predicted Endpoint (-log C)	5.06501	5.06501	3.89291
Distance	0.625	0.625	0.685
Reference	CPDB	CPDB	CPDB

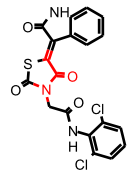
Model Applicability

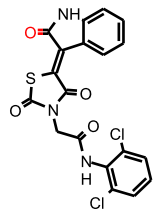
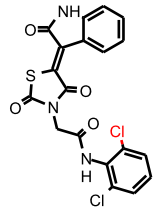
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

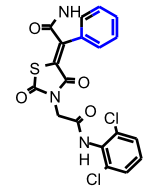
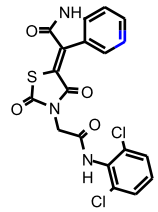
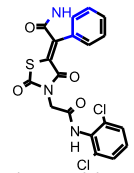
Feature Contribution

Top features for positive contribution

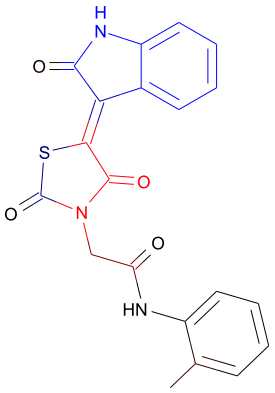
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]N1[*][*]C(=[*])C1= O	0.357

FCFP_6	1	 [*]=O	0.234
FCFP_6	32	 [*]Cl	0.154

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[c]([*]):[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.323

Compound 12



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

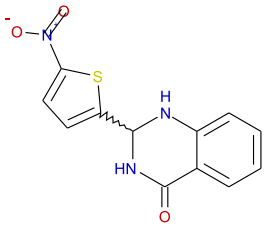
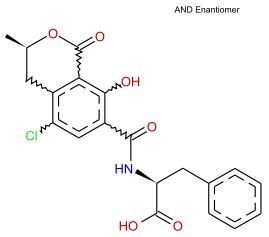
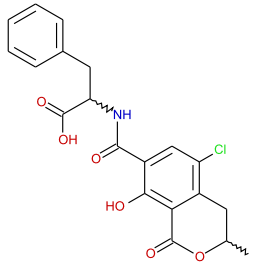
Model Prediction

Prediction: 18.3
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.0145

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	1,2-Dihydro-2-(5-nitro-2-thi-enyl) quinazolin-4(3H)-one	542	Ochratoxin A
Structure			
Actual Endpoint (-log C)	5.25509	6.59334	6.47264
Predicted Endpoint (-log C)	3.89291	5.06501	5.06501
Distance	0.634	0.645	0.645
Reference	CPDB	CPDB	CPDB

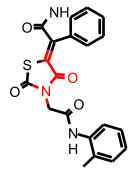
Model Applicability

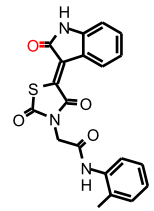
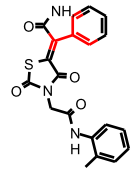
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

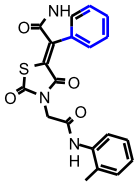
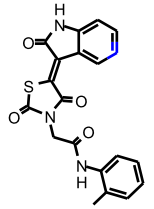
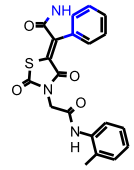
Feature Contribution

Top features for positive contribution

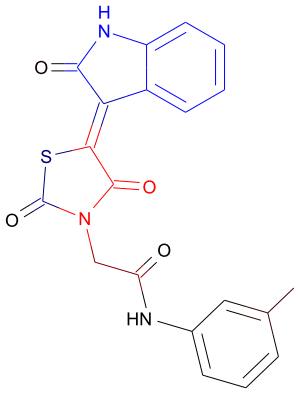
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]N1[*][*]C(=[*])C1= O	0.357

FCFP_6	1	 [*]=O	0.234
FCFP_6	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.323

Compound 13



$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

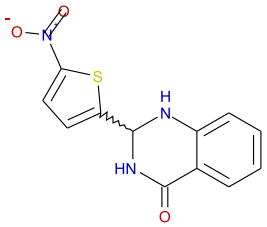
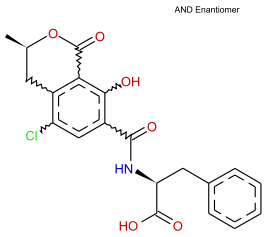
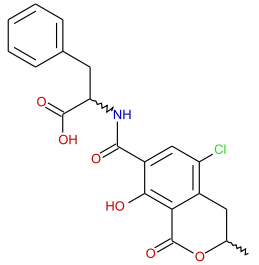
Model Prediction

Prediction: 18.3
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.0145

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	1,2-Dihydro-2-(5-nitro-2-thi-enyl) quinazolin-4(3H)-one	542	Ochratoxin A
Structure			
Actual Endpoint (-log C)	5.25509	6.59334	6.47264
Predicted Endpoint (-log C)	3.89291	5.06501	5.06501
Distance	0.639	0.648	0.648
Reference	CPDB	CPDB	CPDB

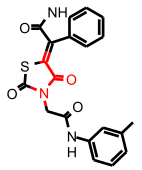
Model Applicability

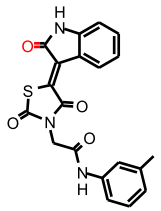
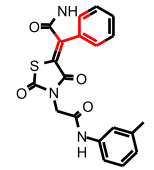
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

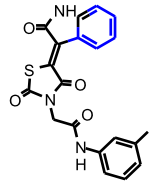
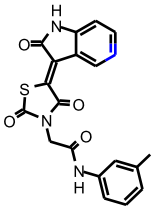
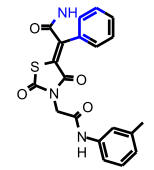
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]N1[*][*]C(=[*])C1=O	0.357

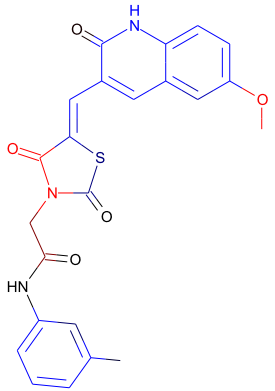
FCFP_6	1	 [*]=O	0.234
FCFP_6	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.323

Compound 14

TOPKAT_Carcinogenic_Potency_TD50_Rat



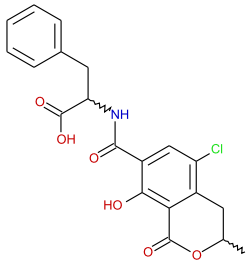
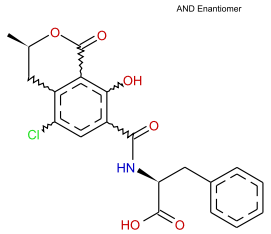
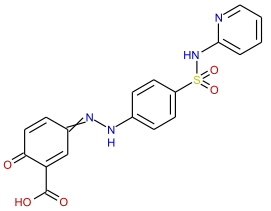
$C_{23}H_{19}N_3O_5S$
Molecular Weight: 449.47905
ALogP: 2.637
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 64.9
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15.4
Mahalanobis Distance p-value: 2.3e-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	Ochratoxin A	542	Salicylazosulfapyridine
Structure			
Actual Endpoint (-log C)	6.47264	6.59334	2.39891
Predicted Endpoint (-log C)	5.06501	5.06501	3.17598
Distance	0.647	0.647	0.671
Reference	CPDB	CPDB	CPDB

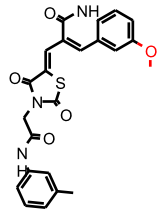
Model Applicability

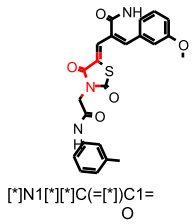
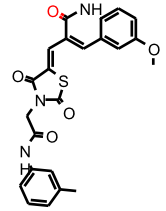
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC7 out of range. Value: -5.6632. Training min, max, SD, explained variance: -5.0422, 6.1749, 1.868, 0.0335.

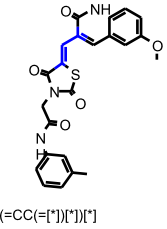
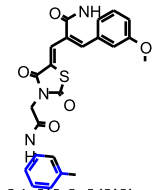
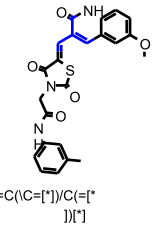
Feature Contribution

Top features for positive contribution

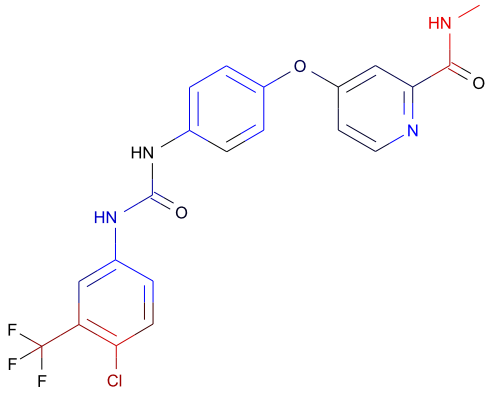
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]OC	0.69

FCFP_6	565998553		0.357
FCFP_6	1		0.234

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724		-0.436
FCFP_6	991735244		-0.422
FCFP_6	436886043		-0.383

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 14.2

Unit: mg/kg_body_weight/day

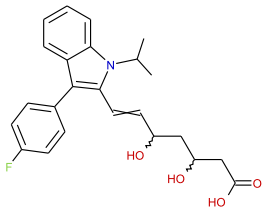
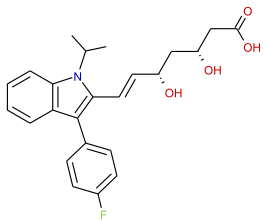
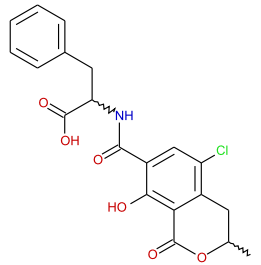
Mahalanobis Distance: 20.4

Mahalanobis Distance p-value: 9.56e-031

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

TOPKAT_Carcinogenic_Potency_TD50_Rat

Structural Similar Compounds

Name	Fluvastatin	913	Ochratoxin A
Structure			
Actual Endpoint (-log C)	3.51742	3.51742	6.47264
Predicted Endpoint (-log C)	5.41573	5.41573	5.06501
Distance	0.597	0.597	0.666
Reference	CPDB	CPDB	CPDB

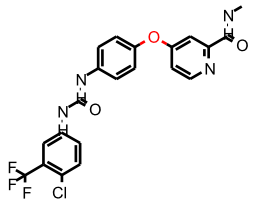
Model Applicability

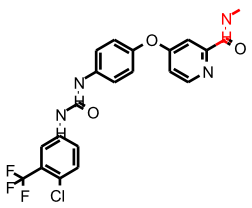
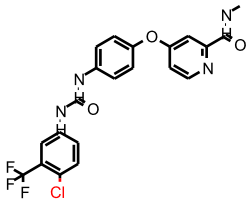
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1029533685: [*]:[c](:[*])C(F)(F)F

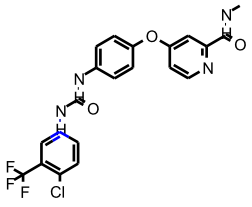
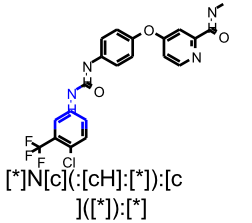
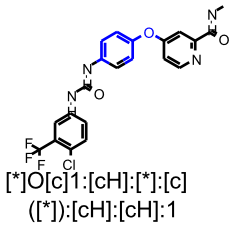
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	0.234

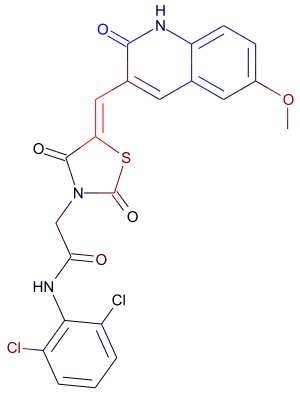
FCFP_6	-885550502	 [*]C(=[*])NC	0.229
FCFP_6	32	 [*]Cl	0.154

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	 [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[c]([*]):[*]	-0.323
FCFP_6	1674451008	 [*]O[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	-0.233

Compound 8

TOPKAT_Chronic_LOAEL



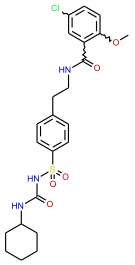
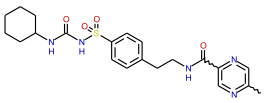
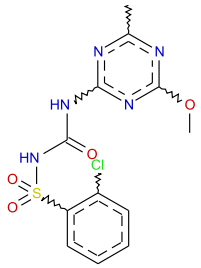
$C_{22}H_{15}Cl_2N_3O_5S$
Molecular Weight: 504.34259
ALogP: 3.479
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.00201
Unit: g/kg_body_weight
Mahalanobis Distance: 33.6
Mahalanobis Distance p-value: 3.79e-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.

