In silico, in vitro VEGFR-2 inhibition, and anticancer activity of a 3-(hydrazonomethyl)naphthalene-2-ol derivative

Eslam B. Elkaeed[a], Reda G.Yousef [b], Hazem Elkady[b], Ahmed B. M. Mehany[c], Bshra A. Alsfouk[d], Dalal Z. Husein[e], Ibrahim M. Ibrahim[f], Ahmed M. Metwaly[g,h]* Ibrahim H. Eissa[b]*

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

[b] Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo11884, Egypt.

[c] Zoology Department, Faculty of Science (Boys), Al-Azhar University, Cairo11884, Egypt.

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

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

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

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

[h] Biopharmaceutical Products Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications (SRTA-City), Alexandria, Egypt.

***Corresponding authors:**

Ibrahim H. Eissa

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

Ahmed M. Metwaly

Pharmacognosy and Medicinal Plants Department, Faculty of Pharmacy (Boys), Al-AzharUniversity, Cairo 11884, Egypt**.** **Email:** ametwaly@azhar.edu.eg

Content

In silico studies

1- Molecular Docking studies

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

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

Docking Setup and Validation of Docking Protocol: The protein-ligand docking studies were carried out using MOE version 2014. Validation of the docking protocol was carried out by redocking the co-crystallized reference ligand (sorafenib) against the isolated pocket of VEGFR-2. The docking protocol was validated by comparing the heavy atoms RMSD value of the re-docked ligand pose with the corresponding co-crystallized reference ligand structure.

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

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

2- ADMET studies

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

3- Toxicity studies

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

4- Molecular Dynamic Simulation:

We used the CHARMM-GUI web server to create the appropriate files for molecular dynamic (MD) simulation of the protein-compound complex. We utilized the TIP3P water model with a padding of 10 to solvate the system in a cubic box before neutralizing it by adding Na+ and Cl- ions to the physiological concentration of 0.154 M. The system was parameterized using the CHAMM36m force field, while the ligand was parameterized using the CHARMM general force field (CGenFF) tool implemented in CHARMM-GUI. To run the simulation with periodic boundary conditions (PBC), GROMACS 2021 was used as an MD engine. First, using the steepest descent algorithm, a minimization step was performed to remove any atom clashes. When the maximal force was less than 100 KJ.mol-1.nm-1, the minimization was considered to be converged. Two equilibration steps were performed starting with a constant number of atoms, constant volume, and constant temperature (NVT) ensemble, followed by a constant number of atoms, constant pressure, and constant temperature (NPT) ensemble. The temperature was set to 310 K and was maintained by the V-rescale algorithm. While the pressure was set to 1 atmospheric

pressure and was maintained through Berendsen barostat. Finally, a 100-ns production run in NVT ensemble was performed. The bond lengths of hydrogen-bonded atoms were constrained using the LINear Constraint Solver (LINCS) method in each of the previous stages. The electrostatics were calculated using the Particle Mesh Ewald (PME) technique with a threshold of 1.2 nm. The Newtonian equations of motion were integrated using the leap-frog algorithm with a time step of 1 femtosecond for the equilibration steps and 2 femtoseconds for the production step. Before analyzing the trajectory, the PBC was removed using GROMACS built-in tool trjconv. The production run was saved each 100 picoseconds with a total of 1000 frames for each system. The analysis of the production trajectory was performed using VMD TK scripts. Root mean square deviation (RMSD), root mean square fluctuation (RMSF), solvent accessible surface area (SASA), radius of gyration (RoG), and the number of hydrogen bonds were calculated. To get a representative frame for each cluster, the trajectory was clustered using TTClust. First, backbone alignment was performed before determining the optimum number of clusters using the elbow method. For each representative frame, protein-ligand interaction profiler (PLIP) was used to detect the number and types of interactions.

5- MM-GBSA Studies

Gmx_MMPBSA package was used to calculate the binding free energy using Molecular Mechanics Generalized Born Surface Area (MM-GBSA) algorithm with decomposition analysis to get the binding energies of amino acids within 10 Å around the ligand. The salt concentration and the method of solvation (igb) were set to 0.154 M and 5, respectively. The internal and external dielectric constant were set to 1.0 and 80.0, respectively, and other options were set as default. MM-GBSA approach is depicted in Equation 1

$\Delta G = < G_{complex} - G_{receptor} - G_{ligand} >$ …….Equation 1

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

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

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

ΔEgas = ΔEele + ΔEvdW ……..Equation 4

ΔEsolv = EGB + ESA…… Equation 5

ESA = γ.SASA ……..Equation 6

Where:

ΔH is the enthalpy which can be calculated from gas-phase energy (Egas) and solvation-free energy (Esol). TΔS is the entropy contribution to the free binding energy. Egas is composed of electrostatic and van der Waals terms; Eele, EvdW, respectively. Esol can be calculated from the polar solvation energy (EGB) and nonpolar solvation energy (ESA) which is estimated from the solvent-accessible surface area.

6- Density Function Theory (DFT) calculations

The Gaussian 09 program was used to perform the quantum chemistry calculations using the DFT method. GaussianView5 was used to display all of the data files. The density function theory (DFT) at 6-311G++(d,p) basis set/B3LYP approach was utilized to optimize organic chemical structure of the compound under investigation and Chem3D 15.0 software was used to create the original chemical structures. Both the Total Electron Density (TED) and the Electrostatic Surface (ESP) maps were examined at the same theoretical level. GaussSum3.0 software was used to compute and evaluate the total density of state (TDOS) for the optimized log file.

Equations of Koopmans' theory: The chemical potential (μ), maximal charge acceptance (ΔNmax), global hardness (η), energy change (ΔE), electronegativity (χ), the global softness (σ), electrophilicity index (ω), ionization potential (IP) and electron affinity (EA)

$IP = -EHOMO$

$EA = -ELUMO$

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

$\chi = -\eta$

$\omega = \mu 2/(2\eta)$ $\sigma = 1/\eta$

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

$\Delta E = -\omega$

Egap= *E*LUMO–*E*HOMO

Biological testing

1- In vitro antitumor assay

This test was carried out on two different human cancer cell lines: MCF-7 and HCT 116. We performed this colorimetric experiment with quantitative assessment of anticancer activity in accordance with the MTT method [1] The assay is based primarily on a biochemical reaction in which mitochondrial succinate dehydrogenase transforms yellow tetrazolium bromide (MTT) to a purple formazan derivative in live cells. The cells were grown in RPMI-1640 media with 10% foetal bovine serum in general. In a 5 percent CO2 incubator at 37 °C, penicillin (100 units/mL) and streptomycin (100g/mL) were introduced. The cells were seeded at a density of 1.0x104 cells/well in a 96-well plate at 37 C for 48 hours under 5% CO2. Following the initial incubation period. The cells were given varying concentrations of the novel compounds and cultured for 24 hours. Then we add 20 l of MTT in a 5 mg/ml solution and let it sit for 4 hours. 100 l of DMSO was applied to each well to dissolve the purple formazan produced. A plate reader (EXL 800, USA) was used to measure and record the colorimetric test at a wavelength of 570 nm. The 50% inhibitory concentration (IC_{50}) was estimated from graphic plots of the dose response curve for each conc. using Graphpad Prism software (San Diego, CA. USA). The data presented are the mean of at least three separate experiments.

2- In vitro VEGFR-2 kinase inhibitory assay

The IC_{50} of the tested compounds were determined against VEGFR-2 kinase with enzyme linked immunosorbent assay (ELISA). We used a human VEGFR-2 ELISA kit. VEGFR-2 specific antibody was seeded on a 96 well microplate and 100 µL of solution of the standard or the tested compound was added. After incubation for 2.5 h. at room temperature and washing, 100 µL of prepared biotin antibody was added. Incubation of the plates for 1 h at room temperature and then washing were carried out before addition of streptavidin solution (100 µL). The mixture was incubated for 1h at room temperature and then washed. After that, 100 µL of tetramethybenzidine (TMB) substrate solution was added before incubation for half an hour at room temperature. Finally, 50 µL of the stop solution was added immediately before the reading at 450 nm. The standard curve was hence drawn, concentration was on the X-axis and the absorbance was on the Y-axis. Percent inhibition was calculated by the comparison of compounds treated to control incubations. The concentration of the test compound causing 50% inhibition (IC_{50}) was calculated

from the concentration–inhibition response curve (triplicate determinations) and the data were compared with Sorafenib as standard VEGFR-2 inhibitor.

3- Safety assay

The safety profile of the tested compound was checked on one non-cancerous cell line (W138) to determine the treatments concentrations that do not depict toxic effects against the tested cells. A portion of 100.0 µl of 6×104 cell/ml cells was seeded into each well of a 96-well plate and then the plates were incubated at 37ºC in a humidified 5.0% CO2 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.

4- Selectivity index (SI)

The selectivity index values of the tested compound 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**.**

4- Spectral data of the final target compound 7

The melting point was carried out by open capillary method on a Gallen kamp Melting point apparatus. The infrared spectra were recorded on pye Unicam SP 1000 IR spectrophotometer using potassium bromide disc technique. Proton magnetic resonance ^{1}HNMR spectra were recorded on a Bruker 400 Megahertz-nuclear magnetic resonance (400 MHZ-NMR) spectrophotometer. Carbon-13 (C13) nuclear magnetic resonance (^{13}CNMR) spectra were recorded on a Bruker 100 Megahertz-nuclear magnetic resonance (100 MHZ-NMR) spectrophotometer. Tetramethylsilane (TMS) was used as internal standard and chemical shifts were measured in δ scale one part per million (ppm). The reactions were monitored by thin-layer chromatography (TLC) using TLC sheets precoated with UV fluorescent silica gel Merck 60 F254 plates and were visualized using ultraviolet (UV) lamp and different solvents as mobile phases.

