Electronic Supplementary Material (ESI) for RSC Advances.

Identification of new theobromine-based derivatives as potent VEGFR-2 inhibitors: Design, semi-synthesis, biological evaluation, and *in silico* studies

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

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

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

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

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

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

[f] Department of Chemistry and Biochemistry, University of Mississippi, MS 38677, USA

[g] Department of Agriculture Chemistry, Faculty of Agriculture, Damietta University, Damietta, Egypt.

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

[i] 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
1. Chemistry
2. Biological testing
3. *In silico* studies
4. Spectral data

S1. Chemistry

All commercially available materials were purchased from commercial sources and used without further purification. All reactions were performed in a temperature-controlled oil/wax bath. Reactions were monitored by analytical thin layer chromatography (TLC), using aluminum-backed plates, cut to size. TLC visualization was achieved by UV. ^{1}H NMR and ^{13}C NMR spectra were recorded on Bruker AVANCE 400 MHz spectrometers at ambient temperature. Spectral data is reported in ppm, coupling constants (*J*) in Hz. Elemental analyses were accomplished using a CHN analyzer. IR spectra were recorded as KBr pellets on a PerkinElmer Spectrum one FT-IR spectrometer. The mass spectra were recorded on an Agilent 6410 triple-quadrupole mass spectrometer equipped with an ESI source. All melting points were taken on a Gallenkamp melting point apparatus and are uncorrected. Nomenclature was determined using ChemBioDraw Ultra 14.0.

S2. Biological testing

S.2.1. In vitro anti-proliferative activity

The *in vitro* antiproliferative activities of all the synthesized compounds against a panel of four human tumor cell lines namely, liver (HepG2) and breast (MCF-7) cancer cells were evaluated quantitatively as described in the literature, using MTT assay protocol. The commercially available drug (sorafenib) was used in this test as positive controls. The anti-proliferative activity was assessed quantitatively as follows.

Human cancer cell lines were dropped in 96-well plates at a density of 3-8 × 10^3 cells/well. Next, the wells were incubated for 24 h in a 5% CO_2 incubator at 37 °C. Then, for each well, the growth medium was exchanged with 0.1 ml of fresh medium containing graded concentrations of the test compounds to be or equal DMSO and incubated for two days. Then 10 µl MTT solution (5 µg/ml) was added to each well, and the cells were incubated for

additional 4 h. The crystals of MTT-formazan were dissolved in 100 µl of DMSO; the absorbance of each well was measured at 490 nm using an automatic ELISA reader system (TECAN, CHE). The IC_{50} values were calculated using the nonlinear regression fitting models (Graph Pad, Prism Version 5). The data represented the mean of three independent experiments in triplicate and were expressed as means ± SD. The IC_{50} value was defined as the concentration at which 50% of the cells could survive.

S.2.2. Safety assay

The safety profile of compound **15a** was checked on one non-cancerous cell line (Vero cells) 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.

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

Inhibitory activity of compound **15a** (The most promising member) against VEGFR-2 was evaluated using Human VEGFR-2 ELISA kit (Enzyme-Linked Immunosorbent Assay). A specific antibody for VEGFR-2 was seeded on a 96-well plate and 100 µL of the standard solution or the tested compound was added, all were incubated at room temperature for 2.5 hours. Then washed, 100 µL of the prepared biotin antibody was added, then incubated at room temperature for additional 1 hour. Washed, 100 µL of streptavidin solution was added then incubated for 45 min. at room temperature. Washed again, 100 µL of TMB Substrate reagent was added and incubated for 30 min. at room temperature. 50 µL of the stop solution was added, then read at 450 nm immediately. The standard curve was drawn, concentrations on the X-axis and the absorbance on the Y-axis.

S.2.4. Apoptosis analysis

To detect the apoptosis induced by compound **15a**, HepG2 cells were seeded and incubated overnight and then treated with compound **15a** at concentrations of 0.76 µM for 72 h. DMSO was chosen as the negative control. After that, the cells were collected and washed with PBS two

successive times. The cells were exposed to centrifugation. Apoptosis detection kit (BD Biosciences, San Jose, CA) was used in this test. According to the manufacturer's protocol the cells were stained by Annexin V-FITC and propidium iodide (PI) in the binding buffer for 20min at room temperature in the dark. Using a flow cytometer, Annexin V-FITC and PI binding were analyzed. flowjo software was used to analyze the frequencies in all quadrants.

S.2.5. Quantitative Real Time Reverse-Transcriptase PCR technique (determination of caspase-3 and caspase-9)

The quantity of caspase-3, caspase-9 mRNA in control and HepG2 treated with compound **15a** was assessed by qRT-PCR . HepG2 cells were treated with compound **15a** at a concentration of 0.76 µM (IC_{50} value). Total RNA from vehicle-treated control (0.01% DMSO) and **15a**-treated HepG2 cells were extracted as-per the manufacturer instructions (RNeasy mini kit, Qiagen, Germany). After RNA extraction, cDNA was prepared using the Revert Aid First Strand cDNA Synthesis kit (Thermo Scientific, USA). Amplification of target cDNA for apoptosis markers and GAPDH [as a normalization (housekeeping) gene] was done using one-step RT-PCR SYBR® Green kit Master Mix (Bio-Rad Laboratories, USA) on Rotor-Gene Q real-time PCR thermal cycler instrument. cDNA (2 µl aliquots) was mixed with 1 µl of forward primer, 1 µl reverse primer, 10 µl master mixture, and the reaction volume was completed to 20 µl with nuclease-free water. All experiments were performed in triplicates.

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

Primers	
Casp3	: F 5’-GGAAGCGAATCAATGGACTCTGG -3’,
Casp3	: R 5'-GCATCGACATCTGTACCAGACC -3'.
Casp9	: F 5’-GTTTGAGGACCTTCGACCAGCT-3’,
Casp9	: R 5'-CAACGTACCAGGAGCCACTCTT-3'.
GAPDH	: F 5’- GTCTCCTCTGACTTCAACAGCG-3’
GAPDH	: R 5’- ACCACCCTGTTGCTGTAGCCAA-3’

S.2.6. Wound healing assay (Migration assay)

HepG2 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 **15a** treatment. After incubation at 37°C and 5.0% CO_2 for 24h, the cells were washed with PBS and the wounds distance was determined as the scratch width of the treated and untreated groups using ImageJ software.

***S3. In silico* studies**

S.3.1. Docking studies

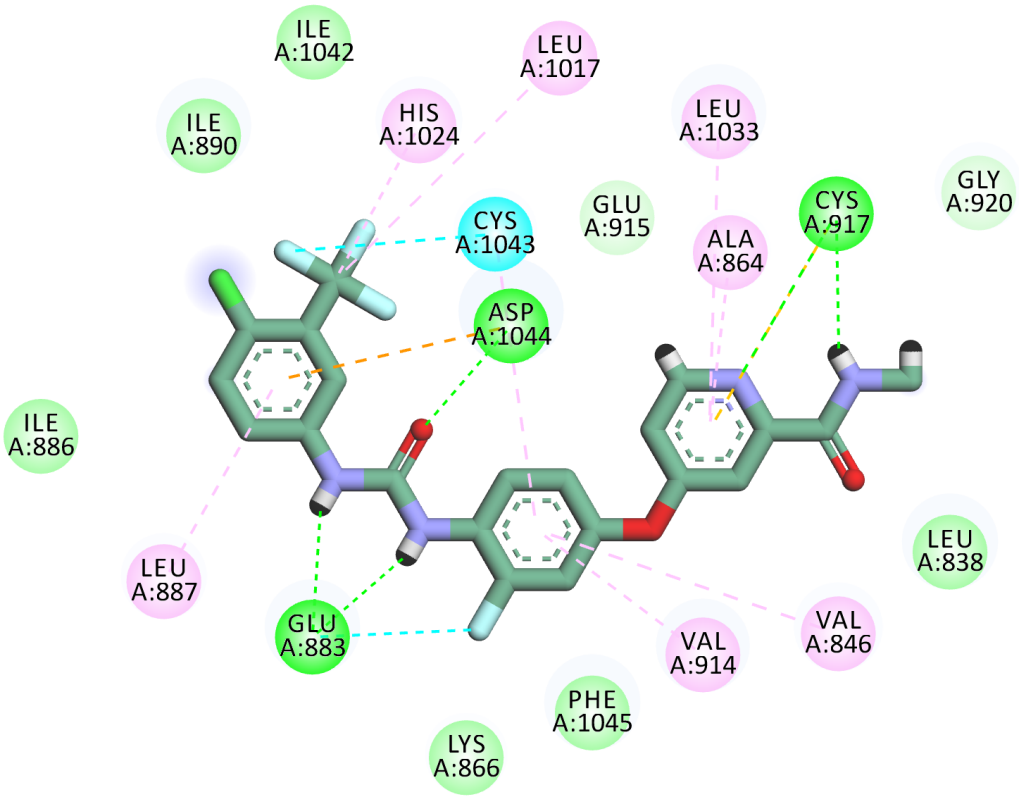
Protein Preparation: The crystal structure of VEGFR-2 [PDB ID: 2OH4, resolution: 2.05 Å] was obtained from Protein Data Bank (https://www.rcsb.org). At first, the crystal structure of the VEGFR-2 complexed with the co-crystallized ligand was prepared by removing crystallographic water molecules. Only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