Structural Similar Compounds

Name	GLYBURIDE	GLIPIZIDE	CHLORSULFURON
Structure			
Actual Endpoint (-log C)	4.21661	3.94991	4.15566
Predicted Endpoint (-log C)	4.21035	3.95594	3.79771
Distance	0.596	0.645	0.698
Reference	UPJ-26452	NDA-17583	EPA COVER SHEET 0027;880301;(1)

Model Applicability

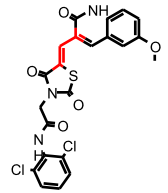
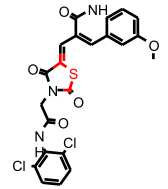
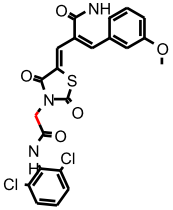
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC22 out of range. Value: -6.8924. Training min, max, SD, explained variance: -4.3287, 5.3383, 1.588, 0.0110.
2. Unknown ECFP_6 feature: -154530762: [*]N[*]
3. Unknown ECFP_6 feature: 912478223: [*]S[*]
4. Unknown ECFP_6 feature: 1335833675: [*]N[c](:[cH]:[*]):[c]([*]):[*]
5. Unknown ECFP_6 feature: 1336666212: [*][c](:[*]):[c](C=[*]):[cH]:[*]
6. Unknown ECFP_6 feature: -1699286547: [*]C(=[*])N[c](:[*]):[*]
7. Unknown ECFP_6 feature: 1298725959: [*]NC(=O)C(=[*])[*]
8. Unknown ECFP_6 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
9. Unknown ECFP_6 feature: 464808839: [*]C(=C[c](:[*]):[*])[*]
10. Unknown ECFP_6 feature: 1182722866: [*]C(=CC(=[*])[*])[*]
11. Unknown ECFP_6 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
12. Unknown ECFP_6 feature: 1945129186: [*]N1[*][*]C(=[*])C1=O
13. Unknown ECFP_6 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
14. Unknown ECFP_6 feature: 190445529: [*]N1[*][*]SC1=O
15. Unknown ECFP_6 feature: 2122741631: [*]=C1[*][*]C(=[*])S1
16. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
17. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
18. Unknown ECFP_6 feature: -1660205591: [*]N[c](:[c]([*]):[*]):[c]([*]):[*]

19. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
20. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
21. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

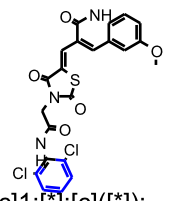
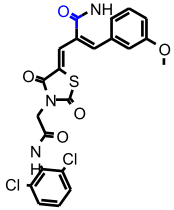
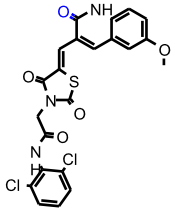
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.16
FCFP_6	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.13
ECFP_6	1559650422	 [*]C[*]	0.129

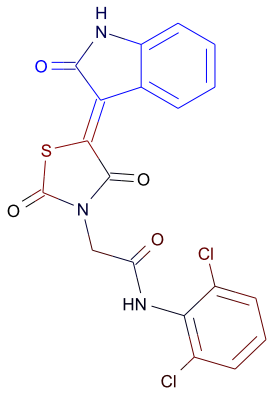
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	991735244	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.134
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11
FCFP_6	1	 [*]=O	-0.102

Compound 9

TOPKAT_Chronic_LOAEL



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 0.0239

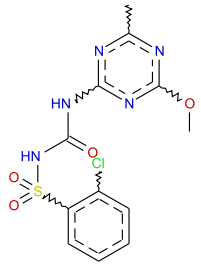
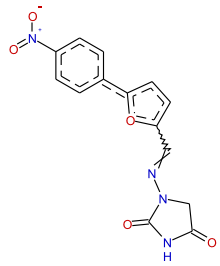
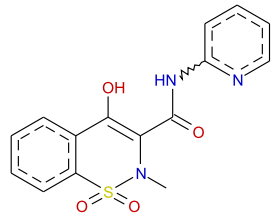
Unit: g/kg_body_weight

Mahalanobis Distance: 30.5

Mahalanobis Distance p-value: 1.3e-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	CHLORSULFURON	DANTROLENE.NA	PIROXICAM
Structure			
Actual Endpoint (-log C)	4.15566	4.19625	5.52028
Predicted Endpoint (-log C)	3.79771	4.62637	4.06087
Distance	0.630	0.655	0.666
Reference	EPA COVER SHEET 0027;880301;(1)	NDA-17443	NDA-18147

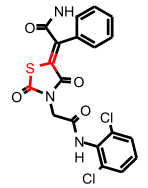
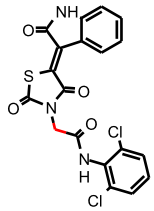
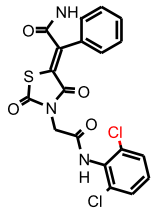
Model Applicability

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

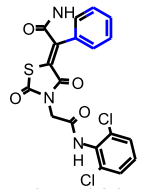
1. OPS PC22 out of range. Value: -5.1197. Training min, max, SD, explained variance: -4.3287, 5.3383, 1.588, 0.0110.
2. Unknown ECFP_6 feature: -154530762: [*]N[*]
3. Unknown ECFP_6 feature: 912478223: [*]S[*]
4. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
5. Unknown ECFP_6 feature: 1335833675: [*]N[c](:[cH]:[*]):[c]([*]):[*]
6. Unknown ECFP_6 feature: -1699286547: [*]C(=[*])N[c](:[*]):[*]
7. Unknown ECFP_6 feature: 1298725959: [*]NC(=O)C(=[*])[*]
8. Unknown ECFP_6 feature: 1790105651: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]
9. Unknown ECFP_6 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
10. Unknown ECFP_6 feature: 2122741631: [*]=C1[*][*]C(=[*])S1
11. Unknown ECFP_6 feature: 190445529: [*]N1[*][*]SC1=O
12. Unknown ECFP_6 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
13. Unknown ECFP_6 feature: 1945129186: [*]N1[*][*]C(=[*])C1=O
14. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
15. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
16. Unknown ECFP_6 feature: -1660205591: [*]N[c](:[c]([*]):[*]):[c]([*]):[*]
17. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

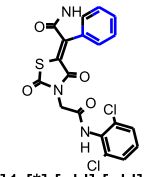
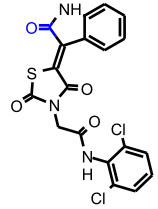
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.13
ECFP_6	1559650422	 [*]C[*]	0.129
FCFP_6	32	 [*]Cl	0.101

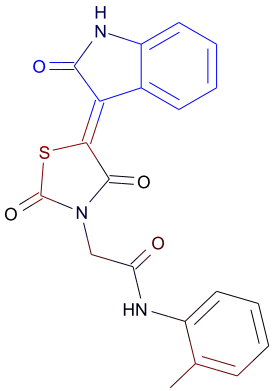
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.134

ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11

Compound 12

TOPKAT_Chronic_LOAEL



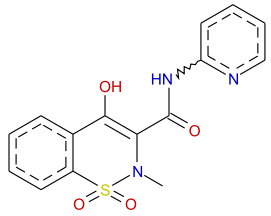
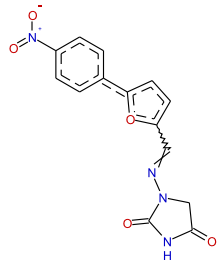
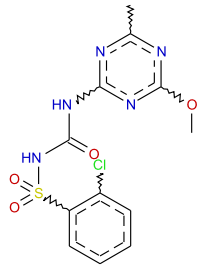
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.0364
Unit: g/kg_body_weight
Mahalanobis Distance: 29.6
Mahalanobis Distance p-value: 7.33e-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

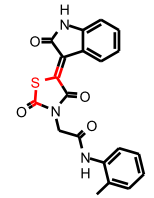
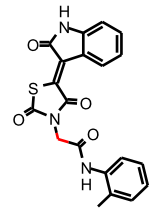
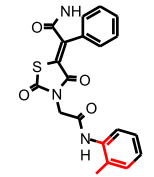
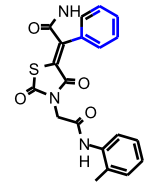
Name	PIROXICAM	DANTROLENE.NA	CHLORSULFURON
Structure			
Actual Endpoint (-log C)	5.52028	4.19625	4.15566
Predicted Endpoint (-log C)	4.06087	4.62637	3.79771
Distance	0.564	0.578	0.586
Reference	NDA-18147	NDA-17443	EPA COVER SHEET 0027;880301;(1)

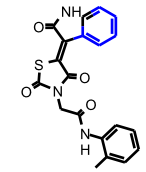
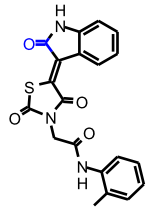
Model Applicability

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

1. OPS PC22 out of range. Value: -5.1654. Training min, max, SD, explained variance: -4.3287, 5.3383, 1.588, 0.0110.
2. Unknown ECFP_6 feature: -154530762: [*]N[*]
3. Unknown ECFP_6 feature: 912478223: [*]S[*]
4. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
5. Unknown ECFP_6 feature: 1335833675: [*]N[c](:[cH]:[*]):[c]([*]):[*]
6. Unknown ECFP_6 feature: -1699286547: [*]C(=[*])N[c](:[*]):[*]
7. Unknown ECFP_6 feature: 1298725959: [*]NC(=O)C(=[*])[*]
8. Unknown ECFP_6 feature: 1790105651: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]
9. Unknown ECFP_6 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
10. Unknown ECFP_6 feature: 2122741631: [*]=C1[*][*]C(=[*])S1
11. Unknown ECFP_6 feature: 190445529: [*]N1[*][*]SC1=O
12. Unknown ECFP_6 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
13. Unknown ECFP_6 feature: 1945129186: [*]N1[*][*]C(=[*])C1=O
14. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
15. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
16. Unknown ECFP_6 feature: 1335108269: [*]N[c](:[cH]:[*]):[c]([*]):[*]

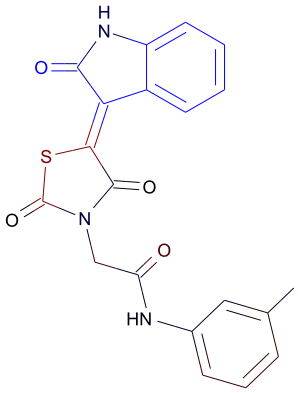
Feature Contribution

Top features for positive contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.13
ECFP_6	1559650422	 [*]C[*]	0.129
ECFP_6	2147419938	 [*][c](:[*]):[c](C):[cH]:[*]	0.098
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	991735244	 [*][c]1:[*]:[c]([*]):[cH]:[cH]:[cH]:1	-0.134

ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11

Compound 13

TOPKAT_Chronic_LOAEL



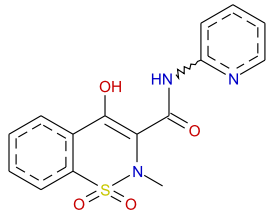
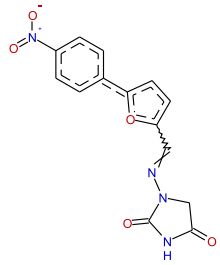
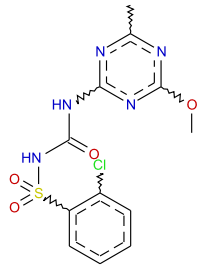
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.043
Unit: g/kg_body_weight
Mahalanobis Distance: 29.6
Mahalanobis Distance p-value: 7.33e-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	PIROXICAM	DANTROLENE.NA	CHLORSULFURON
Structure			
Actual Endpoint (-log C)	5.52028	4.19625	4.15566
Predicted Endpoint (-log C)	4.06087	4.62637	3.79771
Distance	0.566	0.579	0.587
Reference	NDA-18147	NDA-17443	EPA COVER SHEET 0027;880301;(1)

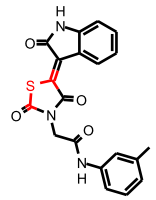
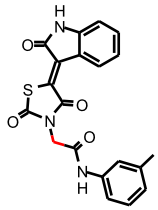
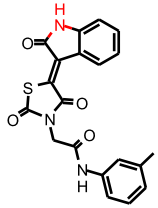
Model Applicability

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

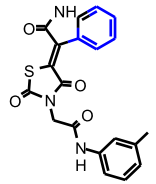
1. OPS PC22 out of range. Value: -5.1654. Training min, max, SD, explained variance: -4.3287, 5.3383, 1.588, 0.0110.
2. Unknown ECFP_6 feature: -154530762: [*]N[*]
3. Unknown ECFP_6 feature: 912478223: [*]S[*]
4. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
5. Unknown ECFP_6 feature: 1335833675: [*]N[c](:[cH]:[*]):[c]([*]):[*]
6. Unknown ECFP_6 feature: -1699286547: [*]C(=[*])N[c](:[*]):[*]
7. Unknown ECFP_6 feature: 1298725959: [*]NC(=O)C(=[*])[*]
8. Unknown ECFP_6 feature: 1790105651: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]
9. Unknown ECFP_6 feature: -631778390: [*]C(=C1S[*][*]C1=[*])[*]
10. Unknown ECFP_6 feature: 2122741631: [*]=C1[*][*]C(=[*])S1
11. Unknown ECFP_6 feature: 190445529: [*]N1[*][*]SC1=O
12. Unknown ECFP_6 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
13. Unknown ECFP_6 feature: 1945129186: [*]N1[*][*]C(=[*])C1=O
14. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
15. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
16. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
17. Unknown ECFP_6 feature: -179515162: [*]:[cH]:[c](C):[cH]:[*]

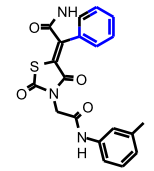
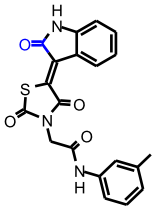
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.13
ECFP_6	1559650422	 [*]C[*]	0.129
FCFP_6	3	 [*]N[*]	0.0924

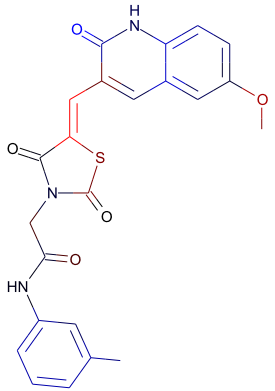
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.134

ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11

Compound 14

TOPKAT_Chronic_LOAEL



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.00426

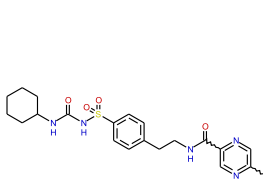
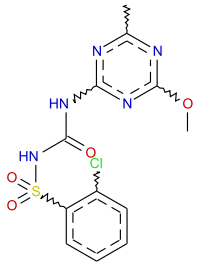
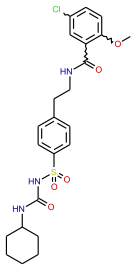
Unit: g/kg_body_weight

Mahalanobis Distance: 31.7

Mahalanobis Distance p-value: 1.04e-027

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

Structural Similar Compounds

Name	GLIPIZIDE	CHLORSULFURON	GLYBURIDE
Structure			
Actual Endpoint (-log C)	3.94991	4.15566	4.21661
Predicted Endpoint (-log C)	3.95594	3.79771	4.21035
Distance	0.590	0.632	0.645
Reference	NDA-17583	EPA COVER SHEET 0027;880301;(1)	UPJ-26452

Model Applicability

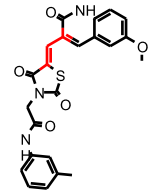
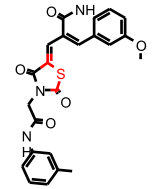
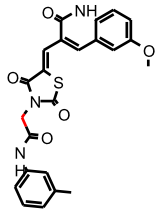
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC22 out of range. Value: -6.9718. Training min, max, SD, explained variance: -4.3287, 5.3383, 1.588, 0.0110.
2. Unknown ECFP_6 feature: -154530762: [*]N[*]
3. Unknown ECFP_6 feature: 912478223: [*]S[*]
4. Unknown ECFP_6 feature: 1335833675: [*]N[c](:[cH]:[*]):[c]([*]):[*]
5. Unknown ECFP_6 feature: 1336666212: [*][c](:[*]):[c](C=[*]):[cH]:[*]
6. Unknown ECFP_6 feature: -1699286547: [*]C(=[*])N[c](:[*]):[*]
7. Unknown ECFP_6 feature: 1298725959: [*]NC(=O)C(=[*])[*]
8. Unknown ECFP_6 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
9. Unknown ECFP_6 feature: 464808839: [*]C(=C[c](:[*]):[*])[*]
10. Unknown ECFP_6 feature: 1182722866: [*]C(=CC(=[*])[*])[*]
11. Unknown ECFP_6 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
12. Unknown ECFP_6 feature: 1945129186: [*]N1[*][*]C(=[*])C1=O
13. Unknown ECFP_6 feature: -661097313: [*]CN1C(=[*])[*][*]C1=[*]
14. Unknown ECFP_6 feature: 190445529: [*]N1[*][*]SC1=O
15. Unknown ECFP_6 feature: 2122741631: [*]=C1[*][*]C(=[*])S1
16. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
17. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
18. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]

19. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
20. Unknown ECFP_6 feature: -179515162: [*]:[cH]:[c](C):[cH]:[*]
21. Unknown ECFP_6 feature: 1307307440: [*]:[c](:[*])OC

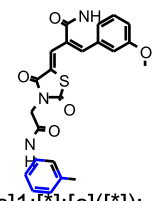
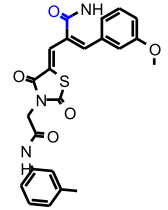
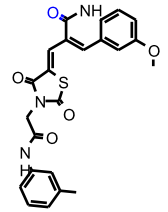
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	451847724	 [*]C(=CC(=[*])[*])[*]	0.16
FCFP_6	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.13
ECFP_6	1559650422	 [*]C[*]	0.129

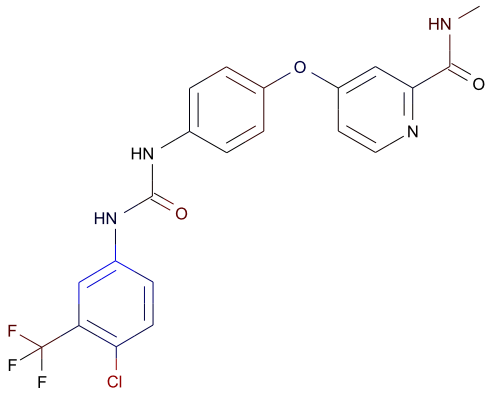
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	991735244	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.134
ECFP_6	2106656448	 [*]C(=O)[*]	-0.11
FCFP_6	1	 [*]=O	-0.102

Sorafenib

TOPKAT_Chronic_LOAEL



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.00483

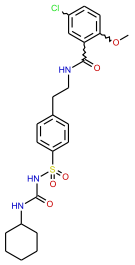
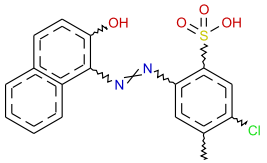
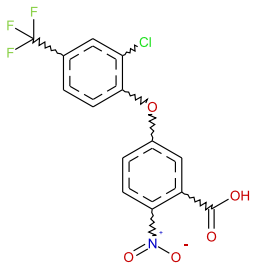
Unit: g/kg_body_weight

Mahalanobis Distance: 30

Mahalanobis Distance p-value: 1.21e-024

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

Structural Similar Compounds

Name	GLYBURIDE	D & C RED 9	SODIUM ACIFLUORFEN
Structure			
Actual Endpoint (-log C)	4.21661	3.87715	4.16036
Predicted Endpoint (-log C)	4.21035	3.6546	4.65915
Distance	0.636	0.722	0.736
Reference	UPJ-26452	NTP REPORT # 225	EPA COVER SHEET 0192;891101;(1)

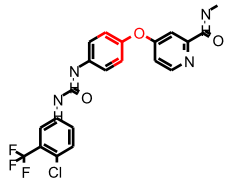
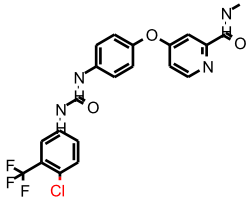
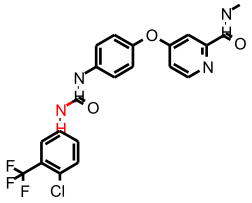
Model Applicability

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

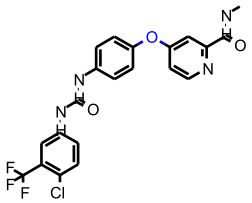
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -1046436026: [*]F
3. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
4. Unknown ECFP_6 feature: 226796801: [*]C([*])([*])F
5. Unknown ECFP_6 feature: 1305253718: [*]:[c](:[*])O[c](:[*]):[*]
6. Unknown ECFP_6 feature: -677309799: [*][c](:[*]):n:[cH]:[*]
7. Unknown ECFP_6 feature: 1338334141: [*]C(=[*])NC
8. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
9. Unknown ECFP_6 feature: 1336678434: [*][c](:[*]):[c](:[cH]:[*])C([*])([*])[*]
10. Unknown ECFP_6 feature: -649580166: [*]NC(=O)N[*]
11. Unknown ECFP_6 feature: -1952889961: [*]:[c](:[*])C(F)(F)F
12. Unknown ECFP_6 feature: 1413420509: [*]C(=[*])[c](:[cH]:[*]):n:[*]
13. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]
14. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
15. Unknown ECFP_6 feature: 864287155: [*]NC

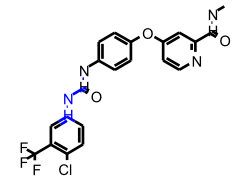
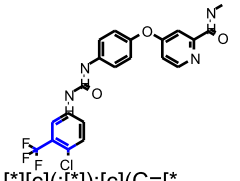
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.106
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

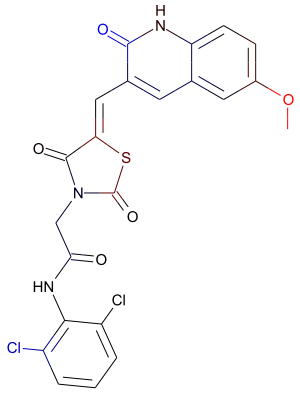
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	-0.102

ECFP_6	-1236483485	 [*]C(=[*])N[c](:[*]): [*]	-0.0747
FCFP_6	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.0713

Compound 8

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.0211

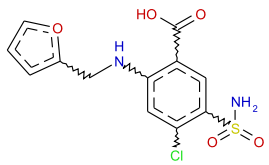
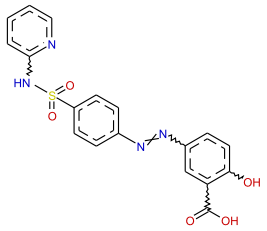
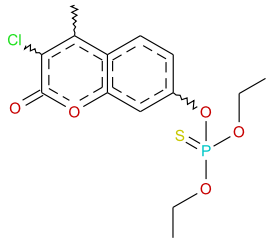
Unit: g/kg_body_weight

Mahalanobis Distance: 11.6

Mahalanobis Distance p-value: 6.59e-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	FUROSEMIDE	SALICYLAZOSULFAPYRIDINE	COUMAPHOS
Structure			
Actual Endpoint (-log C)	4.04236	3.375	5.60537
Predicted Endpoint (-log C)	2.8614	2.80292	4.15004
Distance	0.730	0.759	0.783
Reference	NCI/NTP TR-356	NCI/NTP TR-457	NCI/NTP TR-96

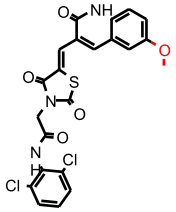
Model Applicability

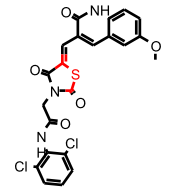
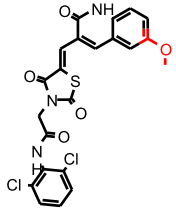
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 5.472. Training min, max, SD, explained variance: -3.3892, 5.0834, 1.644, 0.0611.
2. OPS PC9 out of range. Value: 3.4423. Training min, max, SD, explained variance: -2.8548, 3.3954, 1.263, 0.0360.