Al-Azhar University
The Regional Center for Mycology and Biotechnology

Requester Data:

Name: **Dr. Ibrahim Hasan Eisa**

Authority: Faculty of Pharmacy, Al-Azhar University

Sample Data:

One sample had been submitted for elemental analysis.

Analysis Report:

Sample Code	C%	H%	N%
Q	73.58	4.85	10.50

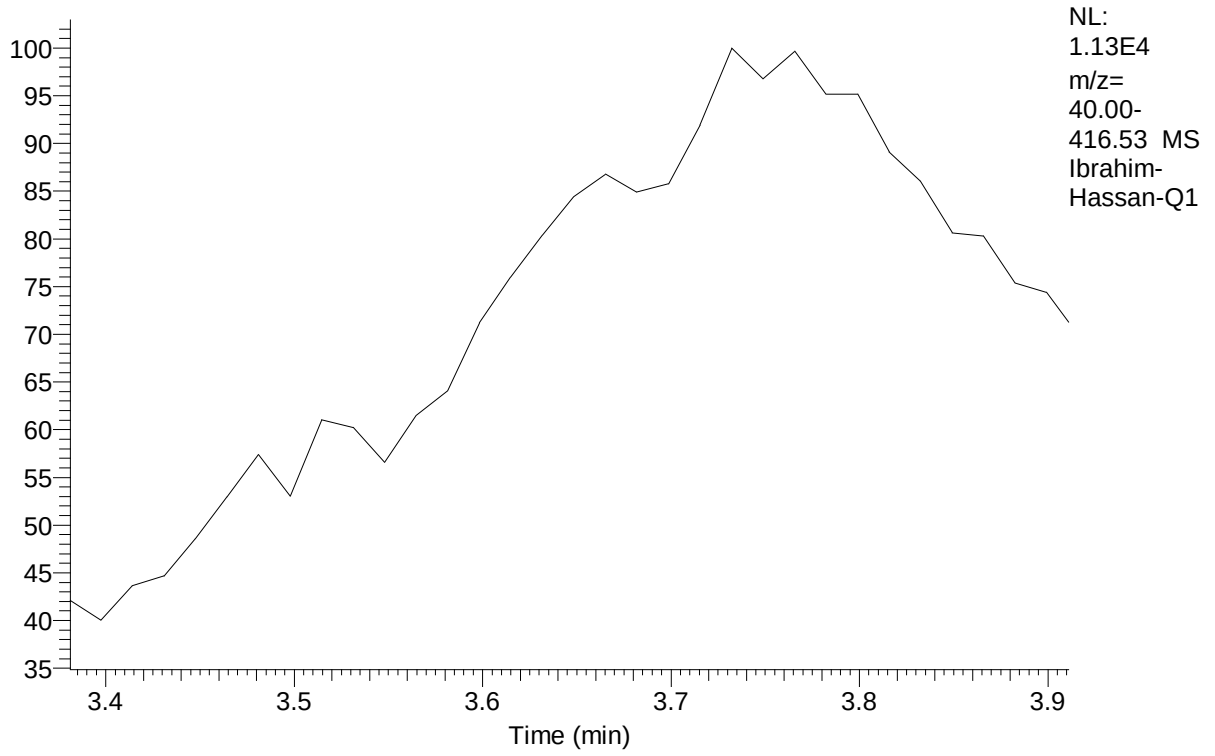
INVESTIGATOR **DIRECTOR**

جامعة الأزهر
المركز الإقليمي للفطريات وتطبيقاتها

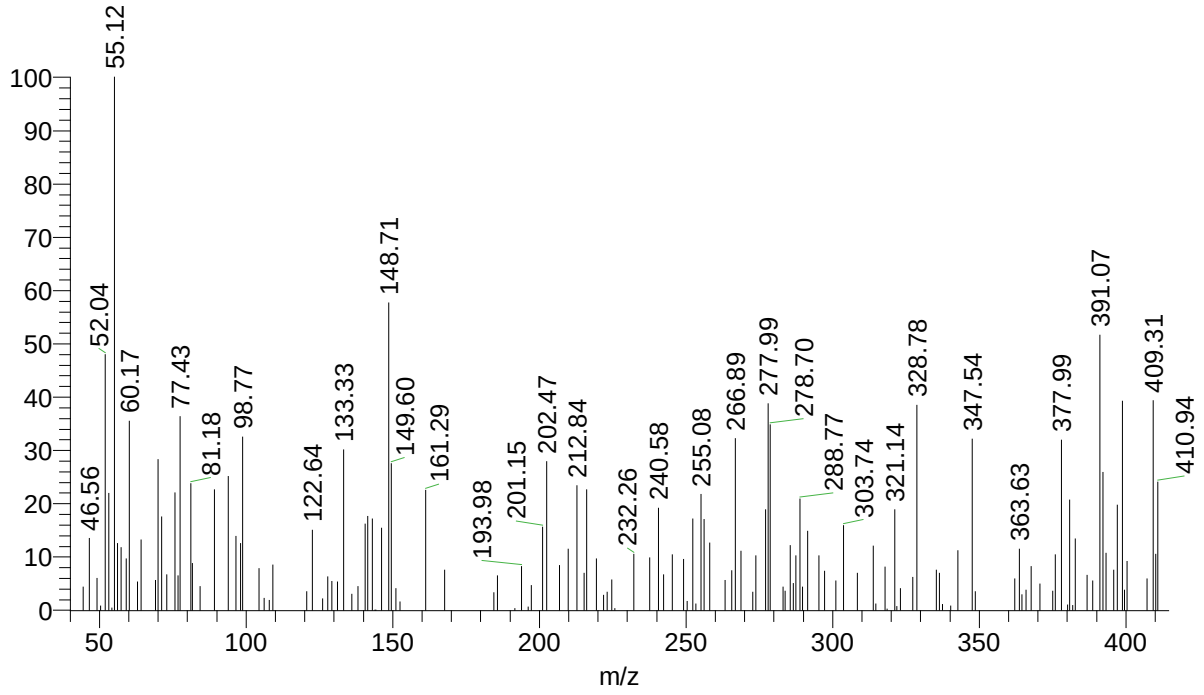
Al-Azhar University Campus - Nasr City, Cairo, Egypt.
Tel: 0202 22620373 Fax : 0202 22620373
E.mail:rcmb@azhar.edu.eg
Website: http://www.azhar.edu.eg.htm * http://www.azhar.edu.eg/pages/fungi_center.htm
Facebook : RCMB AZHAR P.O. box mail : 11751 Nasr City Cairo, Egypt.

Mass Spectroscopy of compound 7

RT: 3.38 - 3.91 SM: 11B



Ibrahim-Hassan-Q1 #157-161 RT: 2.64-2.71 AV: 5 SB: 26 1.21-1.34 , 0.87-1.14 NL: 1.76E2
T: + c EI Full ms [40.00-1000.00]



m/z	Intensity	Relative
44.50	7.7	4.35
46.56	23.7	13.45
49.10	10.5	5.98
50.40	1.4	0.79
52.04	84.3	47.95
53.26	38.6	21.95
54.28	0.7	0.41
55.12	175.9	100.00
56.11	22.0	12.53
57.47	20.7	11.80
59.16	17.0	9.67
60.17	62.4	35.46
63.06	9.3	5.30
64.18	23.2	13.18
69.12	9.8	5.58
70.05	49.7	28.25
71.25	30.9	17.57
72.88	11.8	6.70
75.81	38.8	22.05
76.85	11.5	6.53
77.43	63.9	36.35
81.18	41.8	23.74

81.68	15.5	8.79
84.30	7.9	4.48
89.14	39.7	22.60
93.88	44.1	25.07
96.47	24.3	13.84
98.08	22.0	12.53
98.77	57.2	32.55
104.44	13.8	7.84
106.24	4.0	2.25
107.99	3.3	1.89
109.22	15.0	8.52
120.75	6.1	3.50
122.64	26.4	15.02
126.06	3.9	2.20
127.84	11.0	6.27
129.24	9.5	5.43
131.20	9.5	5.38
133.33	52.9	30.10
136.00	5.3	3.01
138.10	7.8	4.46
140.67	28.4	16.14
141.49	31.0	17.64
143.09	30.2	17.19
143.85	0.1	0.04
146.28	27.1	15.44

148.71	101.5	57.71
149.60	48.4	27.54
151.10	7.1	4.04
152.43	2.8	1.59
161.29	39.7	22.57
167.63	13.3	7.54
184.46	5.8	3.32
185.77	11.5	6.52
191.61	0.5	0.30
193.98	14.4	8.21
196.24	1.1	0.65
197.20	8.1	4.63
201.15	27.5	15.65
202.47	49.0	27.86
206.87	14.8	8.39
209.81	20.2	11.46
212.84	41.2	23.42
215.31	12.3	6.98
216.16	39.8	22.65
219.38	17.0	9.68
221.95	5.0	2.86
223.11	6.0	3.39
224.68	10.1	5.72
225.75	0.7	0.38
232.26	18.5	10.52

237.61	17.3	9.81
240.58	33.8	19.19
242.41	11.7	6.65
245.40	18.4	10.47
249.14	16.8	9.53
250.31	2.9	1.68
252.26	30.2	17.19
253.44	2.2	1.24
255.08	38.3	21.78
256.20	29.9	17.03
258.15	22.3	12.66
263.36	9.9	5.61
265.64	13.1	7.42
266.89	56.7	32.21
268.74	19.5	11.08
272.75	6.1	3.44
273.81	18.1	10.26
277.05	33.3	18.92
277.99	68.2	38.79
278.70	61.2	34.80
283.02	7.7	4.40
283.86	6.3	3.56
285.60	21.4	12.15
286.54	8.8	5.02
287.50	18.0	10.25