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

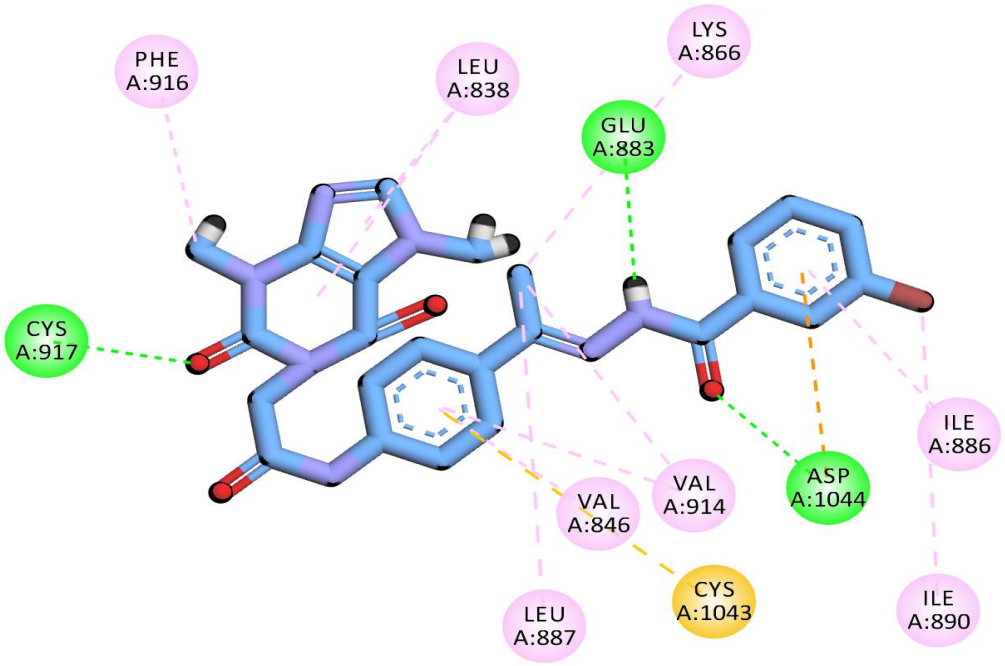
Docking Setup and Validation of Docking Protocol: The protein-ligand docking studies were carried out using MOE version 2019. Validation of the docking protocol was carried out by redocking the co-crystallized reference ligand against the isolated pocket of VEGFR-2. The

docking protocol was validated by comparing the heavy atoms RMSD value of the re-docked ligand pose with the corresponding co-crystallized reference ligand structure.

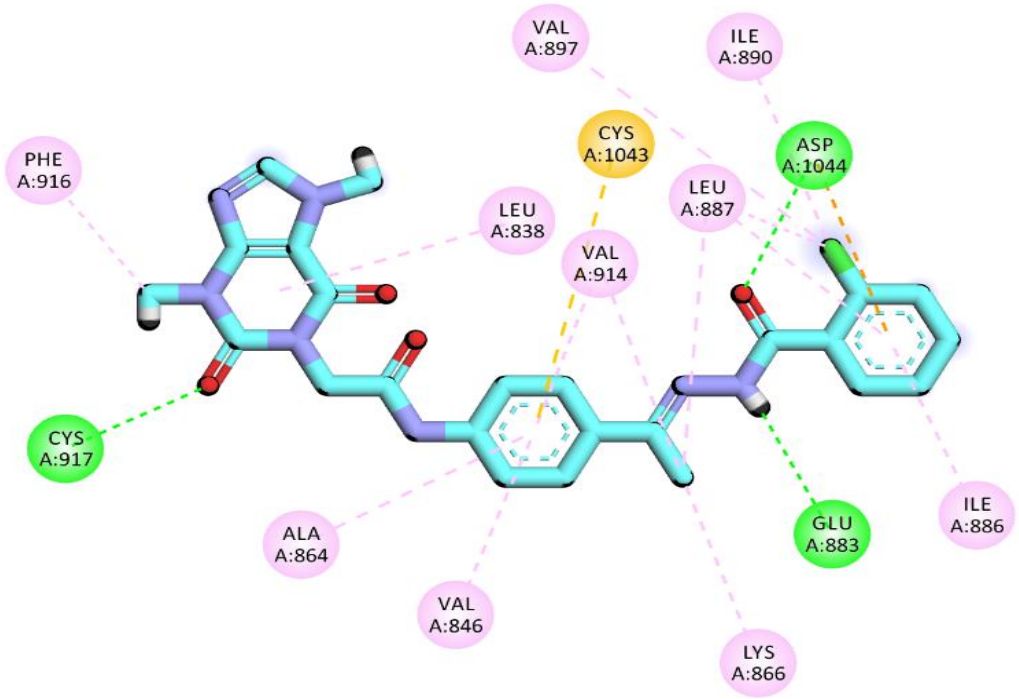
The docking setup for the tested compounds was established according to the protocol followed in the validation step. For each docking run, 30 docked solutions were generated using ASE for scoring function and rigid receptor for refinement. The pose with ideal binding mode was selected for further investigations. The docking results were visualized using Discovery Studio (DS) 4.0. Analysis of the docking results was carried out by comparing the interactions and docking score obtained for the docked ligands with that of the re-docked reference molecule.



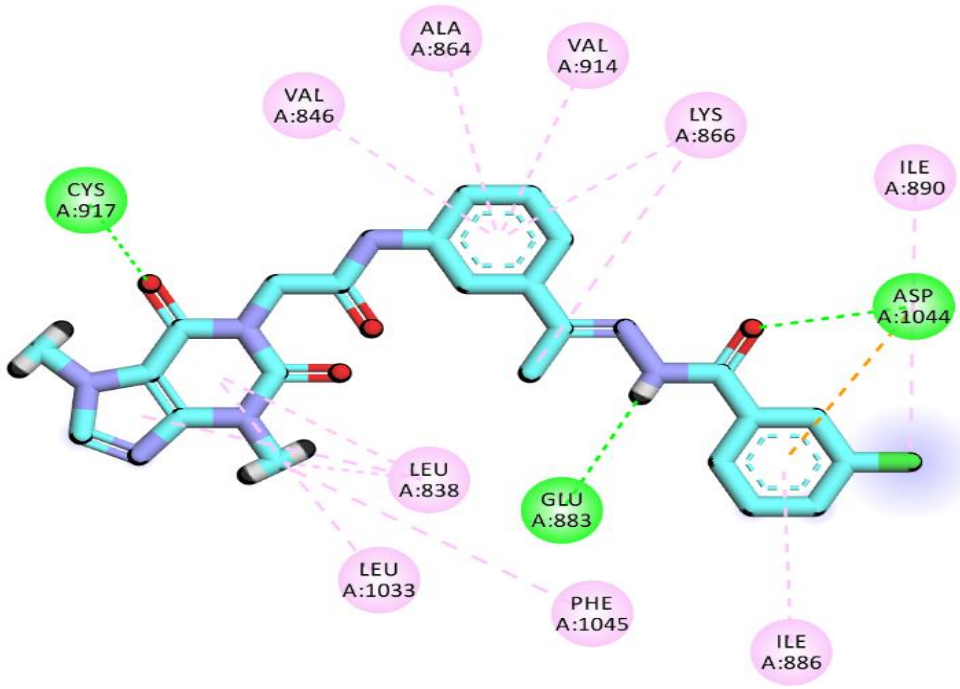
S3.1.1. 2D interaction of sorafenib with the active site of VEGFR-2.



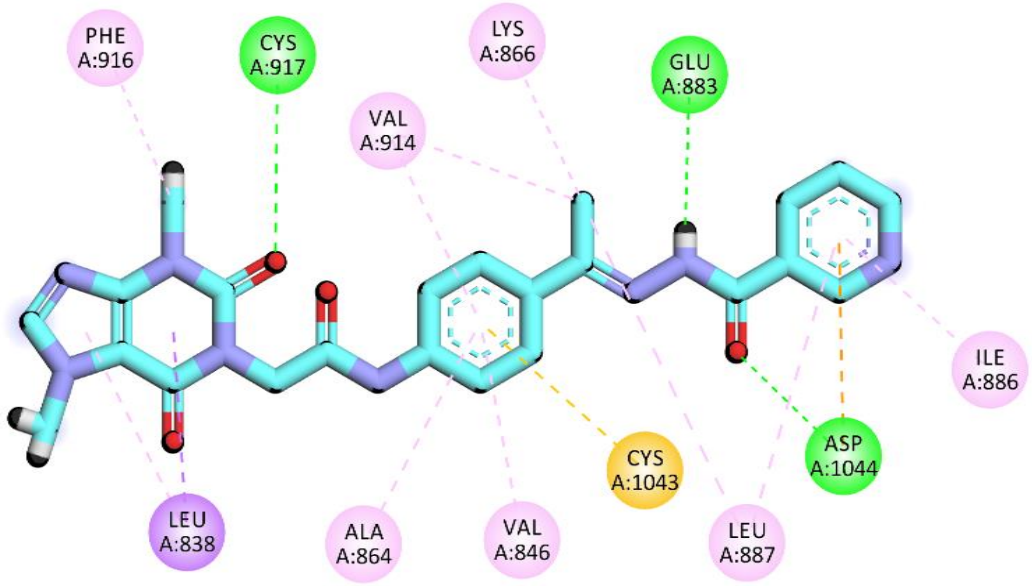
S3.1.2. 2D interaction of compound **15a** in the VEGFR-2 binding site.



S3.1.3. 2D diagram of compound **15b** in the VEGFR-2 binding site.



S3.1.4. 2D interactions of compound **16** in the VEGFR-2 binding site.



S3.1.5. 2D interactions of compound **17** in the VEGFR-2 binding site.