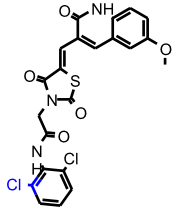
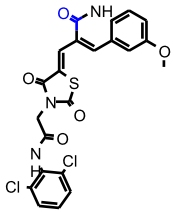
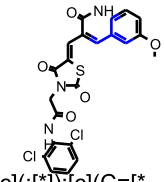
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

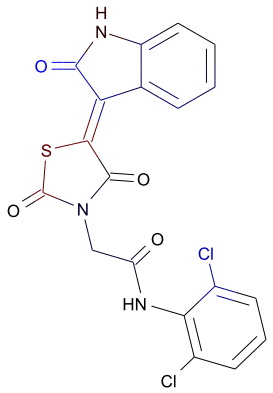
FCFP_2	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.095
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	 [*]:[c](:[*])Cl	-0.134
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.0829

Compound 9

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 0.0478

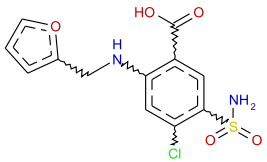
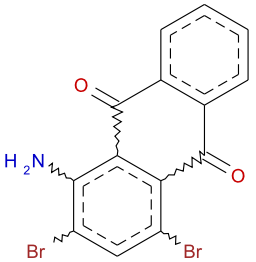
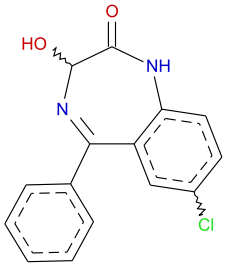
Unit: g/kg_body_weight

Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 1.82e-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	FUROSEMIDE	1-AMINO-2,4-DIBROMOANTHRAQUINONE	OXAZEPAM
Structure			
Actual Endpoint (-log C)	4.04236	2.82966	3.05262
Predicted Endpoint (-log C)	2.8614	3.92444	3.13073
Distance	0.639	0.676	0.696
Reference	NCI/NTP TR-356	NCI/NTP TR-383	NCI/NTP TR-468

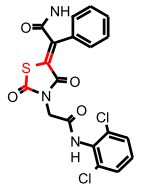
Model Applicability

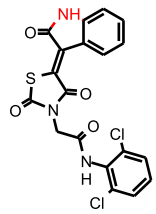
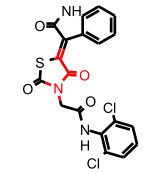
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 5.6206. Training min, max, SD, explained variance: -3.3892, 5.0834, 1.644, 0.0611.

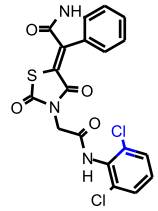
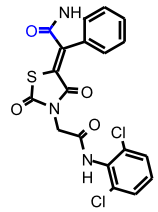
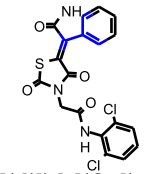
Feature Contribution

Top features for positive contribution

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

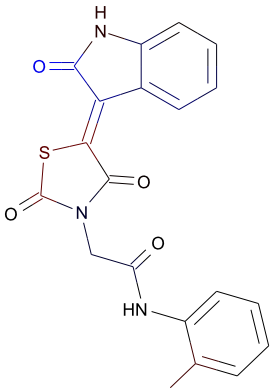
FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	565998553	 [*]N1[*][*]C(=[*])C1=O	0.00813

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	71476542	 [*]:[c](:[*])Cl	-0.134
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.0829

Compound 12

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



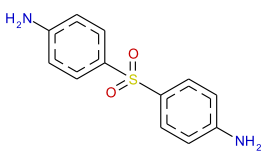
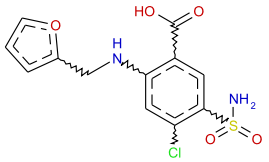
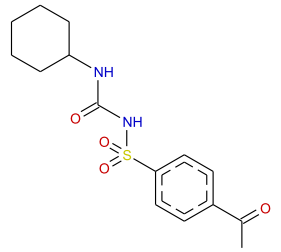
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.0398
Unit: g/kg_body_weight
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 1.74e-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	DAPSONE	FUROSEMIDE	ACETOHEXAMIDE
Structure			
Actual Endpoint (-log C)	3.66258	4.04236	2.55683
Predicted Endpoint (-log C)	3.26993	2.8614	3.62413
Distance	0.594	0.621	0.634
Reference	NCI/NTP TR-20	NCI/NTP TR-356	NCI/NTP TR-050

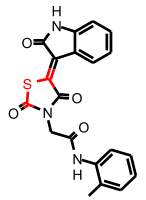
Model Applicability

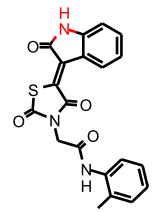
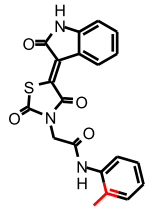
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

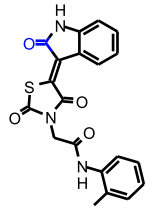
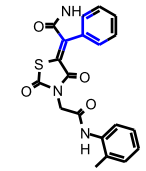
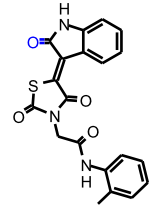
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

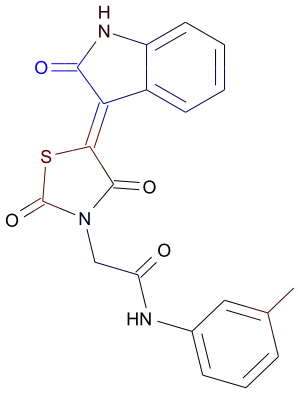
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.095

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	136120670	 [*]:[c](:[*])C	0.064

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.0829
FCFP_2	1	 [*]=O	-0.0796

Compound 13

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



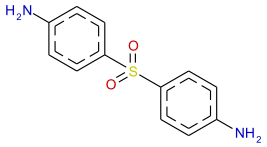
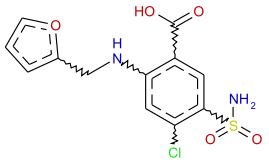
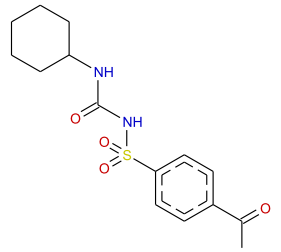
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.0398
Unit: g/kg_body_weight
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 1.74e-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	DAPSONE	FUROSEMIDE	ACETOHEXAMIDE
Structure			
Actual Endpoint (-log C)	3.66258	4.04236	2.55683
Predicted Endpoint (-log C)	3.26993	2.8614	3.62413
Distance	0.594	0.621	0.634
Reference	NCI/NTP TR-20	NCI/NTP TR-356	NCI/NTP TR-050

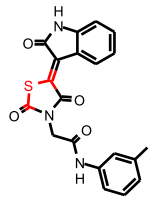
Model Applicability

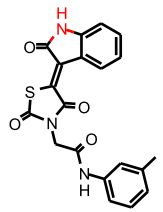
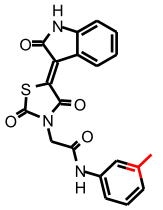
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

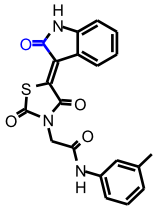
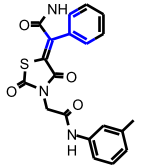
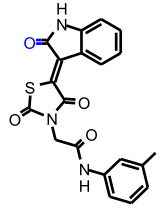
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

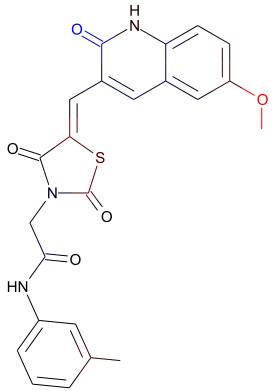
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.095

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	136120670	 [*]:[c](:[*])C	0.064

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.0829
FCFP_2	1	 [*]=O	-0.0796

Compound 14

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{23}H_{19}N_3O_5S$

Molecular Weight: 449.47905

ALogP: 2.637

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.0178

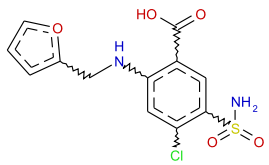
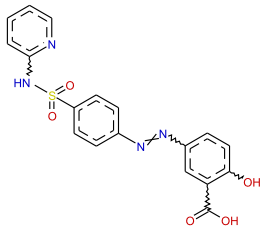
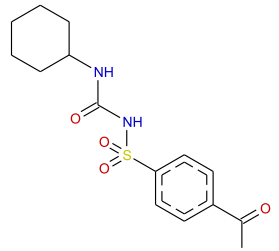
Unit: g/kg_body_weight

Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 5.02e-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	SALICYLAZOSULFAPYRIDINE	ACETOHEXAMIDE
Structure			
Actual Endpoint (-log C)	4.04236	3.375	2.55683
Predicted Endpoint (-log C)	2.8614	2.80292	3.62413
Distance	0.682	0.726	0.753
Reference	NCI/NTP TR-356	NCI/NTP TR-457	NCI/NTP TR-050

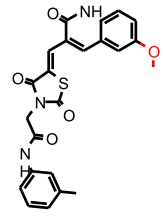
Model Applicability

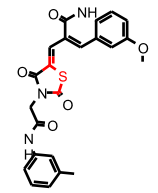
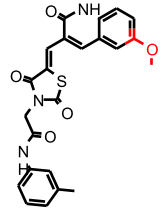
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

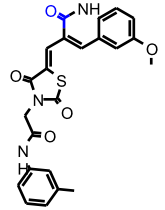
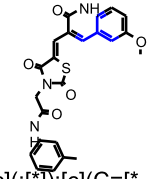
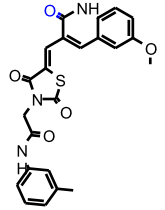
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]OC	0.173

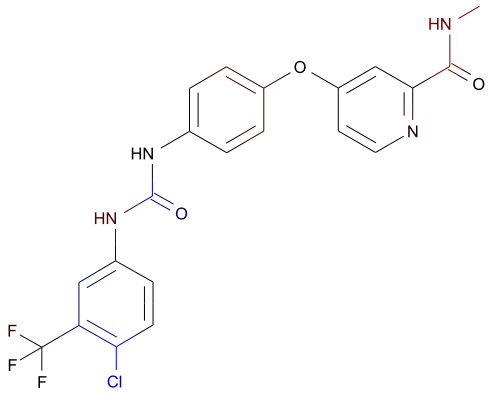
FCFP_2	-1143715940	 [*]=C1[*][*]C(=[*])S1	0.095
FCFP_2	1036089772	 [*]:[c](:[*])OC	0.0749

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.0829
FCFP_2	1	 [*]=O	-0.0796