288.77	36.7	20.86
289.79	7.7	4.37
291.43	26.1	14.87
295.36	17.9	10.19
297.31	12.9	7.36
301.14	9.7	5.53
303.74	27.9	15.86
308.41	12.2	6.95
313.84	21.2	12.07
314.77	2.2	1.22
317.89	14.2	8.10
318.56	0.4	0.26
321.14	33.2	18.88
321.96	1.2	0.69
323.04	7.2	4.09
327.36	10.9	6.21
328.78	67.7	38.49
335.36	13.2	7.50
336.48	12.2	6.96
337.37	2.0	1.13
340.18	1.4	0.82
342.76	19.7	11.21
347.54	56.5	32.10
348.59	6.1	3.46
362.02	10.4	5.92

363.63	20.2	11.49
364.55	5.2	2.94
365.88	6.6	3.77
367.76	14.5	8.24
370.58	8.6	4.91
375.00	6.4	3.61
375.95	18.4	10.44
377.99	56.1	31.92
380.03	1.8	1.01
380.87	36.5	20.75
381.77	1.5	0.87
382.65	23.6	13.40
386.78	11.6	6.62
388.69	9.8	5.56
391.07	90.8	51.66
392.19	45.5	25.88
393.28	18.9	10.75
395.89	13.2	7.49
397.12	34.8	19.77
398.78	68.9	39.21
399.56	6.7	3.83
400.33	16.1	9.16
407.14	10.3	5.86
409.31	69.2	39.36
410.16	18.5	10.54

410.94 42.2 24.02

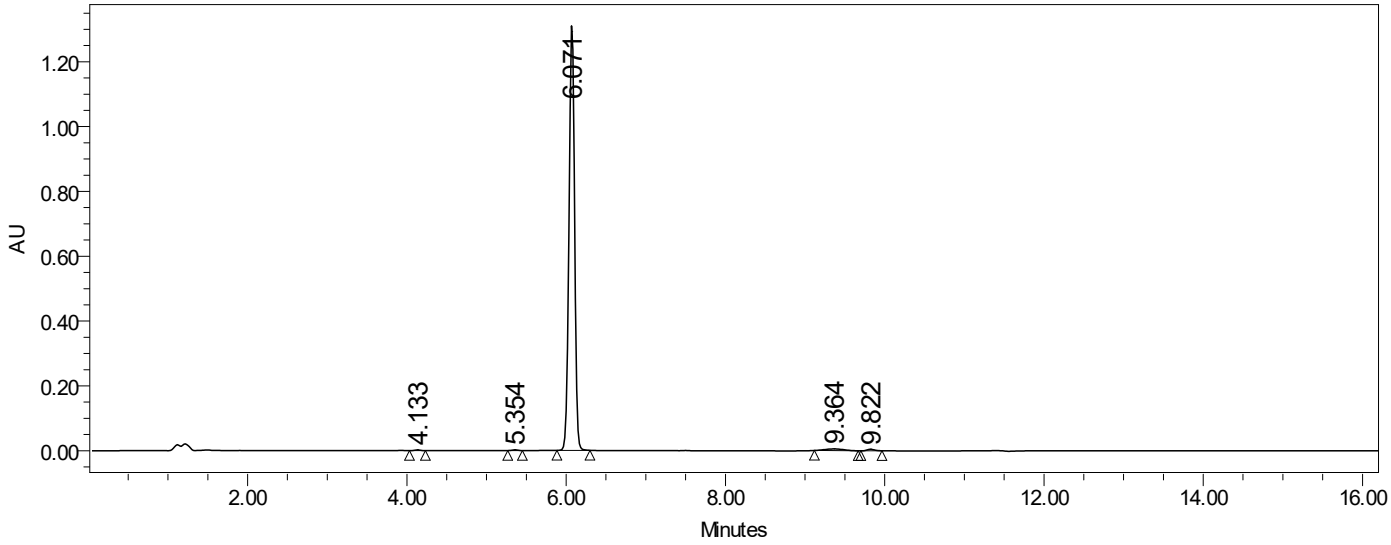
Empower™ 3 SOFTWARE

HPLC of compound 7

Multi Sample Summary

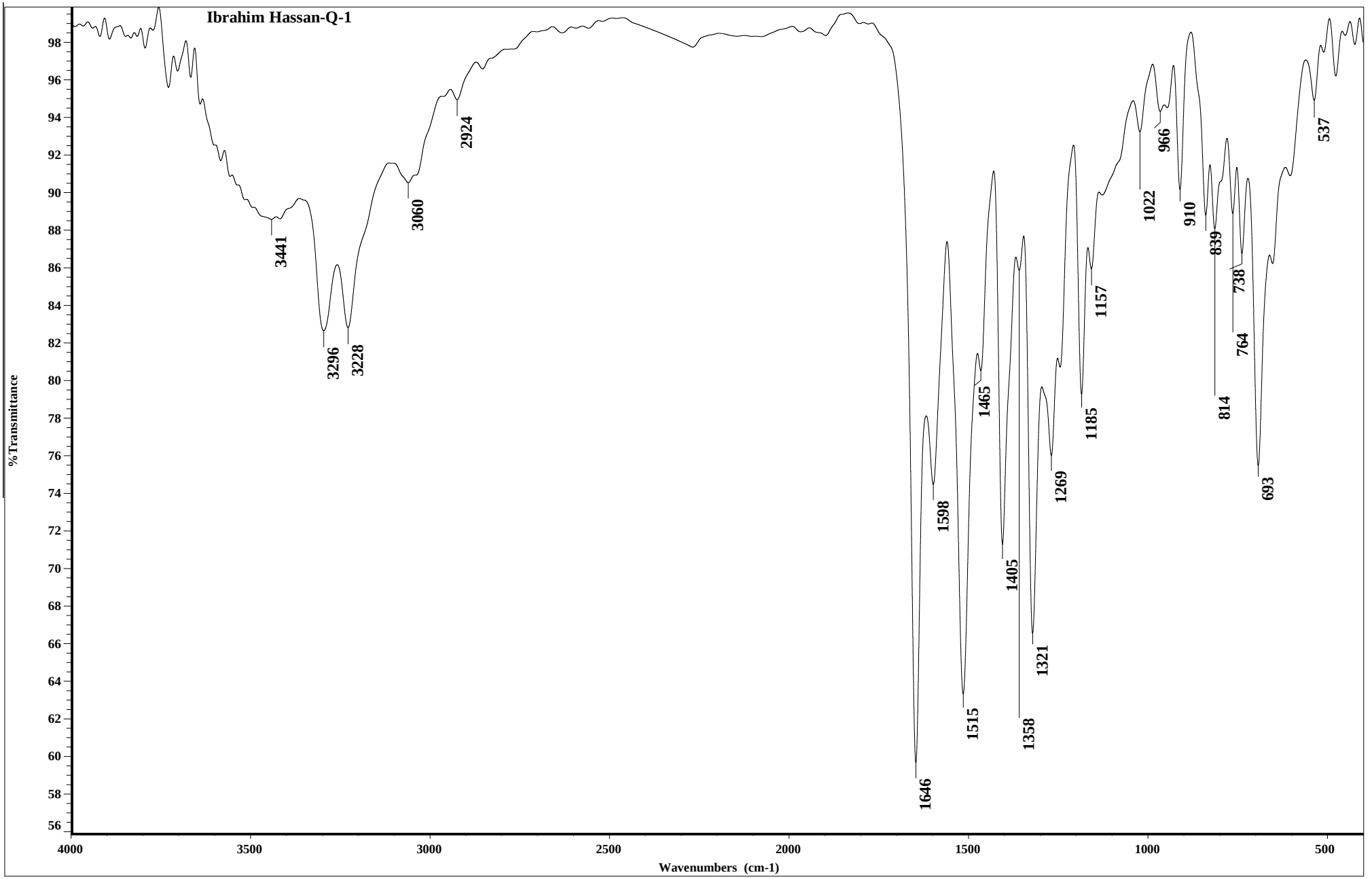
SAMPLE INFORMATION

Sample Name:	impurty 6/8	Acquired By:	System
Sample Type:	Unknown	Sample Set Name:	
Vial:	2	Acq. Method Set:	Organic
Injection #:	1	Processing Method:	Default
Injection Volume:	10.00 ul	Channel Name:	258.0nm@2
Run Time:	20.0 Minutes	Proc. Chnl. Descr.:	W2996 PDA 258.0 nm (PDA 190.0 to
Date Acquired:	8/6/2022 2:59:04 PM EET		
Date Processed:	8/6/2022 3:20:28 PM EET		



	RT	Area	% Area	Height
1	4.133	11163	0.17	2343
2	5.354	10626	0.16	2175
3	6.071	6384148	97.89	1298143
4	9.364	86364	1.32	5219
5	9.822	29504	0.45	4778