S.3.2. MD simulations

To assess the robustness of the VEGFR-2_**15a** complex and to compare it with the structure of the VEGFR-2_Sorafenib (reference molecule) and apo protein (the protein before docking), a 100 ns classical unbiased MD simulation was run in GROMACS 2021 for each system. We utilized the solution builder module of the CHARMM-GUI server to generate the input files. Using the transferable intermolecular potential 3 points (TIP3P) water model, we solvated and centered the each system in a cube of 8.3 nm in length with 1 nm of padding, and then neutralized the system with NaCl ions at a concentration of 0.154 M. The VEGFR-2 protein's amino acid parameters, the TIP3P water model, and the neutralizing ions were all obtained from the CHARMM36m force field. Parameters for the compound **15a** and Sorafenib were obtained with the help of the CHARMM general force field (CGenFF).

We utilized GROMACS 2021 to perform the dynamics and used periodic boundary conditions (PBCs) in all three spatial dimensions. The potential energy of the system was minimized so that atomic collisions may be avoided. During the equilibration process, the temperature was brought to 310 K and the pressure was brought to 1 atm. In detail, the minimizing step was set to be converged at 100,000 minimization steps or when the maximum force on any atom was 100 $KJ.mol^{-1}.nm^{-1}$. By using the Velocity Rescale technique, we were able to achieve thermal equilibrium in a canonical (NVT) ensemble. For pressure equilibration, we employed a Berendsen barostat with the isothermal-isobaric (NPT) ensemble. The production run in an NPT ensemble was started for 100 ns with the Nose-Hoover thermostat at 310 degrees Kelvin and the Parrinello-Rahman barostat set at 1 atm. The LINear Constraint Solver (LINCS) was used to impose length constraints on the hydrogen-bonded atoms. We utilized Particle Mesh Ewald (PME) to calculate the electrostatics with a threshold of 1.2 nm. By using a time step of 1 femtosecond during equilibration and 2 femtoseconds during production, the Newtonian equations of motion were integrated using the leap-frog algorithm. One thousand frames were captured at 0.1 ns intervals throughout the simulation. We centered the protein in the middle of the simulation box and made it whole again (removed the effects of the PBC) using the trjconv command and then analyzed it using VMD TK scripts. VEGFR-2, **15a,** and Sorafenib root mean square deviation (RMSD) values were determined. Root mean square fluctuation (RMSF), the radius of gyration (RoG), the

ligand-protein center of mass separation, and the number of hydrogen bonds between each ligand and the protein were all computed to see their dynamic behavior with time.

MM-GBSA

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

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

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

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

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

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

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

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

Where:

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

ProLIF Analysis

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

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

Principal Component Analysis:

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

Essential subspace size was determined based on the cumulative eigenvalues with respect to the number of eigenvectors used, where the variance maintained by the selected eigenvectors was shown. Additionally, the scree plot was made by plotting the eigenvalue of each eigenvector against its index number.

To make a direct comparison between the frames in the reduced subspace, we first merged the apo-protein and complex trajectories, then aligned them to the apo-protein configuration we obtained after equilibration, then constructed a new C matrix for the combined trajectories, and finally projected each trajectory onto the new C matrix. As a means of gauging the degree to which the two trajectories are similar, we plotted the projection on the first three eigenvectors by using different pair combinations of eigenvectors.

S.3.4. 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 (ΔN_{max}), global hardness (η), energy change (ΔE), electronegativity (χ), the global softness (σ), electrophilicity index (ω), ionization potential (IP) and electron affinity (EA)

$IP = -E_{HOMO}$

$EA = -E_{LUMO}$

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

$\eta = (IP-EA)$

$\chi = -\eta$

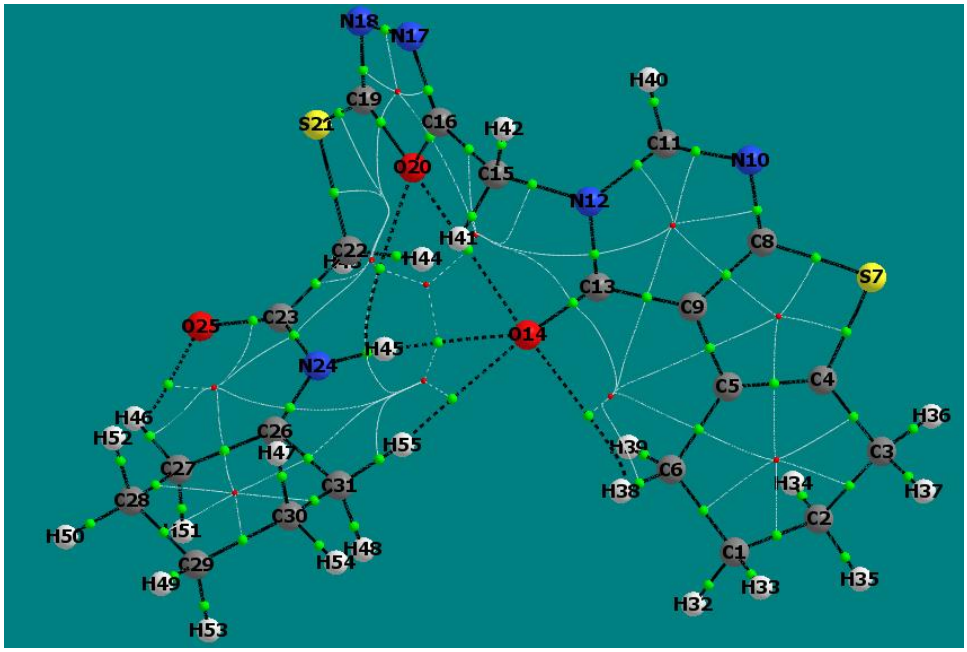
$\omega = \mu^2/(2\eta)$

$\sigma = 1/\eta$

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

$\Delta E = -\omega$

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



QTAIM analysis: molecular graph showing the bonding critical points (BCPs) for compound **15a**.

The QTAIM parameters (a.u.) at bond critical points (BCPs) of **15a**.

BCP #	Atoms	(ρ)	($\nabla^2\rho$)	K(r)	G(r)	V(r)	H(r)
1	N1 - C2	0.352145	-1.06484	0.496979	0.23077	-0.72775	-0.49698
2	C2 - N3	0.317941	-0.73416	0.457061	0.273522	-0.73058	-0.45706
3	N1 - C5	0.335873	-1.00592	0.435063	0.183584	-0.61865	-0.43506
4	C5 - C6	0.323138	-0.91608	0.342533	0.113513	-0.45605	-0.34253
5	N4 - C5	0.306468	-0.86102	0.411731	0.196475	-0.60821	-0.41173
6	N3 - C6	0.295186	-0.69173	0.399687	0.226754	-0.62644	-0.39969
7	C6 - C7	0.293145	-0.79626	0.280738	0.081672	-0.36241	-0.28074
8	C7 - N8	0.277213	-0.73348	0.323491	0.140121	-0.46361	-0.32349
9	N4 - C9	0.312245	-0.91964	0.40409	0.17418	-0.57827	-0.40409
10	N8 - C9	0.298059	-0.84511	0.359575	0.148298	-0.50787	-0.35958
11	C7 - O10	0.408671	-0.31429	0.688523	0.60995	-1.29847	-0.68852
12	C9 - O11	0.401887	-0.42566	0.672446	0.56603	-1.23848	-0.67245
13	N8 - C12	0.244313	-0.57819	0.276931	0.132382	-0.40931	-0.27693
14	C13 - N14	0.311455	-0.87859	0.420546	0.2009	-0.62145	-0.42055
15	C12 - H38	0.288802	-1.01713	0.288357	0.034074	-0.32243	-0.28836
16	C12 - C13	0.243593	-0.55834	0.191945	0.05236	-0.24431	-0.19195
17	N14 - H40	0.335076	-1.7095	0.477909	0.050535	-0.52844	-0.47791
18	O11 - H40	0.02629	0.095956	-0.0021	0.021886	-0.01978	0.002102