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.0885

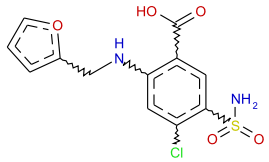
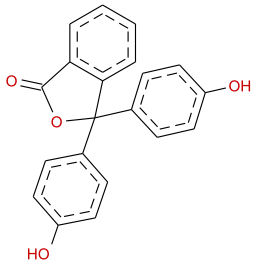
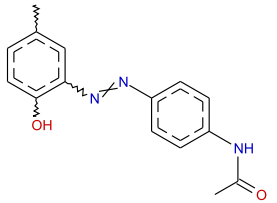
Unit: g/kg_body_weight

Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 1.76e-009

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

Structural Similar Compounds

Name	FUROSEMIDE	PHENOLPHTHALEIN	DISPERSE YELLOW 3
Structure			
Actual Endpoint (-log C)	4.04236	2.20184	2.77703
Predicted Endpoint (-log C)	2.8614	2.8857	2.80195
Distance	0.741	0.780	0.799
Reference	NCI/NTP TR-356	NCI/NTP TR-465	NCI/NTP TR-222

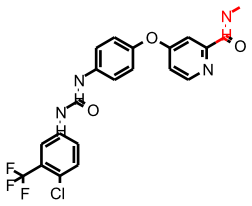
Model Applicability

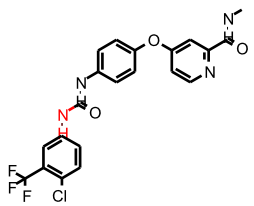
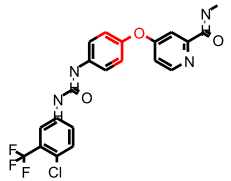
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

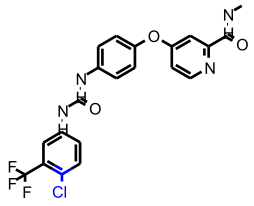
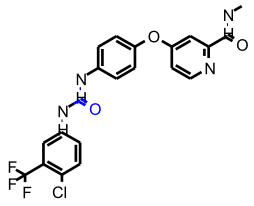
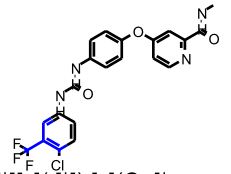
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

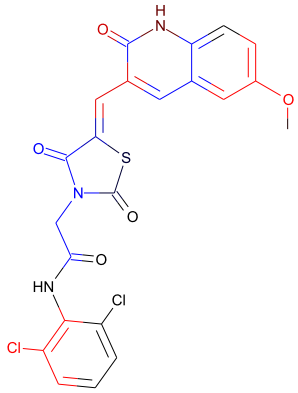
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-885550502	 [*]C(=[*])NC	0.115

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.0611

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_2	71476542	 [*]:[c](:[*])Cl	-0.134
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.0829

Compound 8

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



$C_{22}H_{15}Cl_2N_3O_5S$

Molecular Weight: 504.34259

ALogP: 3.479

Rotatable Bonds: 5

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.00054

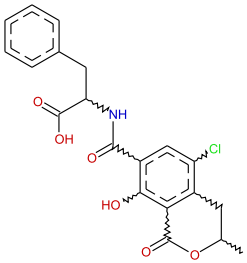
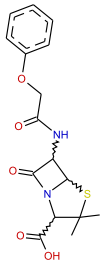
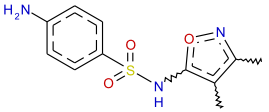
Unit: g/kg_body_weight

Mahalanobis Distance: 12

Mahalanobis Distance p-value: 1.01e-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	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	0.667	0.808	0.994
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

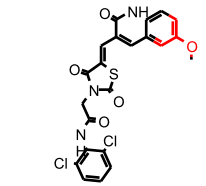
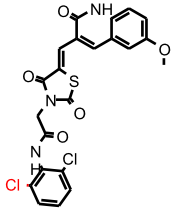
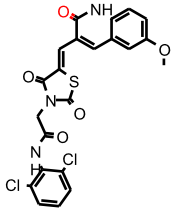
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 504.34. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. OPS PC5 out of range. Value: -4.1507. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
3. OPS PC10 out of range. Value: 2.77. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
4. Unknown FCFP_2 feature: 436915834: [*]\C=C\1/S[*][*]C1=[*]
5. Unknown FCFP_2 feature: -1986158408: [*]N1[*][*]SC1=O

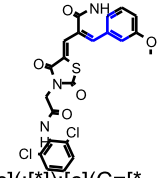
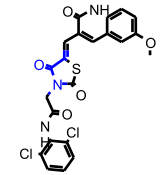
Feature Contribution

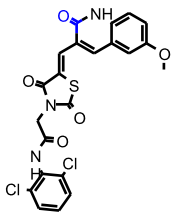
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511

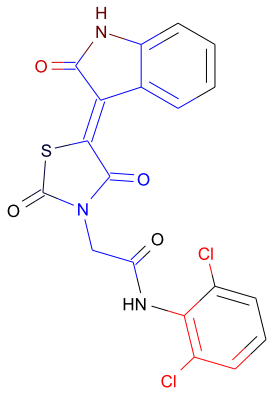
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.406
FCFP_2	565998553	 [*]N1[*][*]C(=[*])C1=O	-0.348

FCFP_2	1872154524	 [*]C(=O)[*]	-0.307

Compound 9

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 0.0102

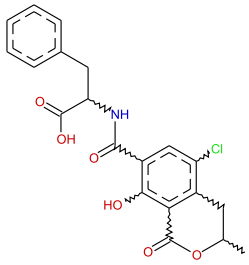
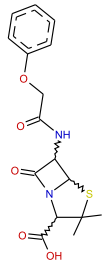
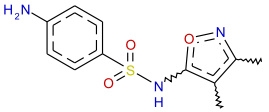
Unit: g/kg_body_weight

Mahalanobis Distance: 8.86

Mahalanobis Distance p-value: 0.000184

Mahalanobis Distance: The Mahalanobis distance (MD) is a generalization of the Euclidean distance that accounts for correlations among the X properties. It is calculated as the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	0.622	0.756	0.795
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

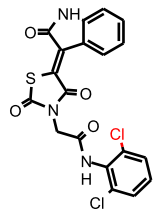
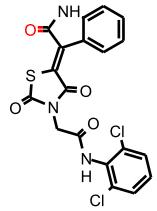
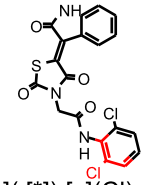
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 448.28. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. OPS PC5 out of range. Value: -3.9859. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
3. OPS PC10 out of range. Value: 2.8356. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
4. Unknown FCFP_2 feature: -1678275541: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]
5. Unknown FCFP_2 feature: 436915834: [*]\C=C\1/S[*][*]C1=[*]
6. Unknown FCFP_2 feature: -1986158408: [*]N1[*][*]SC1=O

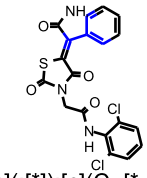
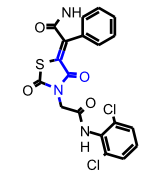
Feature Contribution

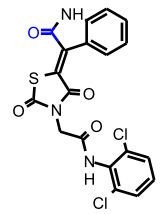
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511
FCFP_2	367998008	 [*][c](:[*]):[c](Cl):[cH]:[*]	0.413

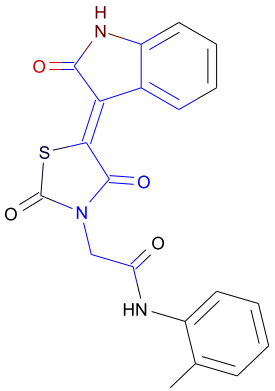
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.406
FCFP_2	565998553	 [*]N1[*][*]C(=[*])C1=O	-0.348

FCFP_2	1872154524	 [*]C(=O)[*]	-0.307

Compound 12

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



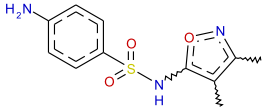
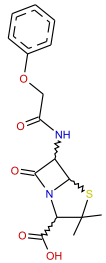
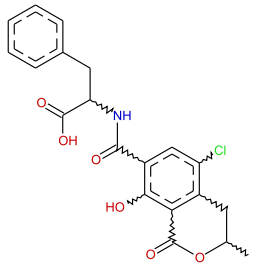
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.138
Unit: g/kg_body_weight
Mahalanobis Distance: 8
Mahalanobis Distance p-value: 0.00206

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

Structural Similar Compounds

Name	SULFISOOXAZOLE	PENICILLIN VK	OCHRATOXIN
Structure			
Actual Endpoint (-log C)	2.82494	2.54455	6.28396
Predicted Endpoint (-log C)	3.0705	3.9702	5.12358
Distance	0.618	0.657	0.682
Reference	NCI/NTP TR-138	NCI/NTP TR-336	NCI/NTP TR-358

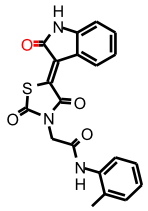
Model Applicability

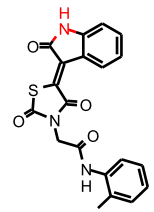
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

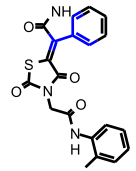
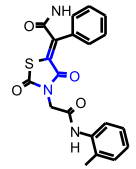
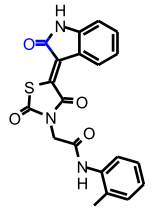
1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1678275541: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]
3. Unknown FCFP_2 feature: 436915834: [*]\C=C\1/S[*][*]C1=[*]
4. Unknown FCFP_2 feature: -1986158408: [*]N1[*][*]SC1=O

Feature Contribution

Top features for positive contribution

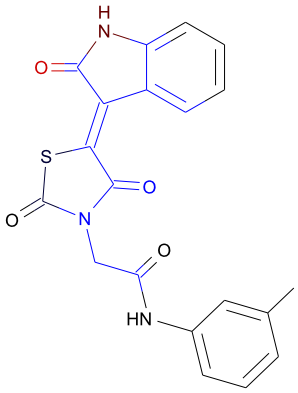
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1	 [*]=O	0.511

FCFP_2	3	 [*]N[*]	0.104

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.406
FCFP_2	565998553	 [*]N1[*][*]C(=[*])C1=O	-0.348
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307

Compound 13

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



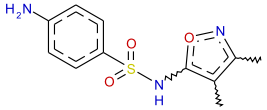
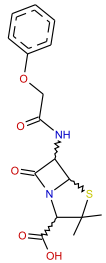
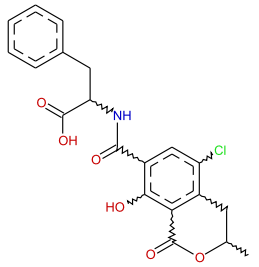
$C_{20}H_{15}N_3O_4S$
Molecular Weight: 393.4158
ALogP: 2.185
Rotatable Bonds: 3
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 0.138
Unit: g/kg_body_weight
Mahalanobis Distance: 8
Mahalanobis Distance p-value: 0.00206

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

Structural Similar Compounds

Name	SULFISOOXAZOLE	PENICILLIN VK	OCHRATOXIN
Structure			
Actual Endpoint (-log C)	2.82494	2.54455	6.28396
Predicted Endpoint (-log C)	3.0705	3.9702	5.12358
Distance	0.618	0.657	0.682
Reference	NCI/NTP TR-138	NCI/NTP TR-336	NCI/NTP TR-358

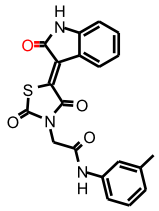
Model Applicability

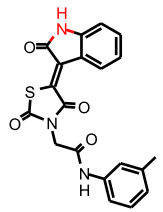
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1678275541: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]
3. Unknown FCFP_2 feature: 436915834: [*]\C=C\1/S[*][*]C1=[*]
4. Unknown FCFP_2 feature: -1986158408: [*]N1[*][*]SC1=O