Spectral data of compound 7

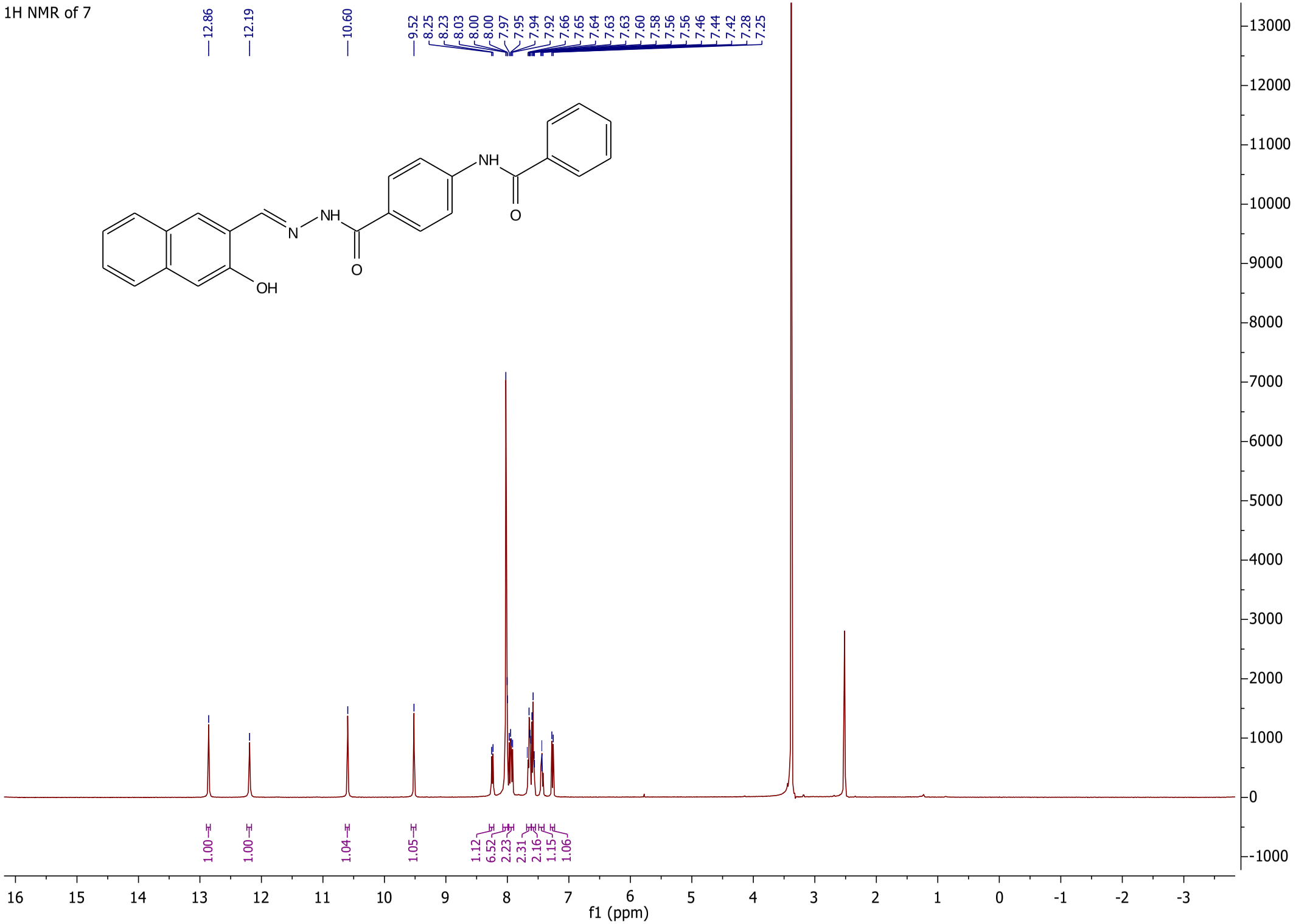


Number of sample scans: 32
Number of background scans: 32
Resolution: 8.000
Sample gain: 1.0
Optical velocity: 0.4747
Aperture: 150.00

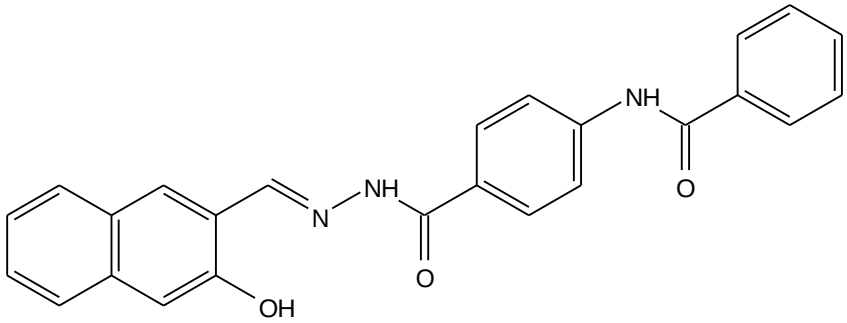
ThermoFisher SCIENTIFIC

Mon Jun 20 12:03:14 2022 (GMT+02:00)

Mansoura University
Faculty of Science
Spectral Analysis Unit
unitofspectra@gmail.com


1H NMR of 7
12.86
12.19
10.60
9.52
8.25
8.23
8.03
8.00
8.00
7.97
7.95
7.94
7.92
7.66
7.65
7.64
7.63
7.63
7.60
7.58
7.56
7.56
7.46
7.44
7.42
7.28
7.25
NH
NH
N
O
O
OH
1.00
1.00
1.04
1.05
1.12
6.52
2.23
2.31
2.16
1.15
1.06
f1 (ppm)

1H NMR of 7

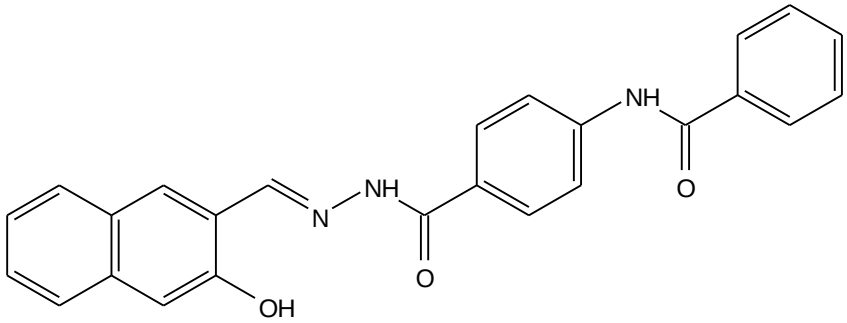


8.25 8.23 8.03 8.00 8.00 7.97 7.95 7.94 7.92 7.66 7.65 7.64 7.63 7.63 7.60 7.58 7.56 7.56 7.46 7.44 7.42 7.28 7.25

1.12 6.52 2.23 2.31 2.16 1.15 1.06

f1 (ppm)