19	C15 - C23	0.303551	-0.83814	0.307231	0.097696	-0.40493	-0.30723
20	N14 - C15	0.285804	-0.78105	0.36808	0.172817	-0.5409	-0.36808
21	C13 - O16	0.407738	-0.23735	0.686246	0.626909	-1.31315	-0.68625
22	O16 - H50	0.017103	0.062213	-0.00209	0.013459	-0.01137	0.002094
23	N4 - C17	0.245777	-0.57459	0.293121	0.149474	-0.4426	-0.29312
24	N3 - C18	0.249278	-0.59364	0.301776	0.153366	-0.45514	-0.30178
25	C15 - C19	0.306358	-0.84787	0.310026	0.098059	-0.40809	-0.31003
26	C19 - C20	0.310831	-0.86724	0.319965	0.103156	-0.42312	-0.31997
27	C22 - C23	0.31167	-0.87322	0.321575	0.10327	-0.42485	-0.32158
28	H48 - H51	0.012224	0.04628	-0.00223	0.009336	-0.0071	0.002234
29	C20 - C21	0.303371	-0.82866	0.304111	0.096947	-0.40106	-0.30411
30	C21 - C22	0.302991	-0.82986	0.303181	0.095716	-0.3989	-0.30318
31	C21 - C24	0.265962	-0.6608	0.230706	0.065507	-0.29621	-0.23071
32	C23 - H50	0.28812	-1.01506	0.289253	0.035488	-0.32474	-0.28925
33	C24 - N25	0.363201	-0.71157	0.565109	0.387217	-0.95233	-0.56511
34	N25 - N27	0.353768	-0.65735	0.336922	0.172583	-0.50951	-0.33692
35	C22 - H49	0.285066	-0.99253	0.284507	0.036374	-0.32088	-0.28451
36	C24 - C26	0.249004	-0.57724	0.203825	0.059514	-0.26334	-0.20383
37	C26 - H53	0.270312	-0.88389	0.263739	0.042766	-0.30651	-0.26374
38	N27 - H54	0.336355	-1.61974	0.457269	0.052333	-0.5096	-0.45727
39	N27 - C28	0.304684	-0.86629	0.404129	0.187557	-0.59169	-0.40413
40	N25 - H55	0.013862	0.050995	-0.00216	0.010591	-0.00843	0.002158
41	C28 - C29	0.260182	-0.64141	0.220781	0.060428	-0.28121	-0.22078
42	C28 - O30	0.40493	-0.29827	0.680068	0.605501	-1.28557	-0.68007
43	C29 - C31	0.30523	-0.83908	0.308267	0.098498	-0.40677	-0.30827
44	C31 - C32	0.309679	-0.86665	0.317097	0.100434	-0.41753	-0.3171
45	C32 - C33	0.308339	-0.85851	0.315388	0.100761	-0.41615	-0.31539
46	C29 - C35	0.304005	-0.83088	0.305946	0.098225	-0.40417	-0.30595
47	C33 - C34	0.309655	-0.86308	0.318268	0.102497	-0.42077	-0.31827
48	C34 - C35	0.312427	-0.87571	0.324143	0.105215	-0.42936	-0.32414
49	C34 - Br36	0.154795	-0.14054	0.086295	0.051162	-0.13746	-0.0863
50	C2 - H37	0.288373	-1.02535	0.288501	0.032163	-0.32067	-0.2885
51	C12 - H39	0.281891	-0.96316	0.278282	0.037493	-0.31578	-0.27828
52	C17 - H41	0.28166	-0.9653	0.278541	0.037216	-0.31576	-0.27854
53	C17 - H42	0.284628	-0.98761	0.283108	0.036206	-0.31931	-0.28311
54	C17 - H43	0.280472	-0.95647	0.276769	0.037652	-0.31442	-0.27677
55	C18 - H44	0.280738	-0.96028	0.278301	0.03823	-0.31653	-0.2783
56	C18 - H45	0.282496	-0.97264	0.279488	0.036329	-0.31582	-0.27949
57	C18 - H46	0.282068	-0.96947	0.278823	0.036454	-0.31528	-0.27882
58	C19 - H47	0.279876	-0.9535	0.278226	0.039851	-0.31808	-0.27823
59	C20 - H48	0.282203	-0.96872	0.282202	0.040023	-0.32223	-0.2822
60	C26 - H51	0.278384	-0.93551	0.2765	0.042623	-0.31912	-0.2765
61	C26 - H52	0.270335	-0.88329	0.264441	0.04362	-0.30806	-0.26444

62	C31 - H55	0.28726	-1.00754	0.288168	0.036282	-0.32445	-0.28817
63	C32 - H56	0.282203	-0.97338	0.281122	0.037777	-0.3189	-0.28112
64	C33 - H57	0.283715	-0.98504	0.282892	0.036631	-0.31952	-0.28289
65	C35 - H58	0.285885	-1.00154	0.285163	0.034777	-0.31994	-0.28516

S.3.5. ADMET studies

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

- **Preparation of the tested compounds:**

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

- **Running of ADMET protocol**

In this protocol, the small molecules panel was utilized with the activation of the ADMET descriptors option. Then, we selected the prepared compounds as the input ligands. Further, all the ADMET parameters (aqueous solubility, Blood brain barrier, intestinal absorption, CYP2D6, and plasma protein binding) were selected. Then, the output of the running protocol was visualized to give the ADMET chart.

S.3.6 Toxicity studies

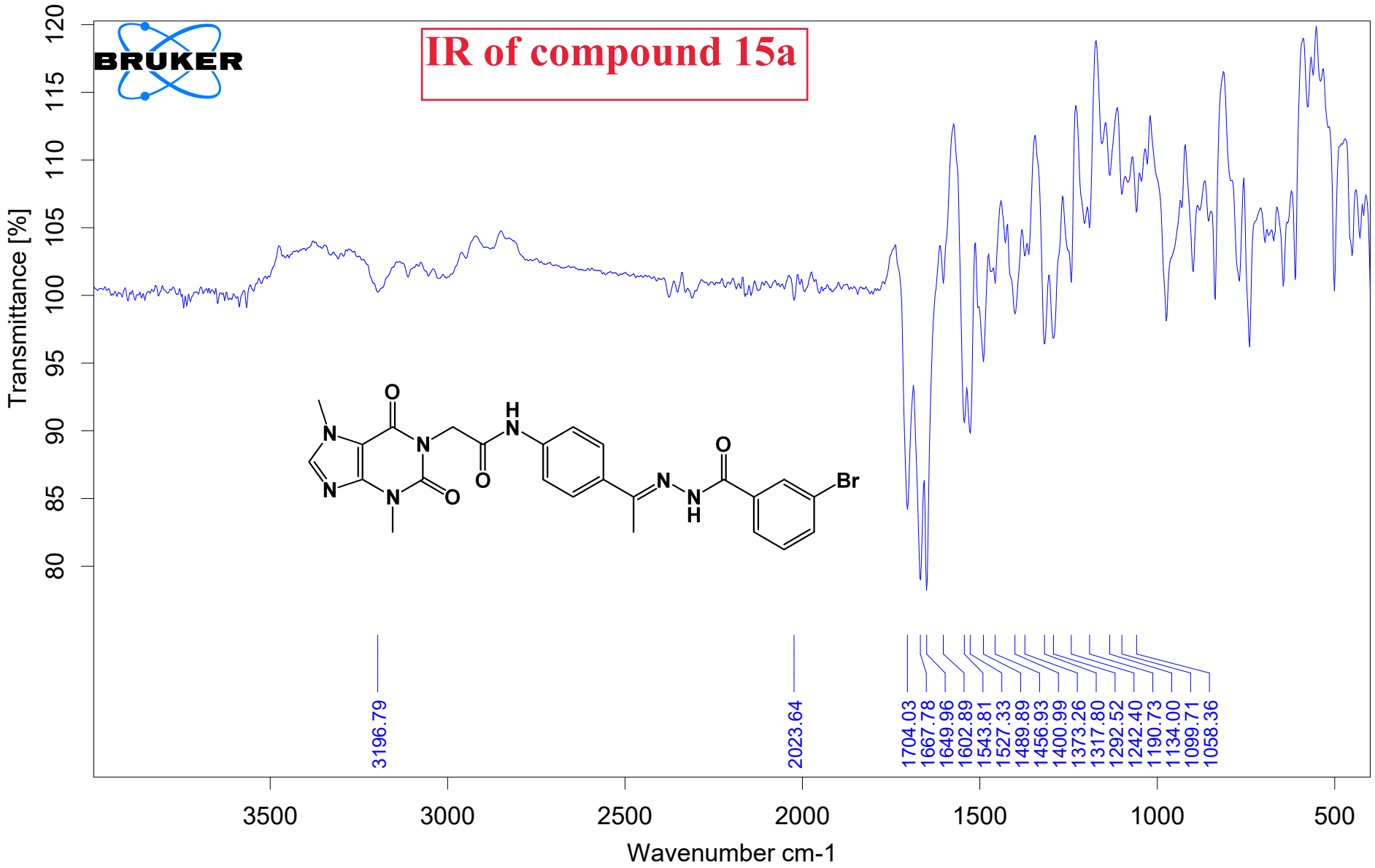
The toxicity parameters of the synthesized compounds were calculated using Discovery studio 4.0. Sorafenib was used as a reference molecule. At first, the CHARMM force field was applied then the compounds were prepared and minimized according to the preparation of small molecule protocol.

- **Preparation of the tested compounds:**

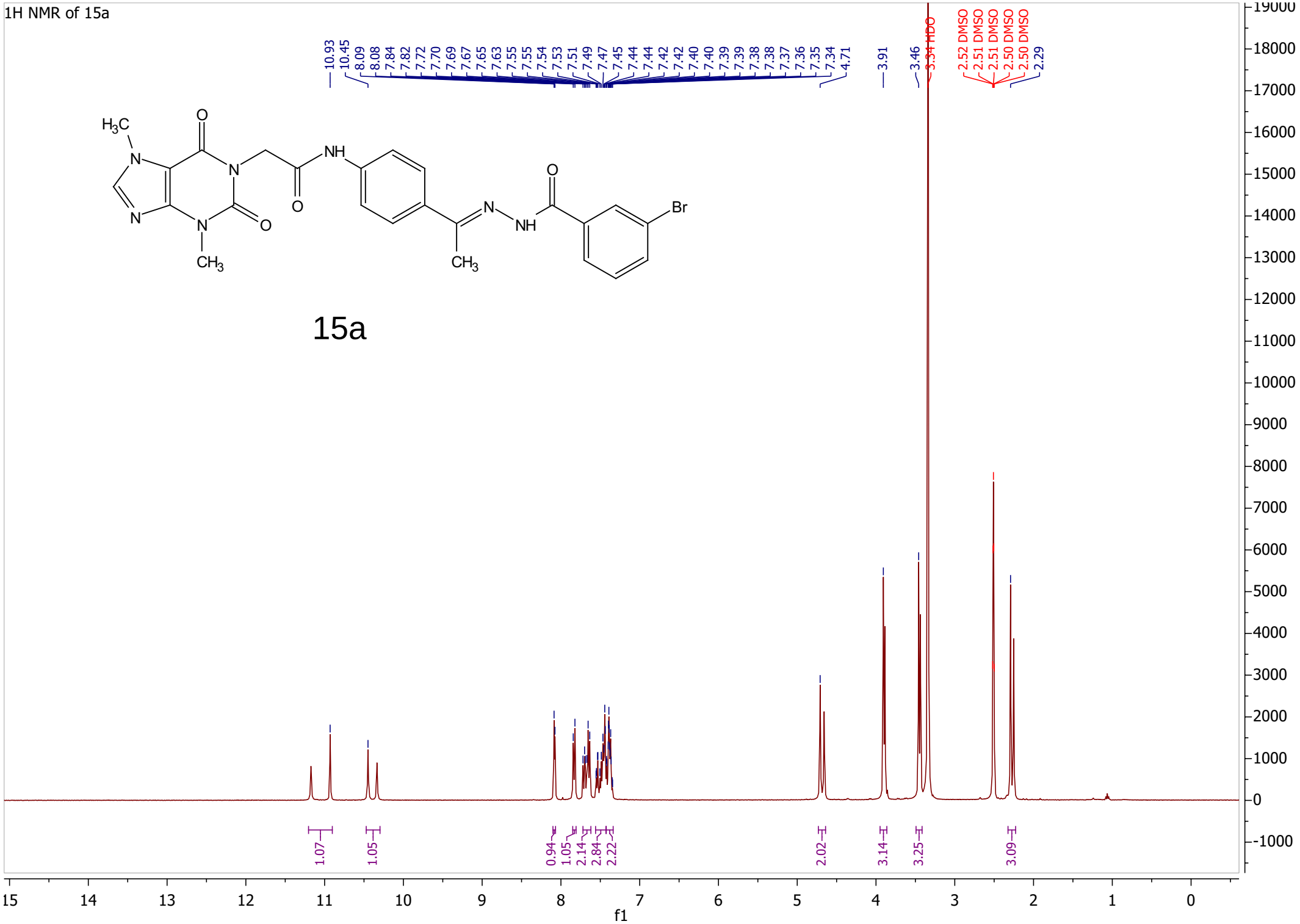
In this protocol, the general-purpose panel was utilized with the activation of the Prepare ligand option. The change ionization was switched on the true option using the Rule based as an ionization method. In Rule based task, we used the carboxylate as an acid ionization. Additionally, the primary,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