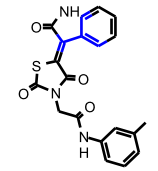
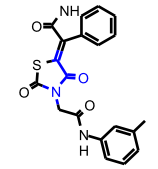
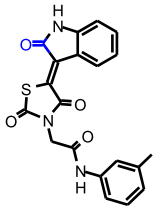
Feature Contribution

Top features for positive contribution

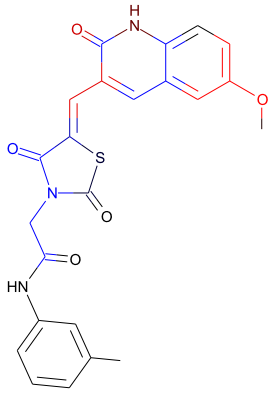
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	1	 [*]=O	0.511

FCFP_2	3	 [*]N[*]	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.406
FCFP_2	565998553	 [*]N1[*][*]C(=[*])C1=O	-0.348
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307

Compound 14



$C_{23}H_{19}N_3O_5S$
Molecular Weight: 449.47905
ALogP: 2.637
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

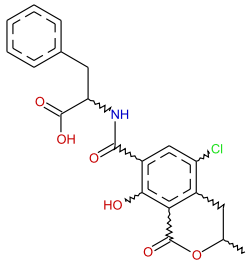
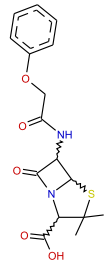
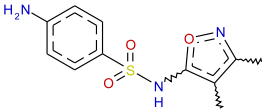
Model Prediction

Prediction: 0.0074
Unit: g/kg_body_weight
Mahalanobis Distance: 10.5
Mahalanobis Distance p-value: 1.31e-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_Rat_Maximum_Tolerated_Dose_Gavage

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	0.658	0.679	0.837
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

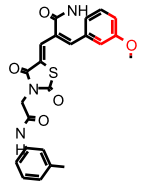
Model Applicability

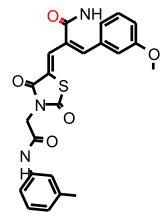
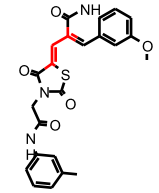
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 449.48. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Unknown FCFP_2 feature: 436915834: [*]\C=C\1/S[*][*]C1=[*]
3. Unknown FCFP_2 feature: -1986158408: [*]N1[*][*]SC1=O

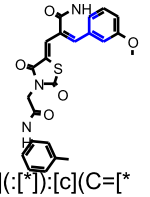
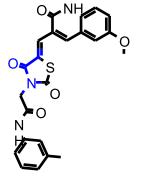
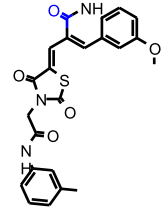
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.672

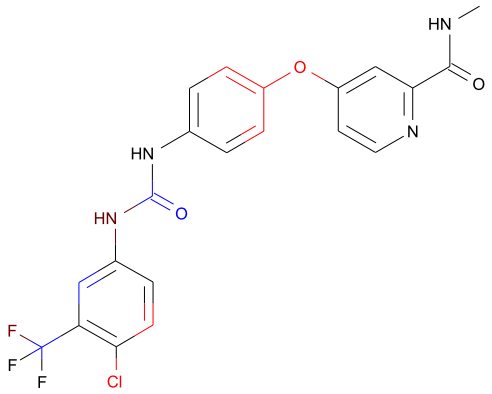
FCFP_2	1	 [*]=O	0.511
FCFP_2	451847724	 [*]C(=CC(=[*])[*])[*]	0.225

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*])):[cH]:[*]	-0.406
FCFP_2	565998553	 [*]N1[*][*]C(=[*])C1=O	-0.348
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.000918

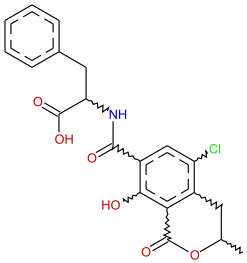
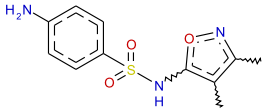
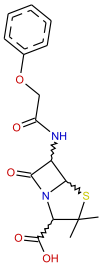
Unit: g/kg_body_weight

Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 4.69e-009

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

Structural Similar Compounds

Name	OCHRATOXIN	SULFISOOXAZOLE	PENICILLIN VK
Structure			
Actual Endpoint (-log C)	6.28396	2.82494	2.54455
Predicted Endpoint (-log C)	5.12358	3.0705	3.9702
Distance	0.758	0.997	1.159
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

Model Applicability

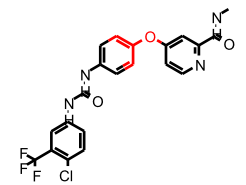
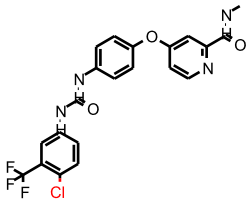
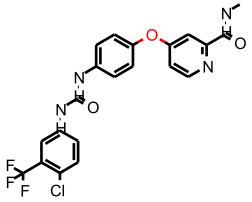
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 464.82. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
3. OPS PC5 out of range. Value: -3.5737. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
4. OPS PC7 out of range. Value: -3.8342. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
5. Unknown FCFP_2 feature: 1499521844: [*]NC(=O)N[*]
6. Unknown FCFP_2 feature: -1029533685: [*]:[c](:[*])C(F)(F)F
7. Unknown FCFP_2 feature: 136686699: [*]NC

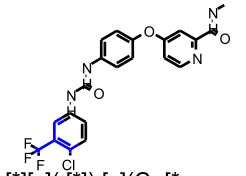
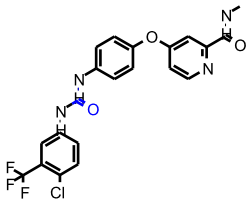
Feature Contribution

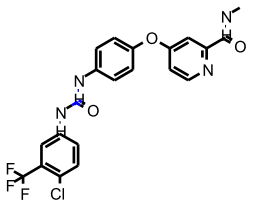
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

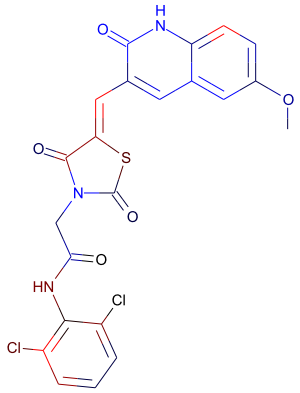
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.672
FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	203677720	 [*][c](:[*]):[c](C=[*]):[cH]:[*]	-0.406
FCFP_2	1872154524	 [*]C(=O)[*]	-0.307

FCFP_2	0	 [*]C(=[*])[*]	-0.29

Compound 8



$C_{22}H_{15}Cl_2N_3O_5S$
Molecular Weight: 504.34259
ALogP: 3.479
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

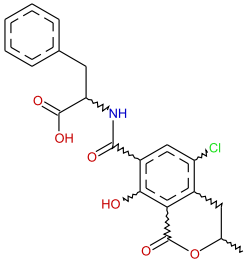
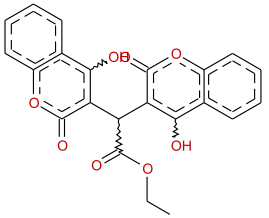
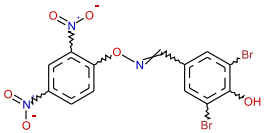
Model Prediction

Prediction: 0.576
Unit: g/kg_body_weight
Mahalanobis Distance: 24.1
Mahalanobis Distance p-value: 3.74e-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.

TOPKAT_Rat_Oral_LD50

Structural Similar Compounds

Name	OCHRATOXIN A	ETHYL-bis-COUMACETATE	BROMOFENOXIME
Structure			
Actual Endpoint (-log C)	4.305	2.687	2.622
Predicted Endpoint (-log C)	3.03558	2.7054	3.41798
Distance	0.644	0.661	0.662
Reference	FCTXAV 6;479;68	FEPRA7 10;303;51	85ARAE 2;203;77

Model Applicability

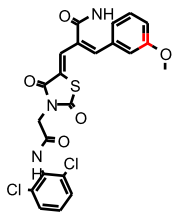
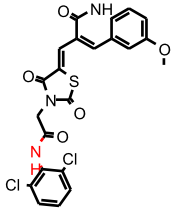
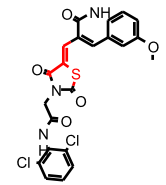
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC58 out of range. Value: -4.1222. Training min, max, SD, explained variance: -4.0974, 5.4085, 1.034, 0.0039.
2. Unknown ECFP_2 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
3. Unknown ECFP_2 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
4. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
5. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
6. Unknown FCFP_6 feature: 451371068: [*]C(=C[c](:[*]):[*])[*]
7. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

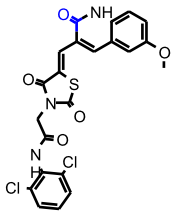
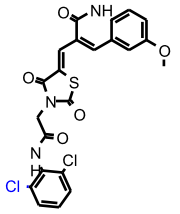
Feature Contribution

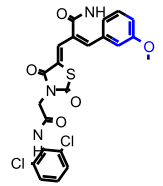
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.184

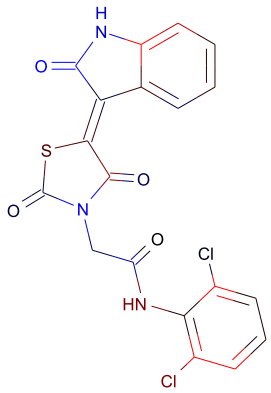
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	-817402818	 [*]Cl	-0.263

ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[c H]:[*]	-0.257

Compound 9

TOPKAT_Rat_Oral_LD50



$C_{19}H_{11}Cl_2N_3O_4S$

Molecular Weight: 448.27933

ALogP: 3.028

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 0.838

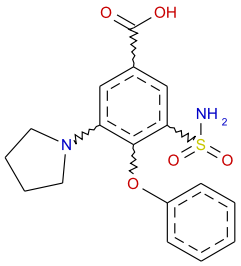
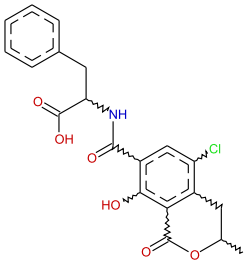
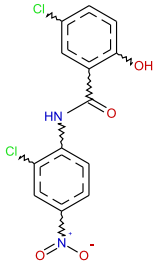
Unit: g/kg_body_weight

Mahalanobis Distance: 21.3

Mahalanobis Distance p-value: 2.05e-013

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

Structural Similar Compounds

Name	PIRETANIDE	OCHRATOXIN A	CLONITRALID
Structure			
Actual Endpoint (-log C)	1.811	4.305	2.816
Predicted Endpoint (-log C)	1.83976	3.03558	2.74335
Distance	0.618	0.622	0.650
Reference	DRFUD4 2;393;77	FCTXAV 6;479;68	85ARAE 3;103;76/77

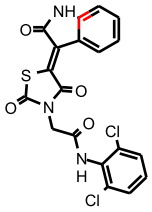
Model Applicability

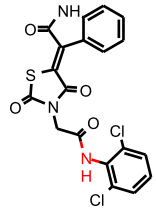
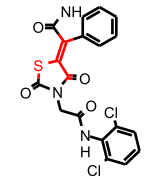
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
4. Unknown FCFP_6 feature: -1678275541: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]
5. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

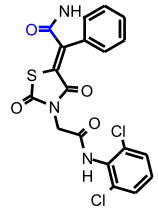
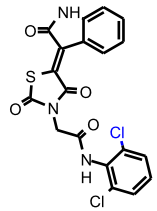
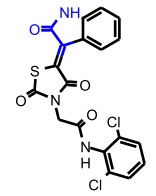
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

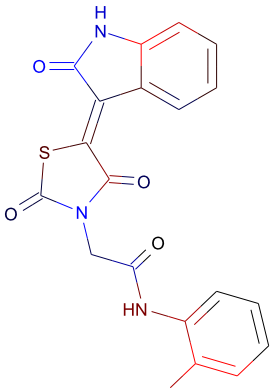
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.184

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	-817402818	 [*]Cl	-0.263
FCFP_6	566058135	 [*]NC(=O)C(=[*])[*]	-0.216

Compound 12

TOPKAT_Rat_Oral_LD50



$C_{20}H_{15}N_3O_4S$

Molecular Weight: 393.4158

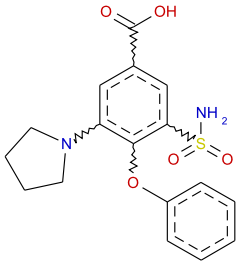
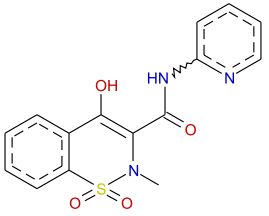
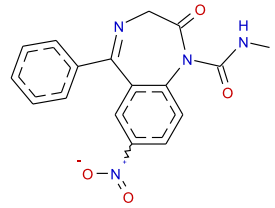
ALogP: 2.185

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Structural Similar Compounds

Name	PIRETANIDE	PIROXICAM	1H-1;4-BENZODIAZEPINE-1-CARBOXAMIDE; 2;3-DIHYDRO-N-METHYL-7-NITRO-2-OXO-5-PHENYL-
Structure			
Actual Endpoint (-log C)	1.811	3.186	2.171
Predicted Endpoint (-log C)	1.83976	2.63418	2.64752
Distance	0.549	0.580	0.581
Reference	DRFUD4 2;393;77	ARZNAD 28;1714;78	TAKHAA 29;153;70

Model Prediction

Prediction: 1.03

Unit: g/kg_body_weight

Mahalanobis Distance: 21.7

Mahalanobis Distance p-value: 5.71e-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.