1H NMR of 7



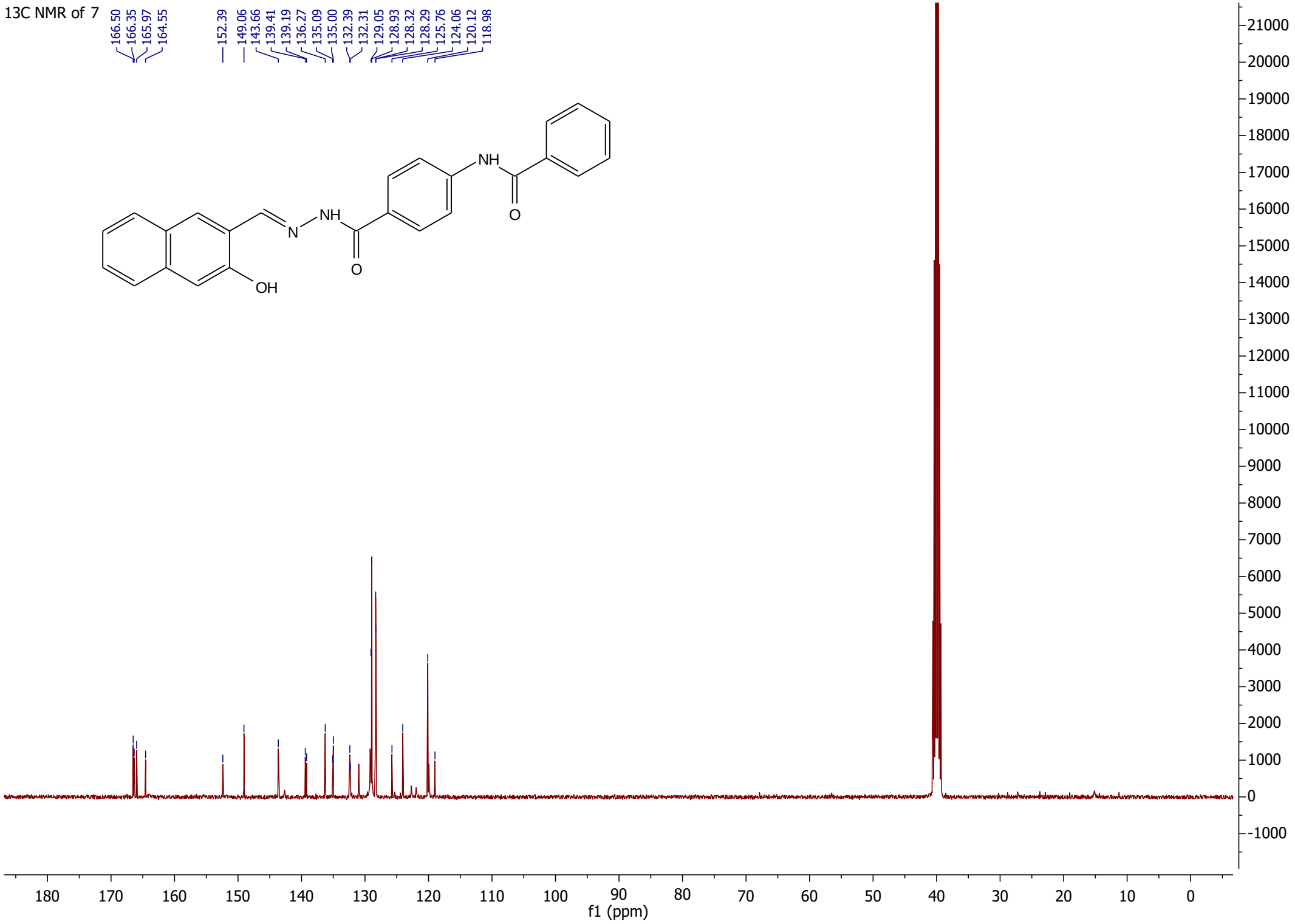
12.86
12.19
10.60
9.52

1.00
1.00
1.04
1.05

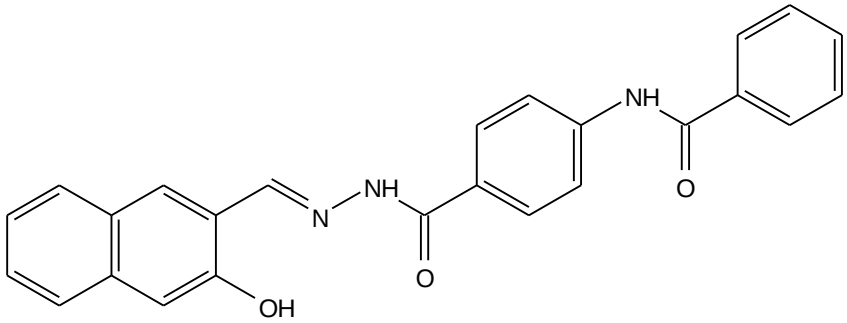
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000

4.4 14.2 14.0 13.8 13.6 13.4 13.2 13.0 12.8 12.6 12.4 12.2 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

f1 (ppm)

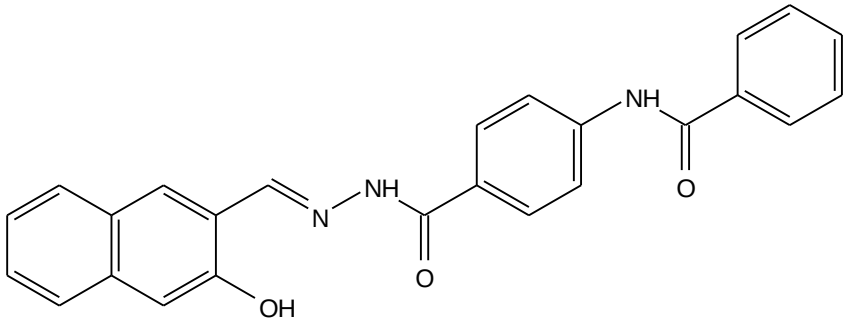

13C NMR of 7
166.50
166.35
165.97
164.55
152.39
149.06
143.66
139.41
139.19
136.27
135.09
135.00
132.39
132.31
129.05
128.93
128.32
128.29
125.76
124.06
120.12
118.98
NH
N
NH
O
O
OH
21000
20000
19000
18000
17000
16000
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

13C NMR of 7


NH
N
NH
O
O
OH

152.39
149.06
143.66
139.41
139.19
136.27
135.09
135.00
132.39
132.31
129.05
128.93
128.32
128.29
125.76
124.06
120.12
118.98
36000
34000
32000
30000
28000
26000
24000
22000
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
-2000
155
150
145
140
135
130
125
120
115
f1 (ppm)

13C NMR of 7



166.50
166.35
165.97
164.55
152.39
149.06
143.66
139.41
139.19
136.27
135.09
135.00

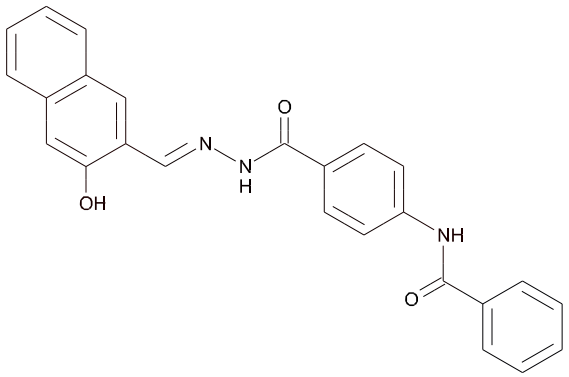
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000

178 176 174 172 170 168 166 164 162 160 158 156 154 152 150 148 146 144 142 140 138 136 134

f1 (ppm)

Toxicity Report

Compound 7



$C_{25}H_{19}N_3O_3$

Molecular Weight: 409.43665

ALogP: 4.134

Rotatable Bonds: 5

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 1.16

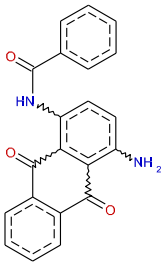
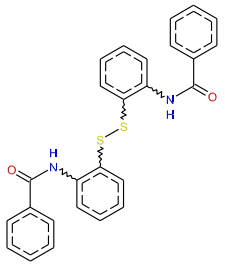
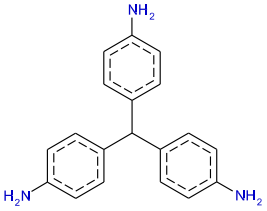
Mahalanobis Distance: 5.41

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.

TOPKAT_Ocular_Irritancy_None_vs_Irritant

Structural Similar Compounds

Name	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	BENZANILIDE;2';2'''-DITHIOBIS-	METHANE;TRIS(4-AMINOPHENYL)-
Structure			
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.713	0.714	0.716
Reference	28ZPAK-;124;72	28ZPAK-;173;72	28ZPAK-;73;72

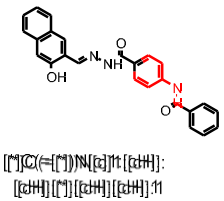
Model Applicability

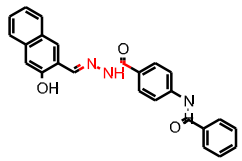
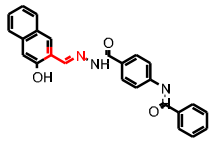
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 581019816: [*]N\N=C\[*]

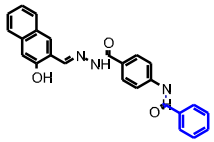
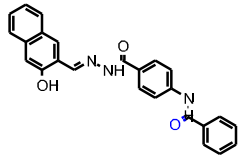
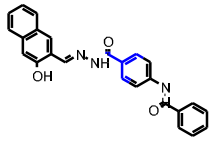
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1838187238	 [*]C(=[*])N[c]1[cH]:[cH][*][cH][cH]1	0.147	17 out of 18

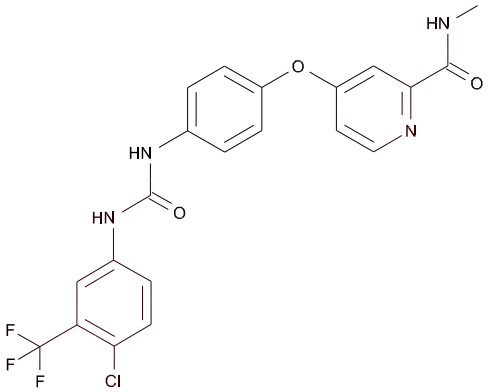
FCFP_12	-885520711	 [*]C(=[*])NN=[*]	0.137	2 out of 2
FCFP_12	-2100785893	 [*]\N=C\[c](:[*]):[*]	0.137	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382

Sorafenib

TOPKAT_Ocular_Irritancy_None_vs_Irritant



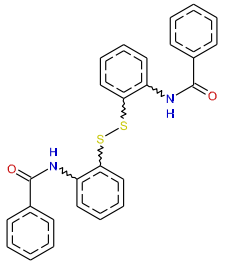
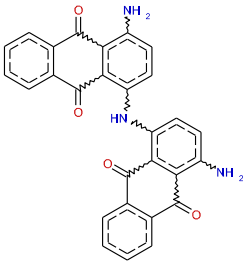
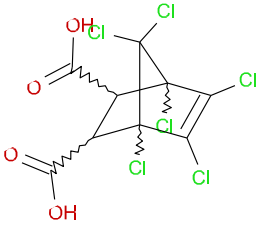
$C_{21}H_{16}ClF_3N_4O_3$
Molecular Weight: 464.82494
ALogP: 4.175
Rotatable Bonds: 6
Acceptors: 4
Donors: 3

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

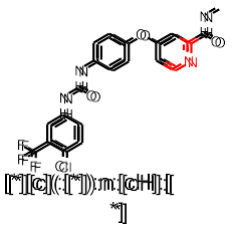
Model Applicability

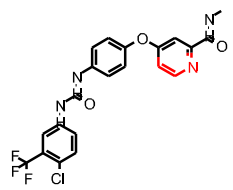
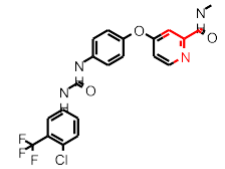
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

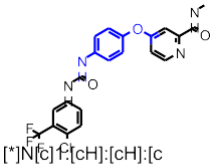
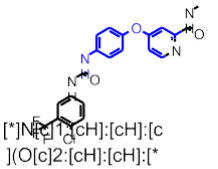
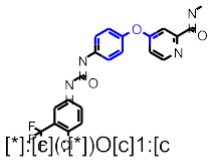
Feature Contribution

Top features for positive contribution

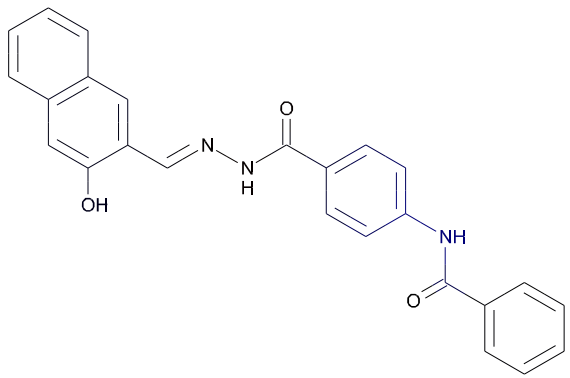
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384		0.208	44 out of 44