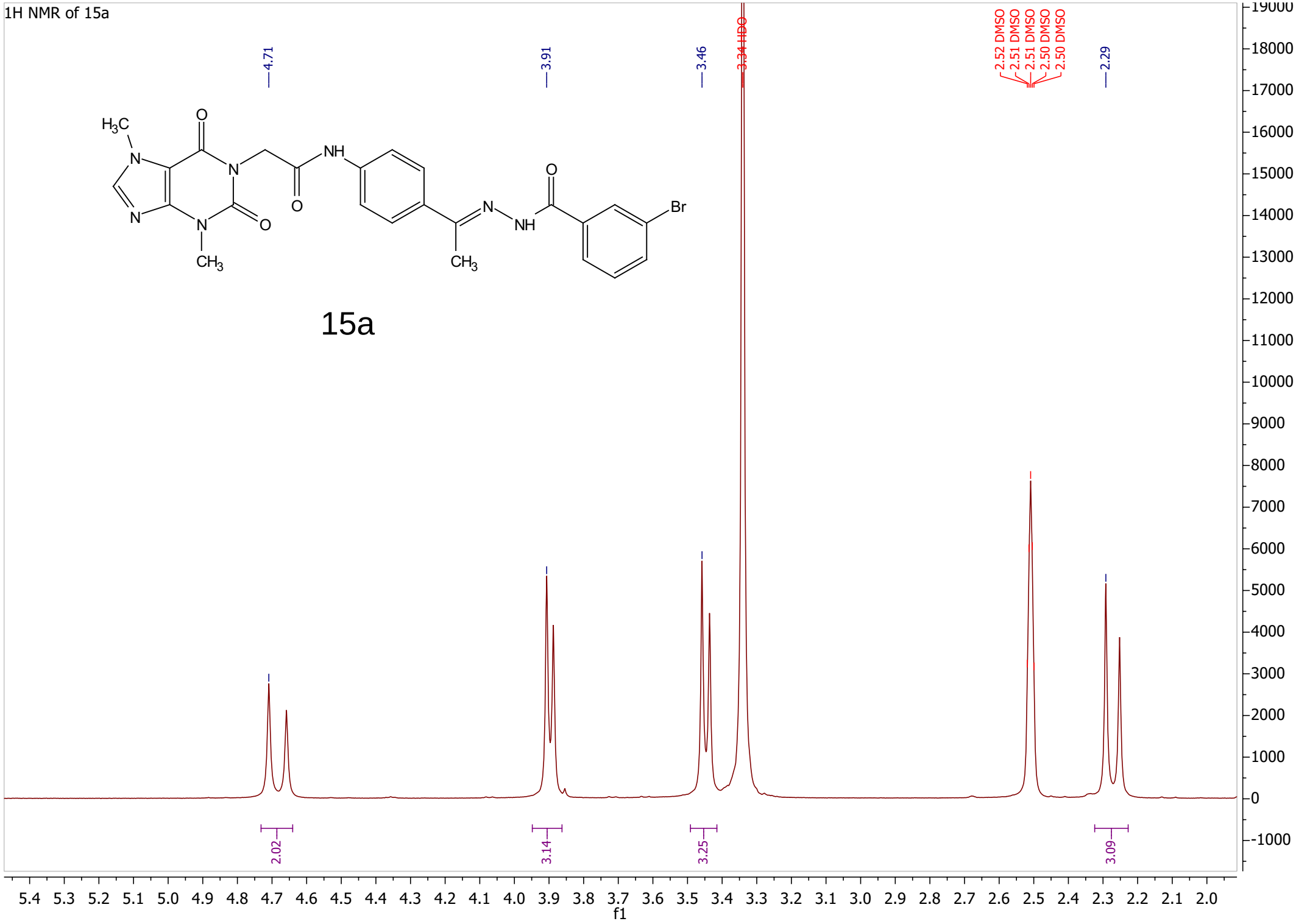
- **Running of Toxicity protocol**

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

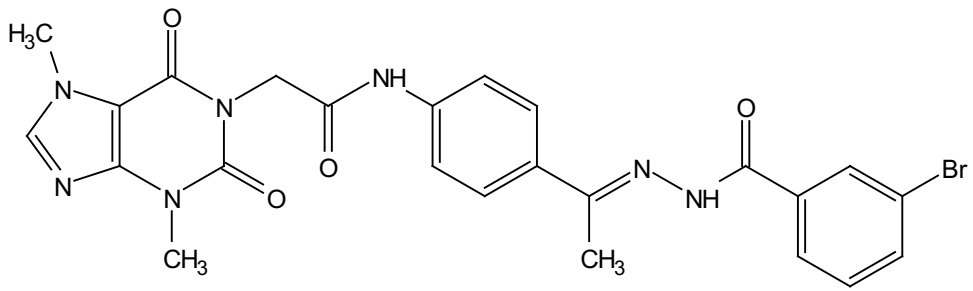


C:\Users\Public\Documents\Bruker\OPUS_8.1.29\DATA\MEAS\XN-3.1	XN-3	Instrument type and / or accessory	2/28/2023


1H NMR of 15a
10.93
10.45
8.09
8.08
7.84
7.82
7.72
7.70
7.69
7.67
7.65
7.63
7.55
7.55
7.54
7.53
7.51
7.49
7.47
7.45
7.44
7.44
7.42
7.42
7.40
7.40
7.39
7.39
7.38
7.38
7.37
7.36
7.35
7.34
4.71
3.91
3.46
3.34 HDO
2.52 DMSO
2.51 DMSO
2.51 DMSO
2.50 DMSO
2.50 DMSO
2.29
H3C
N
O
N
NH
O
N
N
O
N
NH
O
Br
CH3
CH3
15a
1.07
1.05
0.94
1.05
2.14
2.84
2.22
2.02
3.14
3.25
3.09
f1

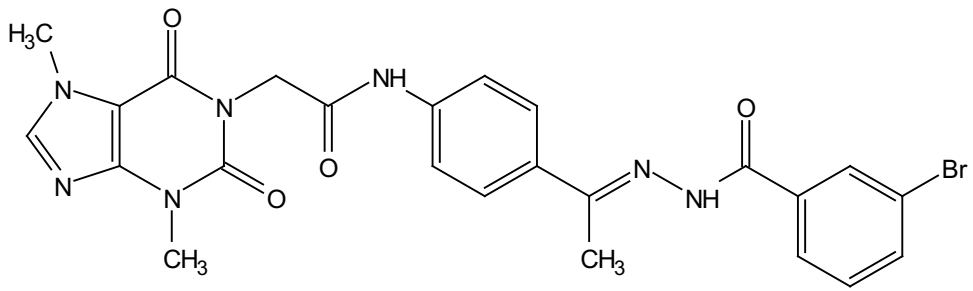

1H NMR of 15a
4.71
3.91
3.46
3.34 HDO
2.52 DMSO
2.51 DMSO
2.51 DMSO
2.50 DMSO
2.50 DMSO
2.29
H3C
N
O
N
NH
O
N
N
O
N
NH
O
Br
CH3
CH3
15a
2.02
3.14
3.25
3.09
f1