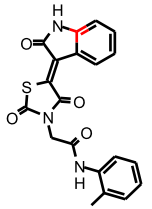
Model Applicability

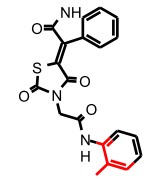
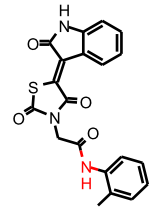
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
4. Unknown FCFP_6 feature: -1678275541: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]

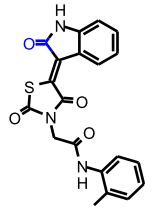
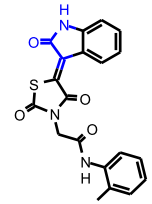
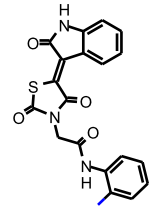
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

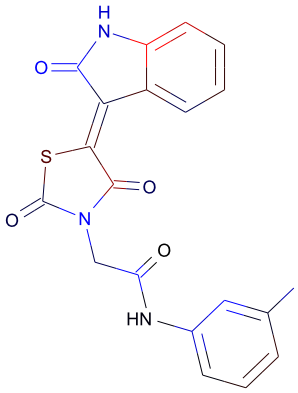
ECFP_6	2147419938	 [*][c](:[*]):[c](C):[cH]:[*]	0.263
ECFP_6	-1897341097	 [*]N[*]	0.216

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
FCFP_6	566058135	 [*]NC(=O)C(=[*])[*]	-0.216
ECFP_6	734603939	 [*]C	-0.201

Compound 13

TOPKAT_Rat_Oral_LD50



$C_{20}H_{15}N_3O_4S$

Molecular Weight: 393.4158

ALogP: 2.185

Rotatable Bonds: 3

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 2.88

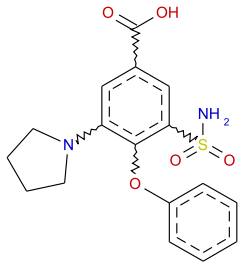
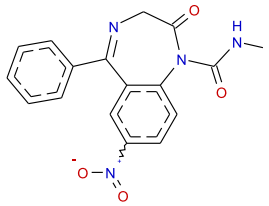
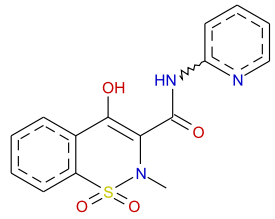
Unit: g/kg_body_weight

Mahalanobis Distance: 21.7

Mahalanobis Distance p-value: 5.71e-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	PIRETANIDE	1H-1;4-BENZODIAZEPINE-1-CARBOXAMIDE; 2;3-DIHYDRO-N-METHYL-7-NITRO-2-OXO-5-PHENYL-	PIROXICAM
Structure			
Actual Endpoint (-log C)	1.811	2.171	3.186
Predicted Endpoint (-log C)	1.83976	2.64752	2.63418
Distance	0.549	0.581	0.582
Reference	DRFUD4 2;393;77	TAKHAA 29;153;70	ARZNAD 28;1714;78

Model Applicability

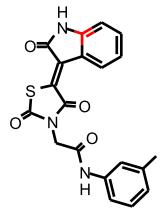
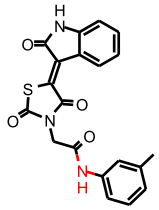
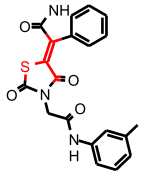
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
4. Unknown FCFP_6 feature: -1678275541: [*]C(=C1C(=[*])[*][*]:[c]1:[*])[*]

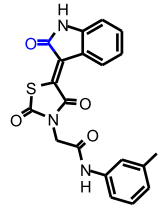
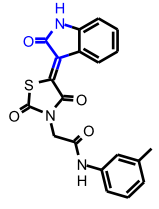
Feature Contribution

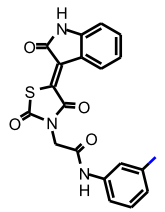
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.184

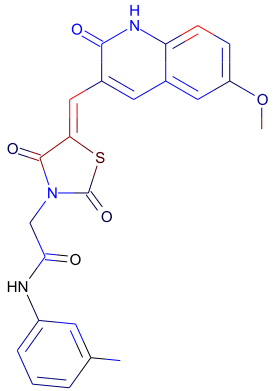
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
FCFP_6	566058135	 [*]NC(=O)C(=[*])[*]	-0.216

ECFP_6	734603939	 [*]C	-0.201

Compound 14

TOPKAT_Rat_Oral_LD50



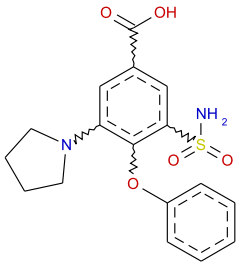
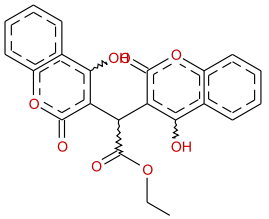
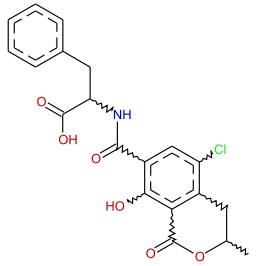
$C_{23}H_{19}N_3O_5S$
Molecular Weight: 449.47905
ALogP: 2.637
Rotatable Bonds: 5
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 1.21
Unit: g/kg_body_weight
Mahalanobis Distance: 23.6
Mahalanobis Distance p-value: 1.42e-023

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

Structural Similar Compounds

Name	PIRETANIDE	ETHYL-bis-COUMACETATE	OCHRATOXIN A
Structure			
Actual Endpoint (-log C)	1.811	2.687	4.305
Predicted Endpoint (-log C)	1.83976	2.7054	3.03558
Distance	0.557	0.586	0.615
Reference	DRFUD4 2;393;77	FEPRA7 10;303;51	FCTXAV 6;479;68

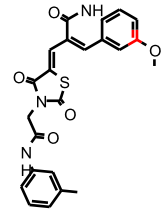
Model Applicability

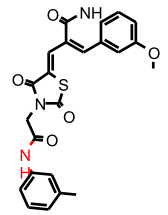
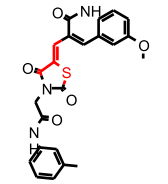
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 2131425032: [*]\C=C(\C=[*])/C(=[*])[*]
3. Unknown ECFP_2 feature: 1000552169: [*]\C=C\1/S[*][*]C1=[*]
4. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
5. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
6. Unknown FCFP_6 feature: 451371068: [*]C(=C[c](:[*]):[*])[*]

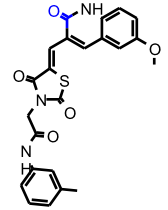
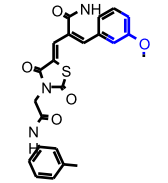
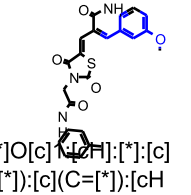
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

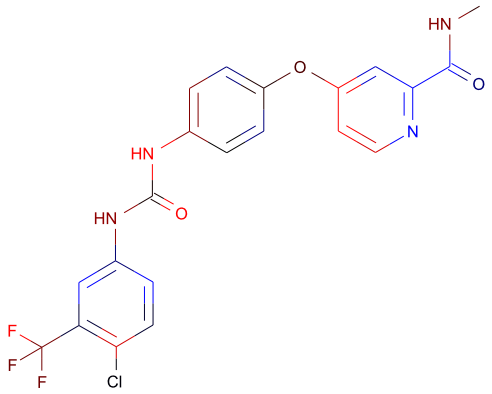
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	436915834	 [*]\C=C\1/S[*][*]C1=[*]	0.184

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[c H]:[*]	-0.257
FCFP_6	1676877079	 [*]O[c]1:[cH]:[*]:[c] ([*]):[c](C=[*]):[cH]:1	-0.254

Sorafenib

TOPKAT_Rat_Oral_LD50



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

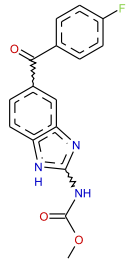
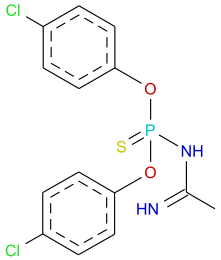
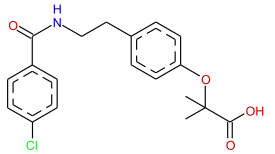
ALogP: 4.175

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

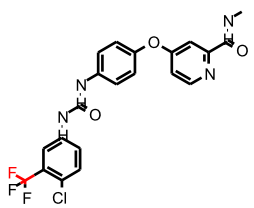
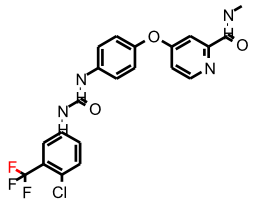
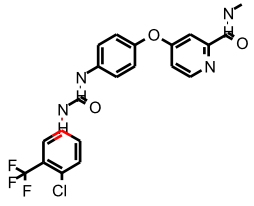
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
3. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl
4. Unknown FCFP_6 feature: 1747237384: [*][c](:[*]):n:[cH]:[*]
5. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
6. Unknown FCFP_6 feature: 136686699: [*]NC

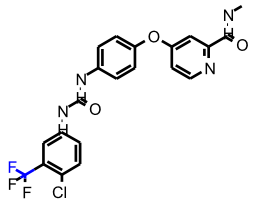
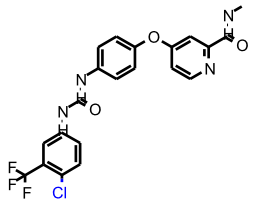
Feature Contribution

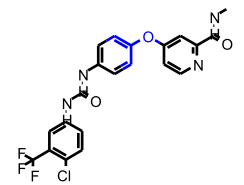
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	71953198	 [*]C([*])([*])F	0.392
ECFP_6	-1046436026	 [*]F	0.349
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	226796801	 [*]C([*])([*])F	-0.32
ECFP_6	-817402818	 [*]Cl	-0.263

ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[c H]:[*]	-0.257

RAW DATA OF VEGFR-2 ASSAY

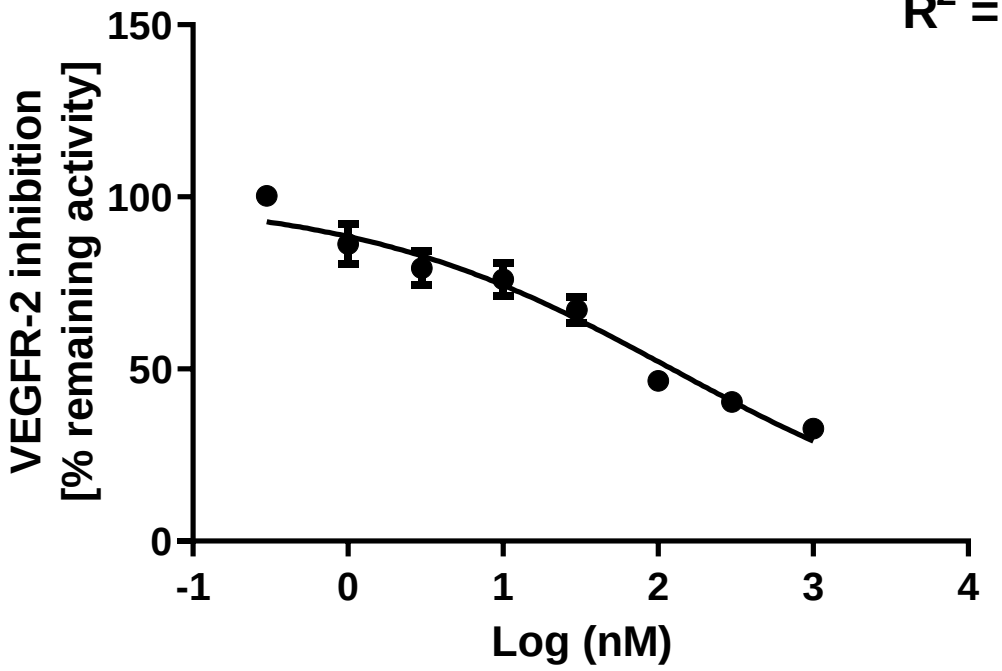
VEGFR-2 assay of compound 8

Best-fit values	
LogIC50	2.089
HillSlope	-0.4244
IC50	122.9

FQ4

IC_{50} = 122.9

R^2 = 0.9185



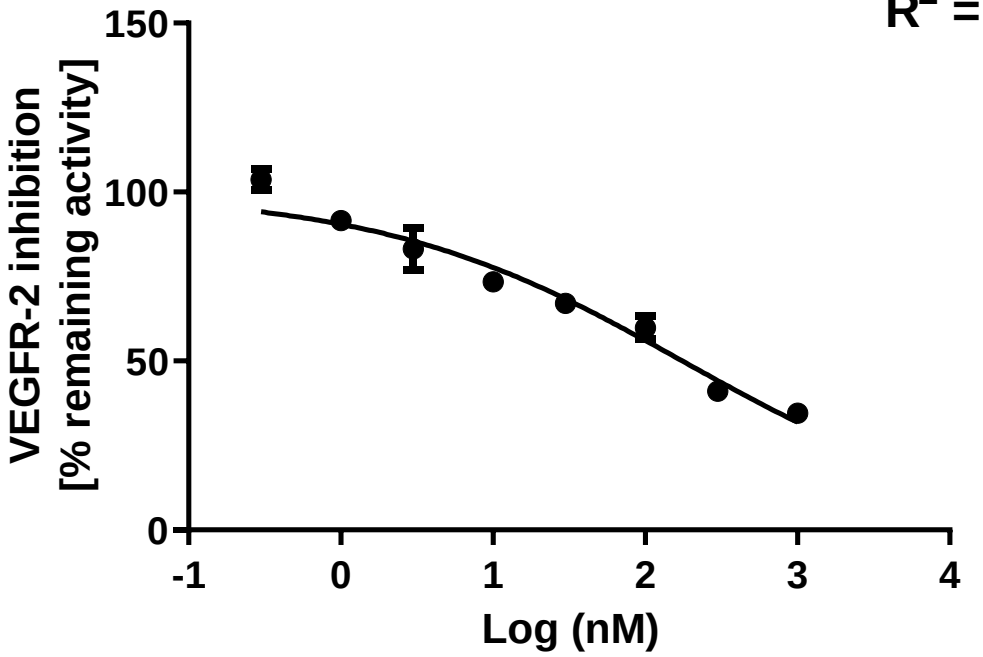
VEGFR-2 assay of compound 9

Best-fit values	
LogIC50	2.244
HillSlope	-0.4357
IC50	175.5

FQ8

IC_{50} = 175.5

R^2 = 0.9260



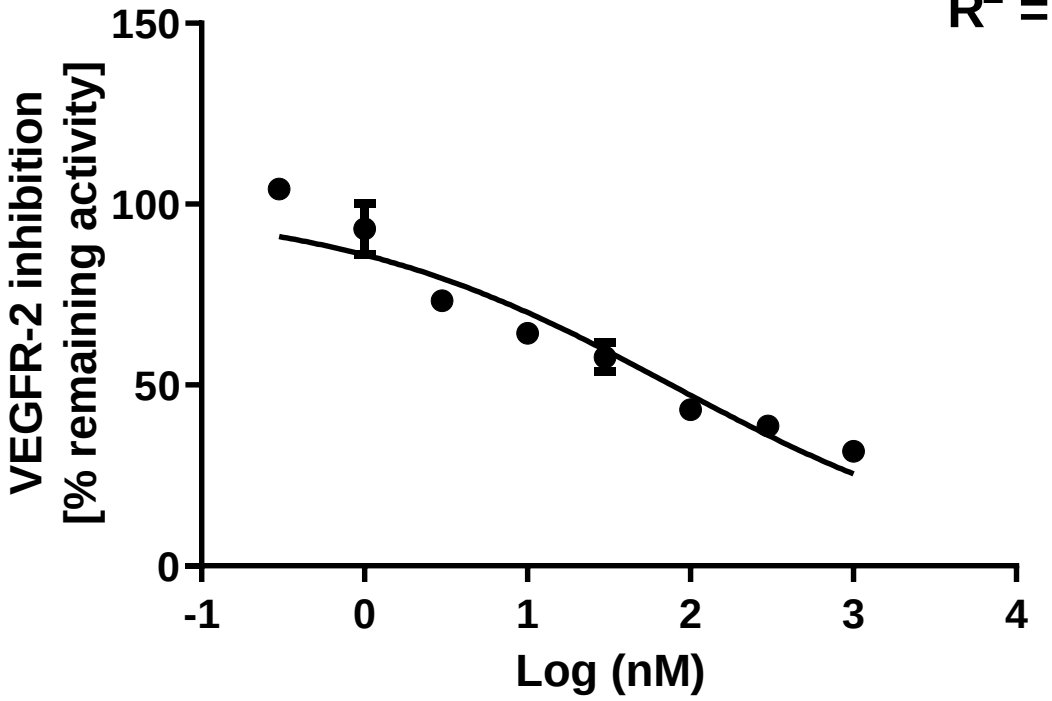
VEGFR-2 assay of compound 12

Best-fit values	
LogIC50	1.884
HillSlope	-0.4179
IC50	76.64

FI-8

IC_{50} = 76.64

R^2 = 0.8911



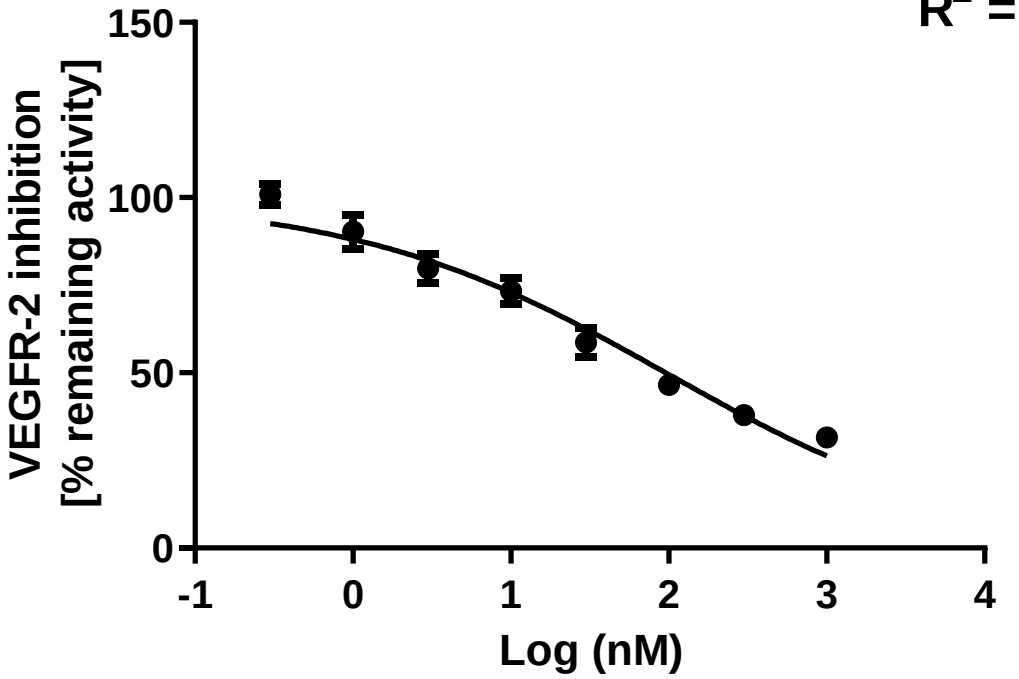
VEGFR-2 assay of compound 13

Best-fit values	
LogIC50	1.982
HillSlope	-0.4387
IC50	95.87

FI-3

IC_{50} = 95.87

R^2 = 0.9341



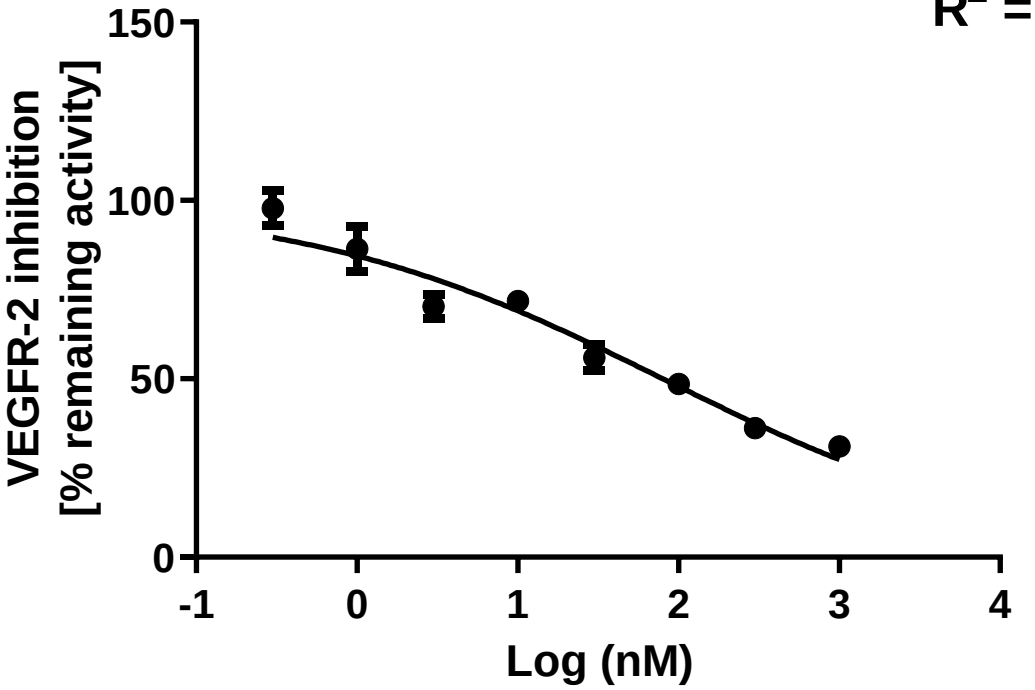
VEGFR-2 assay of compound 14

Best-fit values	
LogIC50	1.904
HillSlope	-0.3860
IC50	80.15

FI-4

IC_{50} = 80.15

R^2 = 0.9088



VEGFR-2 assay of compound Sorafenib

Best-fit values	
LogIC50	1.730
HillSlope	-0.4752
IC50	53.65

Sorafinib

IC_{50} = 53.65

R^2 = 0.9248