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-747629521	 [*]N[c]1:[cH]:[cH]:[c](O[c](:[*]):[*]):[cH]:[cH]:1	-0.268	1 out of 2
FCFP_12	702861189	 [*]N[c]1:[cH]:[cH]:[c](O[c]2:[cH]:[cH]:[*]:[c]([*]):[cH]:2):[cH]:[cH]:1	-0.268	1 out of 2
FCFP_12	-1270820019	 [*]:[c](:[*])O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0	7 out of 9

Compound 7



$C_{25}H_{19}N_3O_3$

Molecular Weight: 409.43665

ALogP: 4.134

Rotatable Bonds: 5

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Irritant

Probability: 0.74

Enrichment: 0.804

Bayesian Score: -3.78

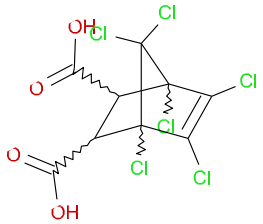
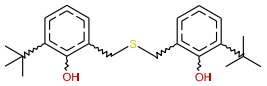
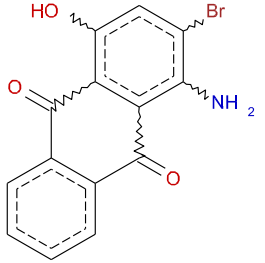
Mahalanobis Distance: 7.15

Mahalanobis Distance p-value: 0.993

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller 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	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	Sulfide, bis(4-t-butyl-m-cresyl)-	1-Amino-2-bromo-4-hydroxyanthraquinone
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.786	0.819	0.847
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	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	28ZPAK -,83,72

Model Applicability

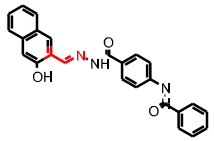
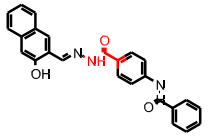
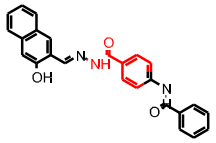
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

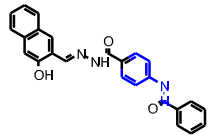
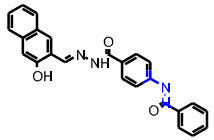
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

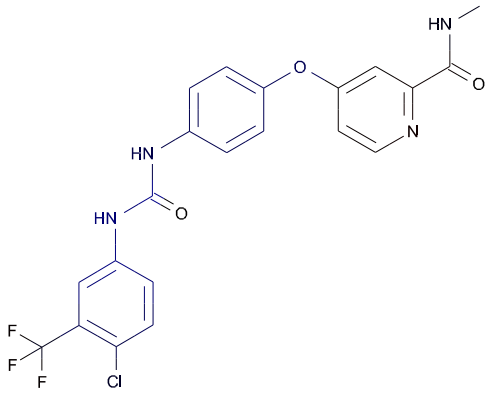
FCFP_12	-2100785893	 [*]\N=C\[c](:[*]):[*]	0.081	11 out of 11
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.0734	5 out of 5
FCFP_12	-581879738	 [*]NC(=O)[c]1:[cH]:[c H]:[*]:[cH]:[cH]:1	0.0658	3 out of 3

Top Features for negative contribution

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

Sorafenib

TOPKAT_Skin_Irritancy_None_vs_Irritant



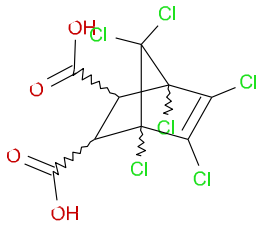
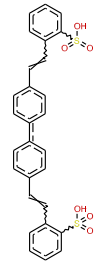
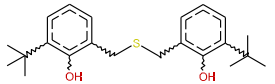
$C_{21}H_{16}ClF_3N_4O_3$
Molecular Weight: 464.82494
ALogP: 4.175
Rotatable Bonds: 6
Acceptors: 4
Donors: 3

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

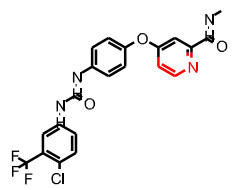
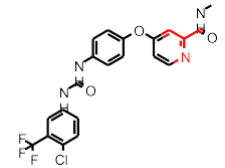
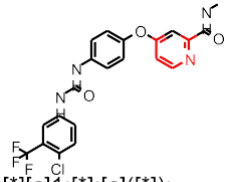
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

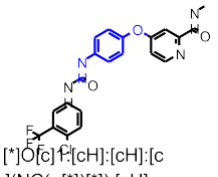
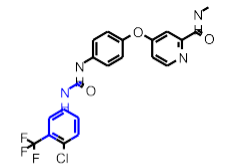
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

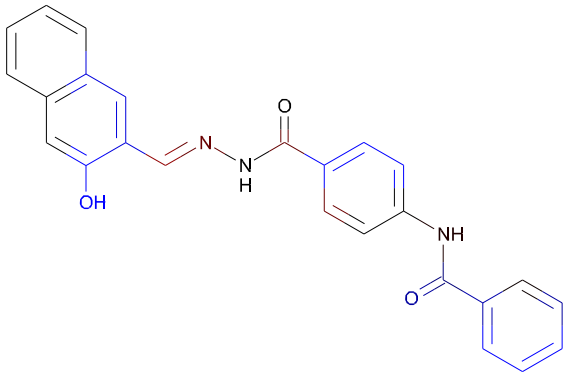
FCFP_12	-124655670	 [*]:[cH]:[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1539132615	 [*]C(=[*])[c](:[cH]:[*]):n:[*]	0.0795	9 out of 9
FCFP_12	-1695756380	 [*][c]1:[*]:[c]([*]):n:[cH]:[cH]:1	0.0772	7 out of 7

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]:[*]:[cH]:[cH]:1	-0.692	5 out of 12

Compound 7

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{25}H_{19}N_3O_3$

Molecular Weight: 409.43665

ALogP: 4.134

Rotatable Bonds: 5

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 38.6

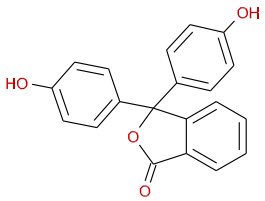
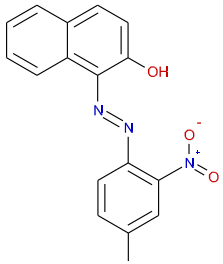
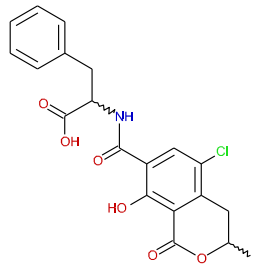
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.7

Mahalanobis Distance p-value: 8.17e-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	Phenolphthalein	646	Ochratoxin A
Structure			
Actual Endpoint (-log C)	2.43468	0.937339	4.79932
Predicted Endpoint (-log C)	3.66084	3.26294	3.6353
Distance	0.742	0.797	0.823
Reference	CPDB	CPDB	CPDB

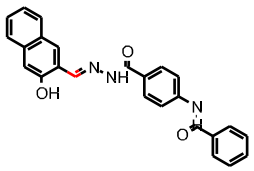
Model Applicability

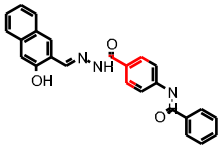
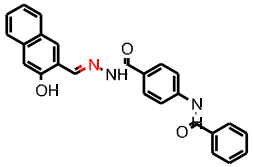
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

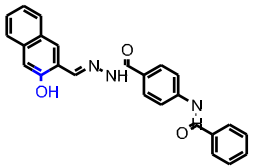
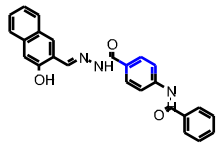
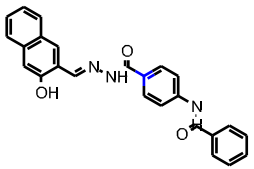
Feature Contribution

Top features for positive contribution

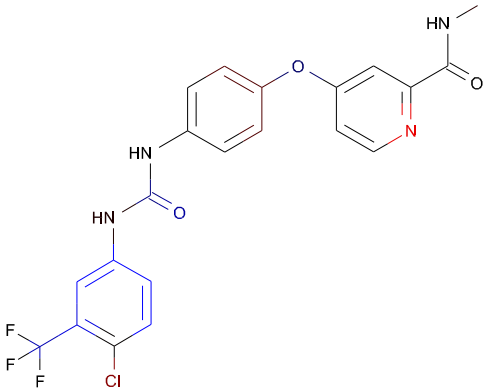
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-1925046727	 [*]C=[*]	0.145

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2019062761	 [*]:[c](:[*])O	-0.258
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

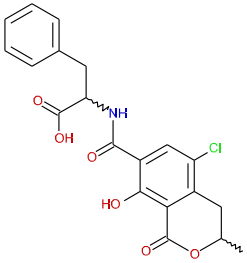
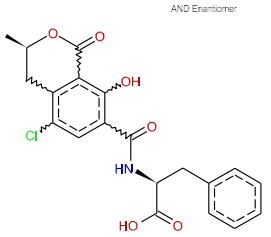
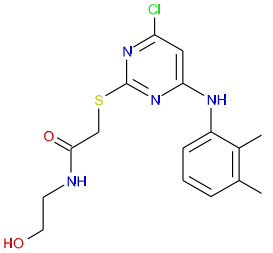
Model Prediction