1H NMR of 15a



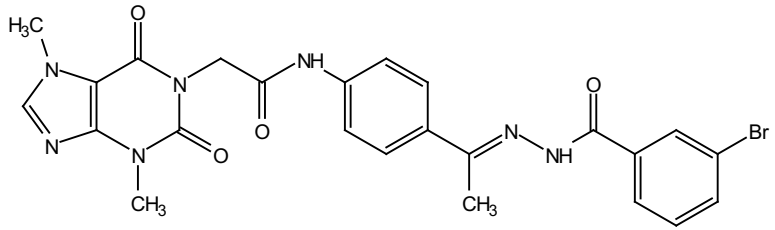
15a

1H NMR of 15a

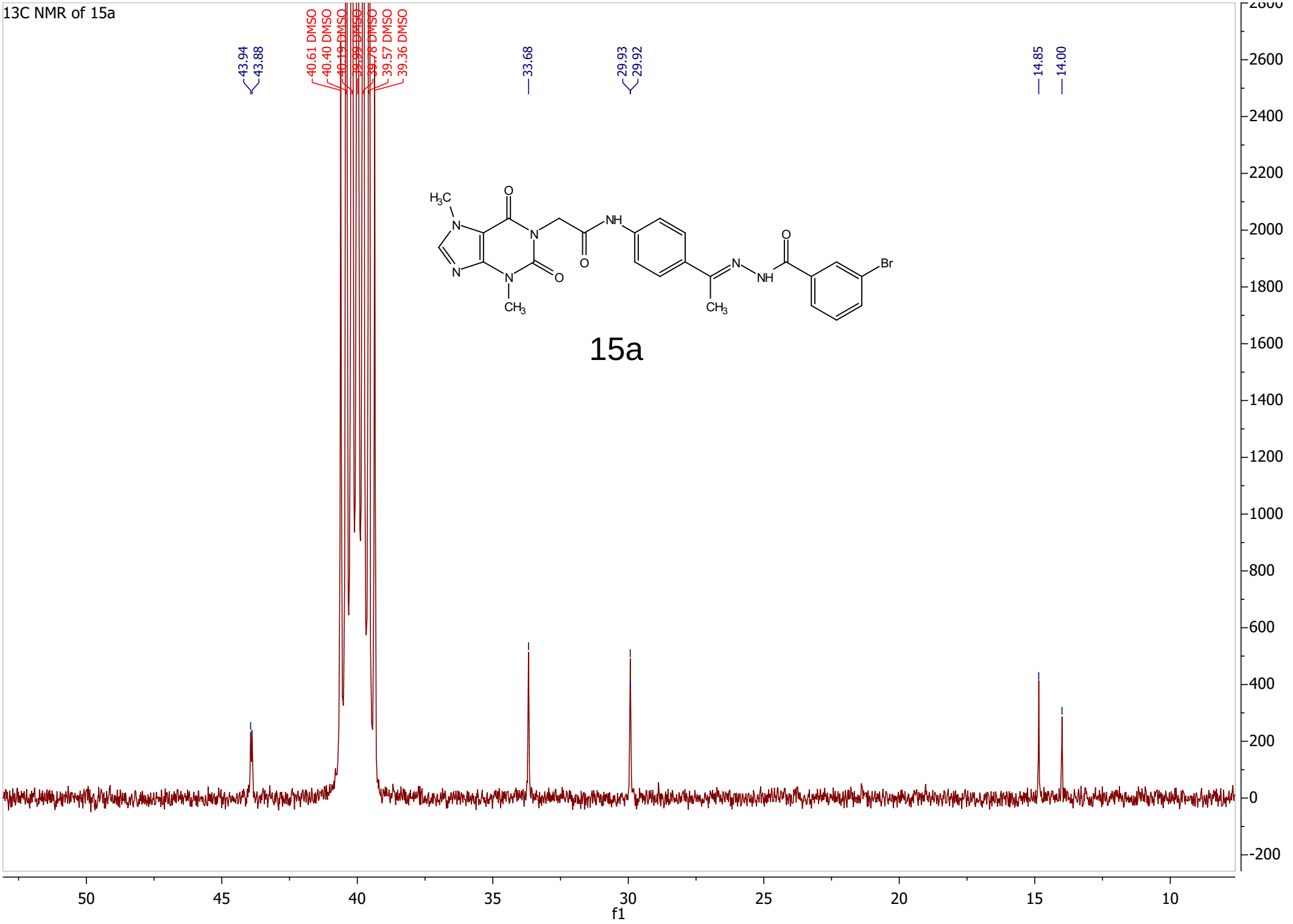


15a

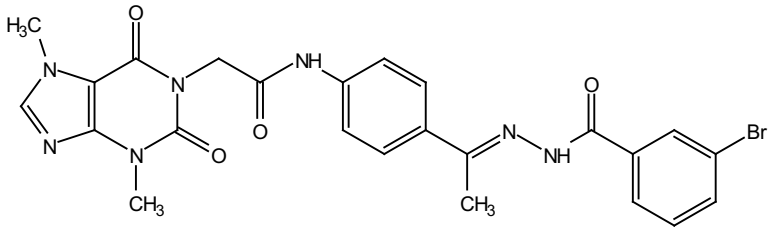
13C NMR of 15a



15a

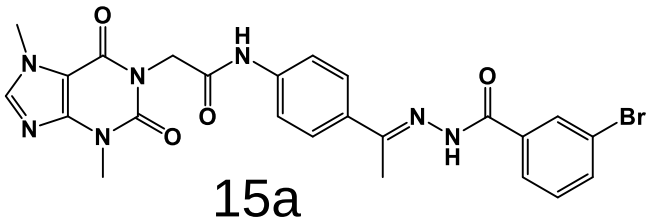

13C NMR of 15a
43.94
43.88
40.61 DMSO
40.40 DMSO
40.19 DMSO
39.99 DMSO
39.78 DMSO
39.57 DMSO
39.36 DMSO
33.68
29.93
29.92
14.85
14.00
H3C
N
N
O
N
NH
O
N
N
O
CH3
N
NH
O
Br
CH3
15a
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
-200
50
45
40
35
30
25
20
15
10
f1