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

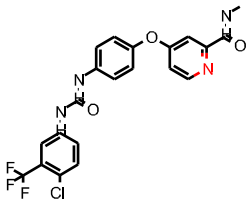
Model Applicability

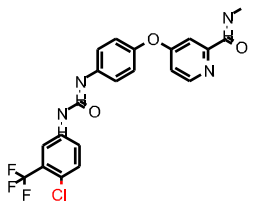
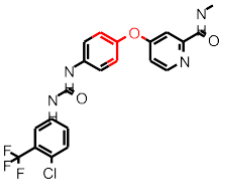
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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:[*]

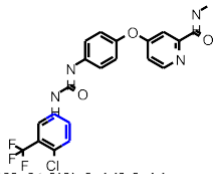
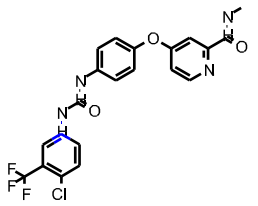
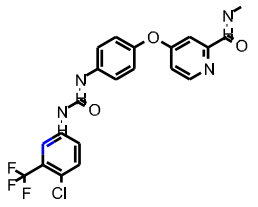
Feature Contribution

Top features for positive contribution

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

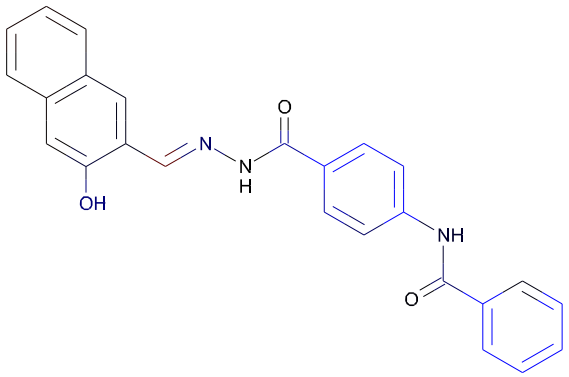
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-176455838	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.0818

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 7

TOPKAT_Chronic_LOAEL



$C_{25}H_{19}N_3O_3$

Molecular Weight: 409.43665

ALogP: 4.134

Rotatable Bonds: 5

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.36

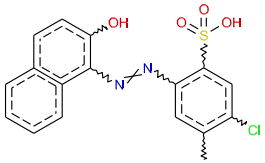
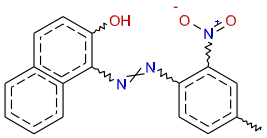
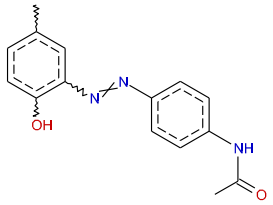
Unit: g/kg_body_weight

Mahalanobis Distance: 26.6

Mahalanobis Distance p-value: 2.06e-018

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

Structural Similar Compounds

Name	D & C RED 9	C.I. PIGMENT RED 3	HC BLUE 1
Structure			
Actual Endpoint (-log C)	3.87715	3.0252	3.0323
Predicted Endpoint (-log C)	3.6546	3.34768	2.7171
Distance	0.693	0.750	0.796
Reference	NTP REPORT # 225	NTP REPORT # 407	NTP REPORT # 222

Model Applicability

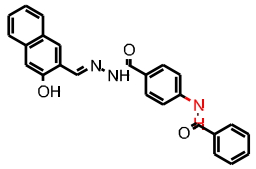
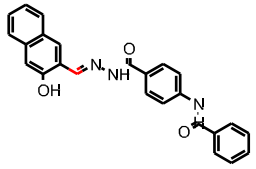
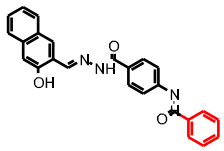
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
3. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
4. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
6. Unknown ECFP_6 feature: 544048674: [*]C(=[*])NN=[*]
7. Unknown ECFP_6 feature: 1814278164: [*]N\N=C\[*]
8. Unknown ECFP_6 feature: -1832102709: [*]\N=C\[c](:[*]):[*]
9. Unknown ECFP_6 feature: 1335702447: [*][c](:[*]):[c](C=[*]):[cH]:[*]
10. Unknown ECFP_6 feature: -178525456: [*]:[cH]:[c](:[cH]:[*]):[c](:[*]):[*]
11. Unknown ECFP_6 feature: 2019062761: [*]:[c](:[*])O

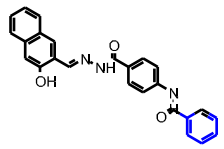
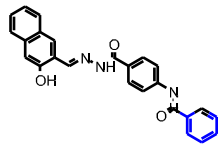
Feature Contribution

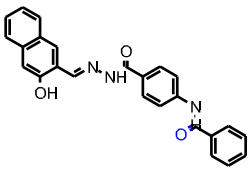
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	3	 [*]N[*]	0.0924
ECFP_6	-1925046727	 [*]C=[*]	0.0915
FCFP_6	-2093839777	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.078

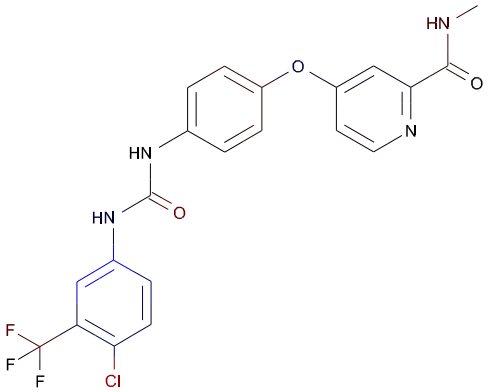
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.134
ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133

FCFP_6	1	 [*]=O	-

Sorafenib

TOPKAT_Chronic_LOAEL



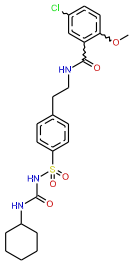
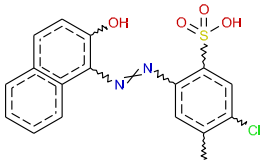
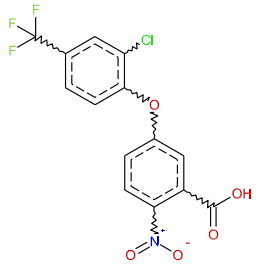
$C_{21}H_{16}ClF_3N_4O_3$
Molecular Weight: 464.82494
ALogP: 4.175
Rotatable Bonds: 6
Acceptors: 4
Donors: 3

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)

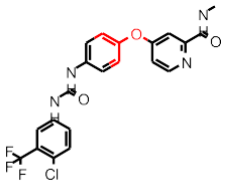
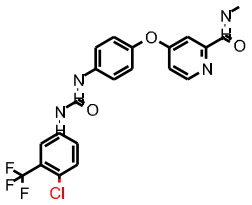
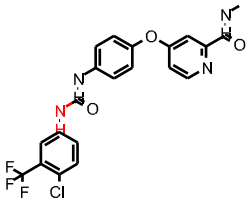
Model Applicability

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

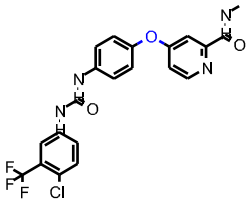
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -1046436026: [*]F
3. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
4. Unknown ECFP_6 feature: 226796801: [*]C([*])([*])F
5. Unknown ECFP_6 feature: 1305253718: [*]:[c](:[*])O[c](:[*]):[*]
6. Unknown ECFP_6 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

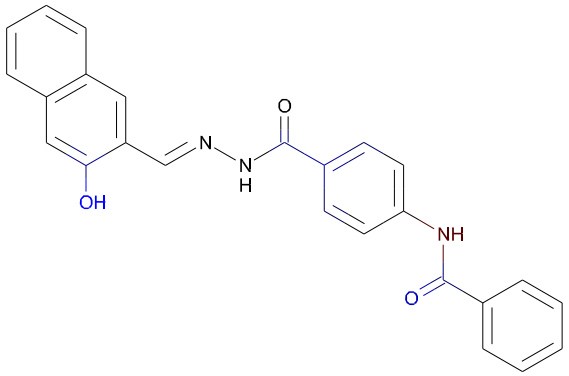
Top features for positive contribution

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

Compound 7



$C_{25}H_{19}N_3O_3$

Molecular Weight: 409.43665

ALogP: 4.134

Rotatable Bonds: 5

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.486

Unit: g/kg_body_weight

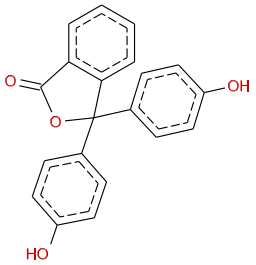
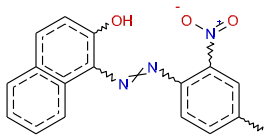
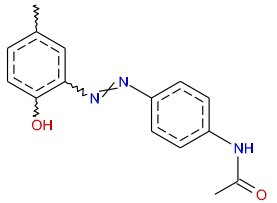
Mahalanobis Distance: 7.85

Mahalanobis Distance p-value: 0.0354

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_Feed

Structural Similar Compounds

Name	PHENOLPHTHALEIN	C.I.PIGMENT RED 3	DISPERSE YELLOW 3
Structure			
Actual Endpoint (-log C)	2.20184	2.65635	2.77703
Predicted Endpoint (-log C)	2.8857	2.97957	2.80195
Distance	0.637	0.694	0.756
Reference	NCI/NTP TR-465	NCI/NTP TR-407	NCI/NTP TR-222

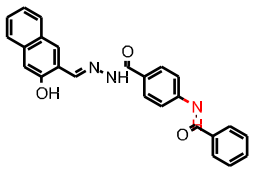
Model Applicability

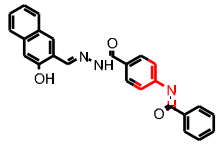
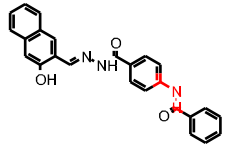
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

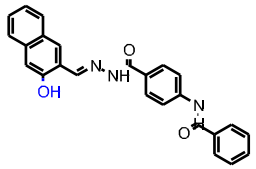
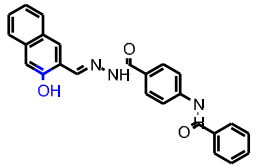
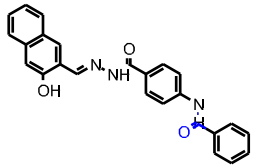
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	3	 [*]N[*]	0.0737

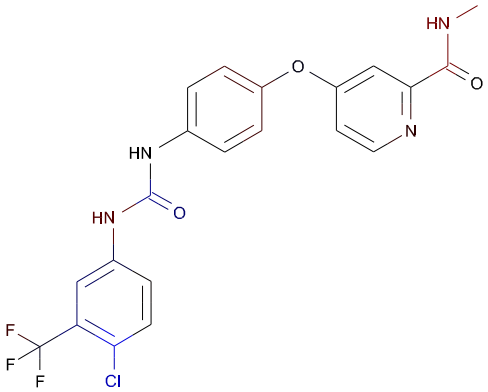
FCFP_2	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	0.00762
FCFP_2	1294255210	 [*]C(=[*])N[c](:[*]): [*]	0.00319

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	7	 [*]O	-0.214
FCFP_2	-549108873	 [*]:[c](:[*])O	-0.127
FCFP_2	1872154524	 [*]C(=O)[*]	-0.105

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



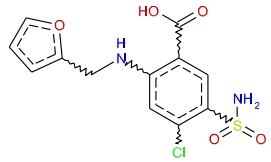
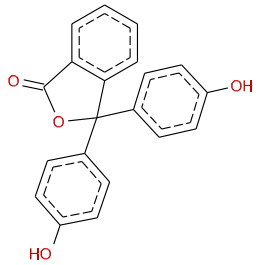
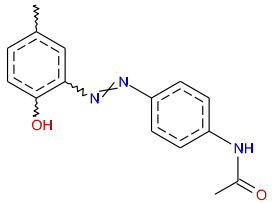
$C_{21}H_{16}ClF_3N_4O_3$
Molecular Weight: 464.82494
ALogP: 4.175
Rotatable Bonds: 6
Acceptors: 4
Donors: 3

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

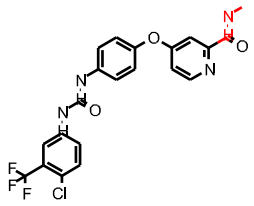
Model Applicability

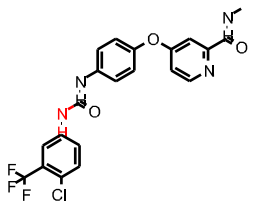
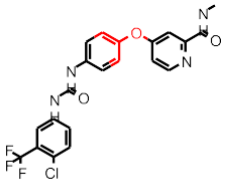
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

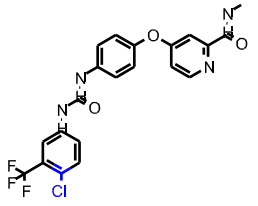
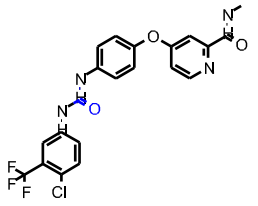
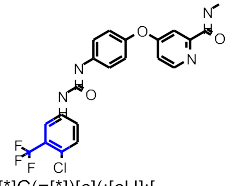
Feature Contribution

Top features for positive contribution

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

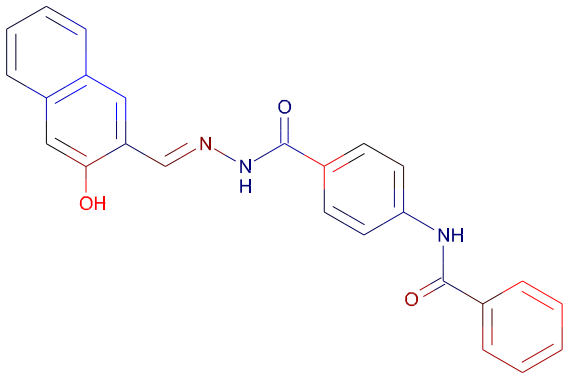
FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	332760439	 [*]O[c](:[cH]:[*]):[cH]:[*]	0.0611

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](:[cH]:[*]):[cH]:[*]	-0.0829

Compound 7

TOPKAT_Rat_Oral_LD50



$C_{25}H_{19}N_3O_3$

Molecular Weight: 409.43665

ALogP: 4.134

Rotatable Bonds: 5

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 1.79

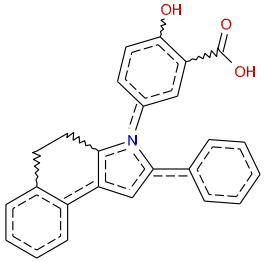
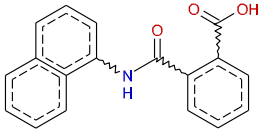
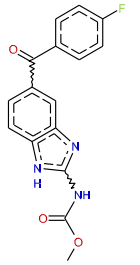
Unit: g/kg_body_weight

Mahalanobis Distance: 19.9

Mahalanobis Distance p-value: 1.24e-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	FENDOSAL	NAPTALAM	FLUBENDAZOLE
Structure			
Actual Endpoint (-log C)	2.928	1.551	2.088
Predicted Endpoint (-log C)	2.59	1.89036	2.69288
Distance	0.707	0.732	0.747
Reference	AGACBH 8;209;78	FMCHA2 -;C206;89	YRTMA6 9;11;78

Model Applicability

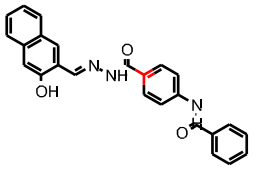
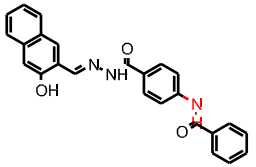
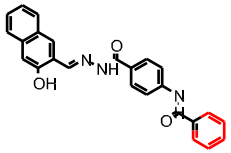
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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: -885520711: [*]C(=[*])NN=[*]
5. Unknown FCFP_6 feature: 581019816: [*]N\N=C\[*]
6. Unknown FCFP_6 feature: -2100785893: [*]\N=C\[c](:[*]):[*]
7. Unknown FCFP_6 feature: 74595001: [*][c](:[*]):[c](O):[cH]:[*]
8. Unknown FCFP_6 feature: -549108873: [*]:[c](:[*])O

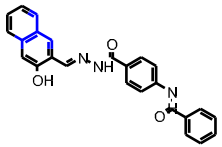
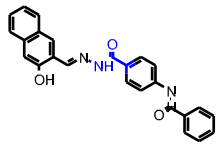
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

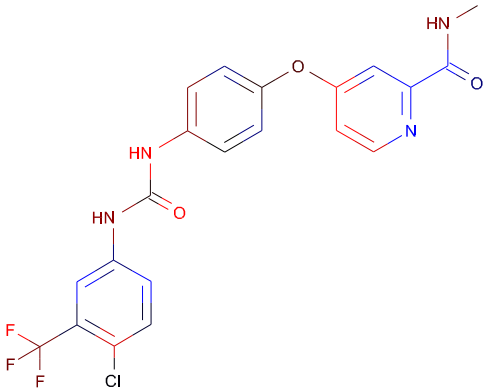
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
ECFP_6	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-178525456	 [*]:[cH]:[c](:[cH]:[*]):[c](:[*]):[*]	-0.157
FCFP_6	-1549103449	 [*]NC(=O)[c](:[*]):[*]	-0.117

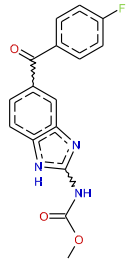
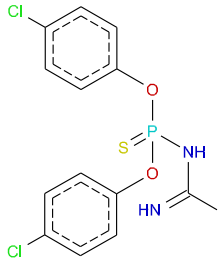
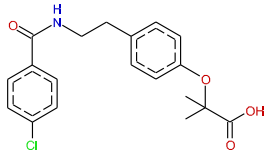
Sorafenib

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

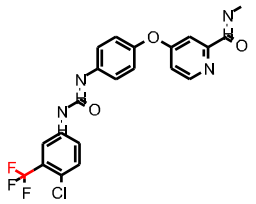
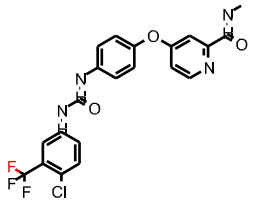
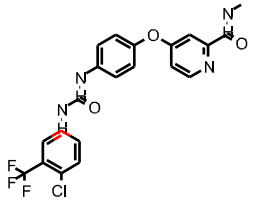
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

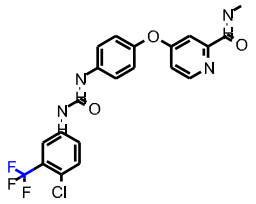
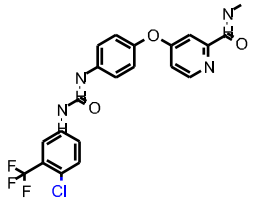
Feature Contribution

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