13C NMR of 15a

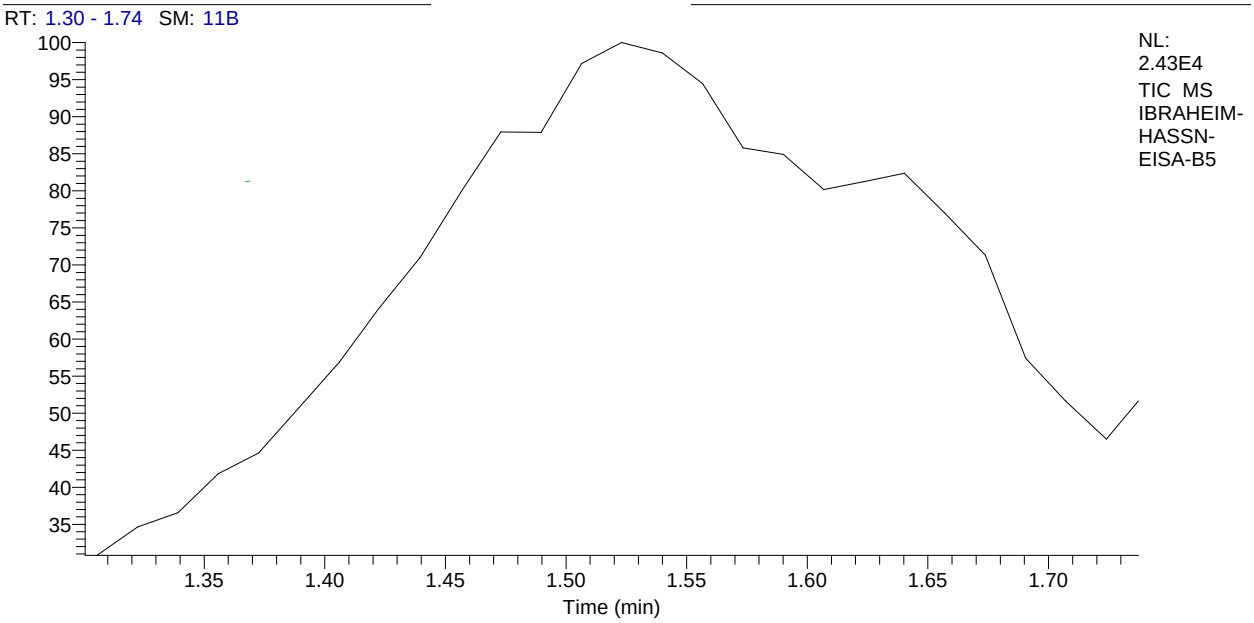


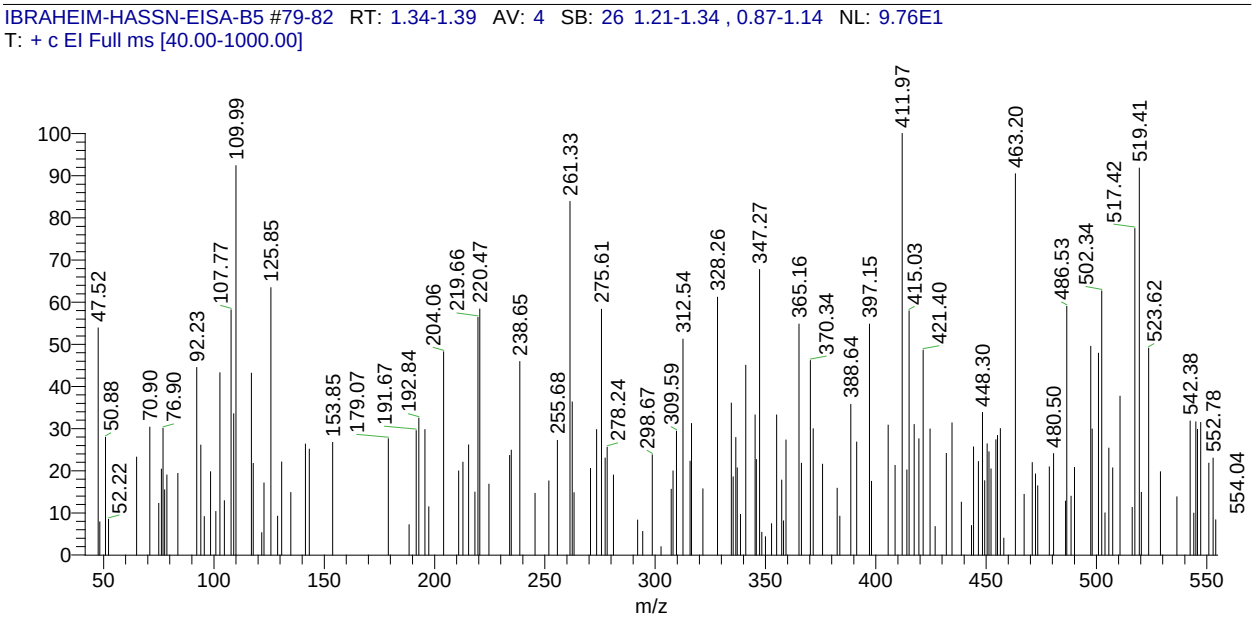
15a

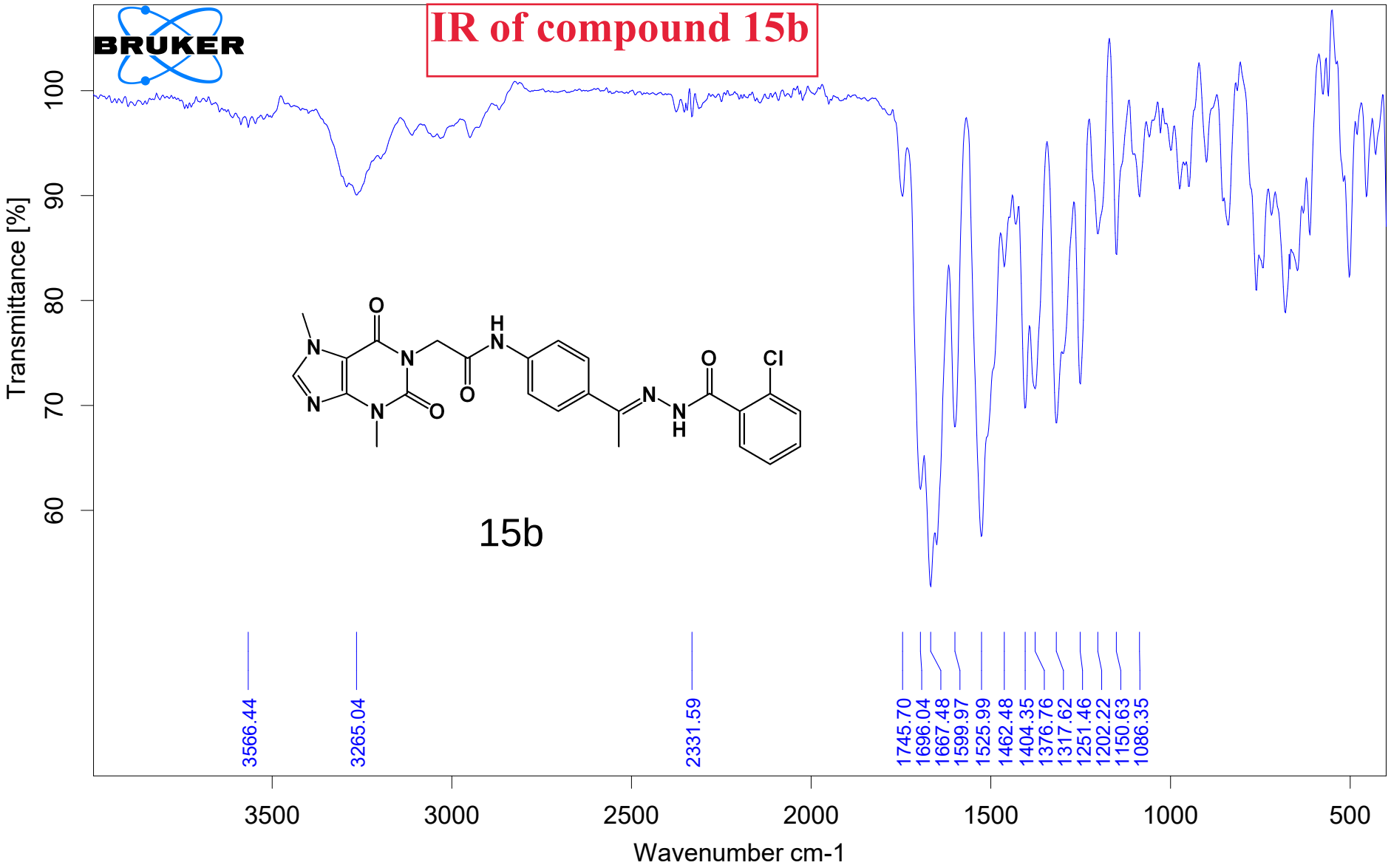
Mass spec. of 15a



15a

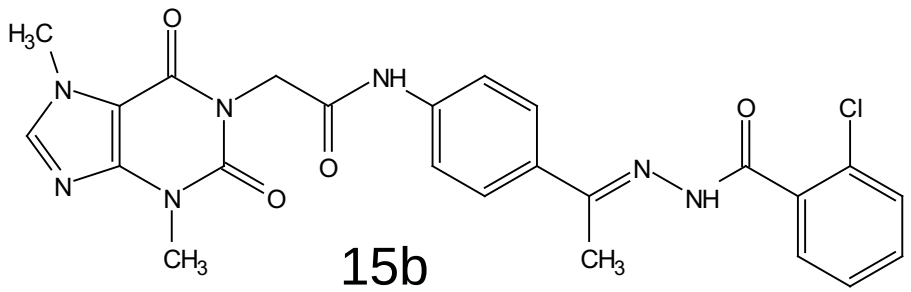






C:\Users\Public\Documents\Bruker\OPUS_8.1.29\DATA\MEAS\XN-3.23	XN-3	Instrument type and / or accessory	2/28/2023

1H NMR of 15b



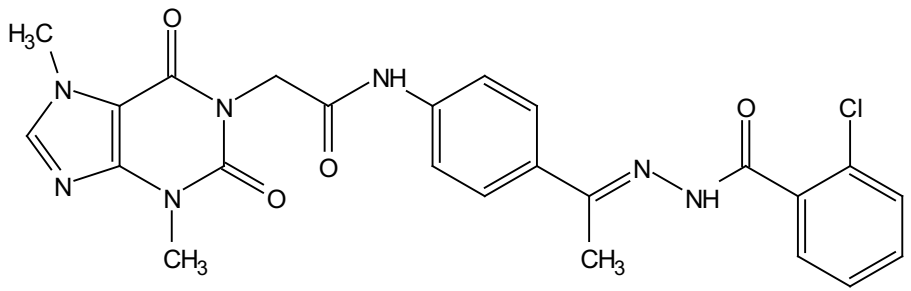
15b

10.92 10.43 8.08 7.97 7.85 7.83 7.66 7.64 7.59 7.57 7.55 7.52 7.45 7.43 7.40 7.38 4.72 3.91 3.47 HDO 3.44 HDO 3.35 HDO 3.33 HDO 2.52 DMSO 2.30

1.11 1.07 1.28 0.99 1.09 5.98 2.00 3.39 3.10

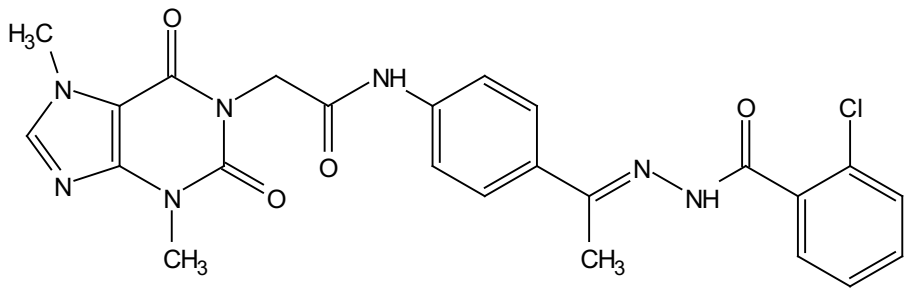
f1

1H NMR of 15b

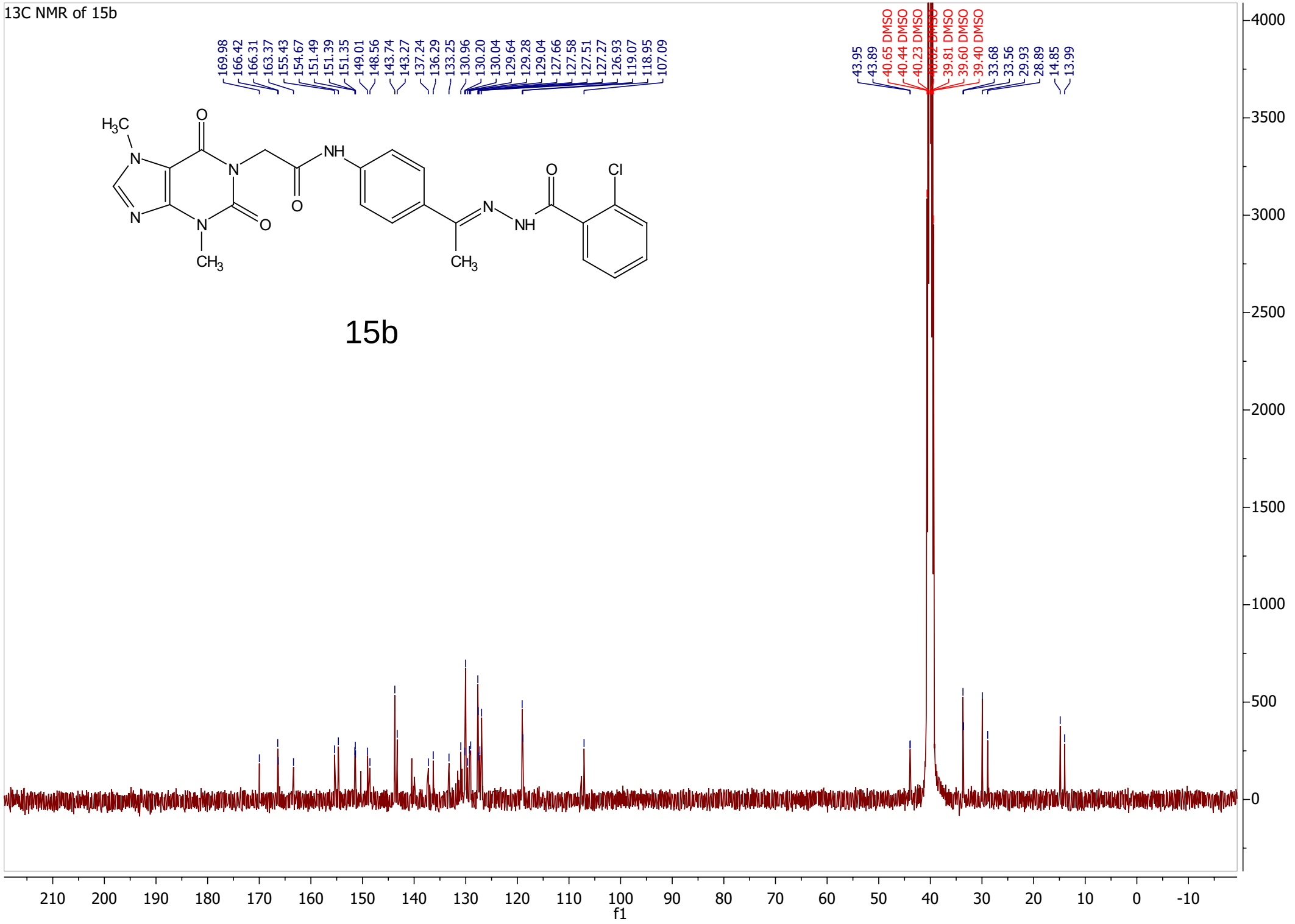


15b

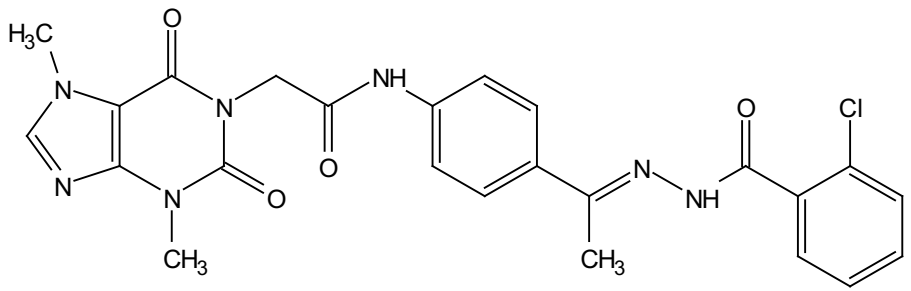
1H NMR of 15b



15b


13C NMR of 15b
169.98
166.42
166.31
163.37
155.43
154.67
151.49
151.39
151.35
149.01
148.56
143.74
143.27
137.24
136.29
133.25
130.96
130.20
130.04
129.64
129.28
129.04
127.66
127.58
127.51
127.27
126.93
119.07
118.95
107.09
43.95
43.89
40.65 DMSO
40.44 DMSO
40.23 DMSO
39.81 DMSO
39.60 DMSO
39.40 DMSO
33.68
33.56
29.93
28.89
14.85
13.99
H3C
N
O
N
NH
Cl
O
N
N
O
N
NH
O
CH3
CH3
15b
4000
3500
3000
2500
2000
1500
1000
500
0
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1

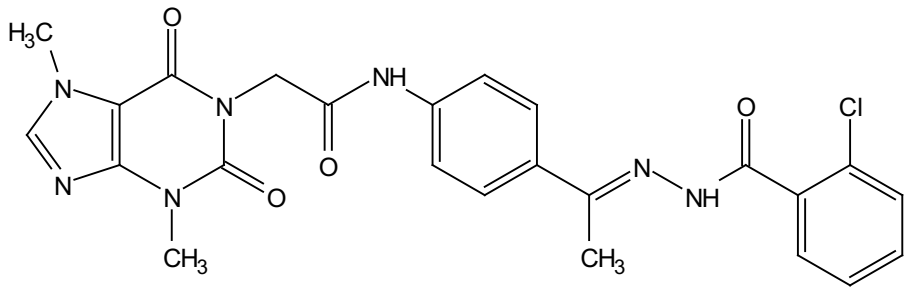
13C NMR of 15b



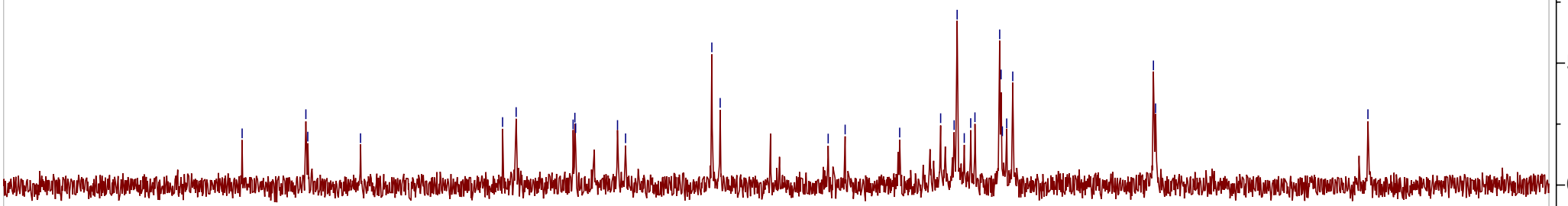
15b

13C NMR of 15b

169.98
166.42
166.31
163.37
155.43
154.67
151.49
151.39
151.35
149.01
148.56
143.74
143.27
137.24
136.29
133.25
130.96
130.20
130.04
129.64
129.28
129.04
127.66
127.58
127.51
127.27
126.93
119.07
118.95
107.09

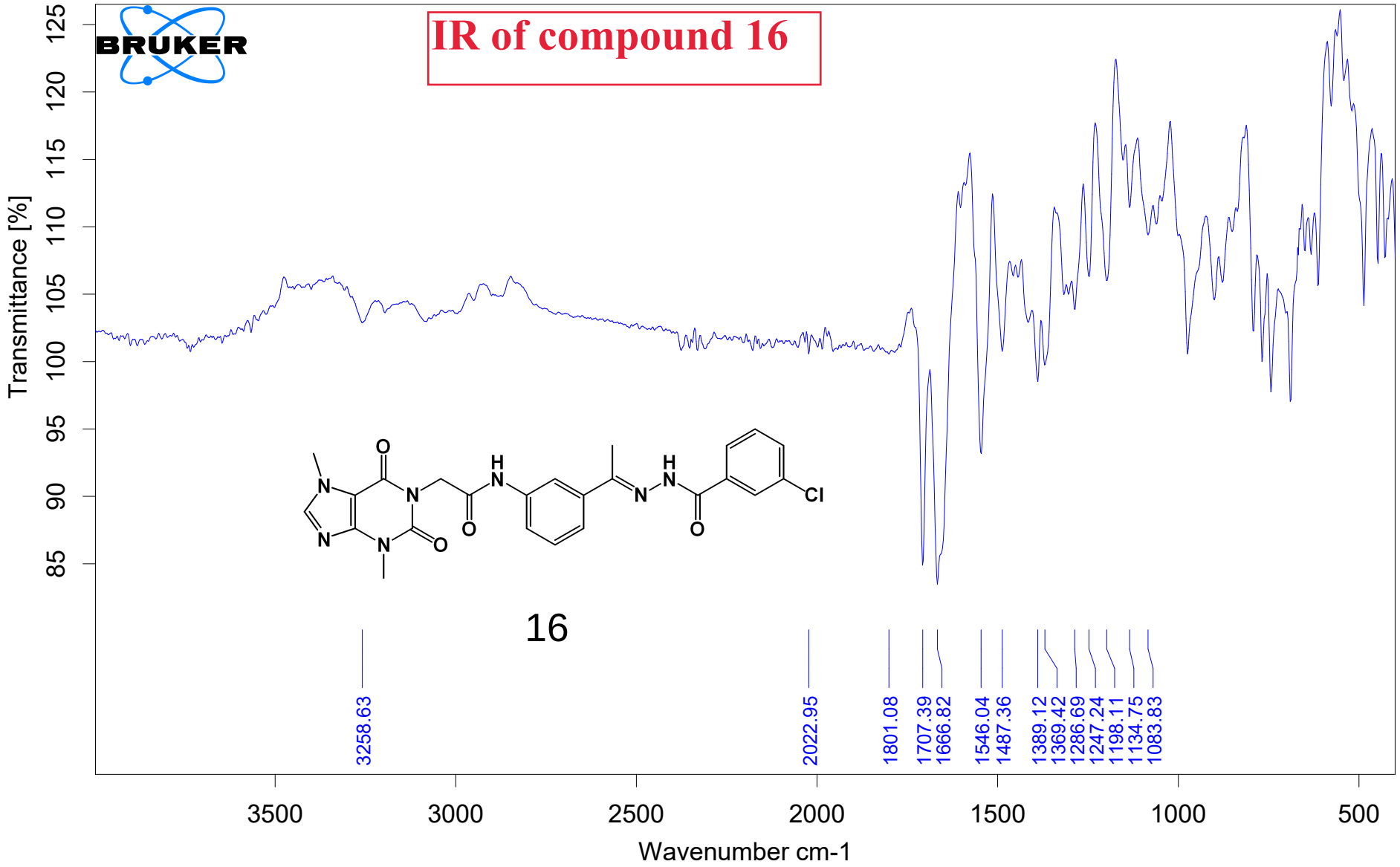


15b



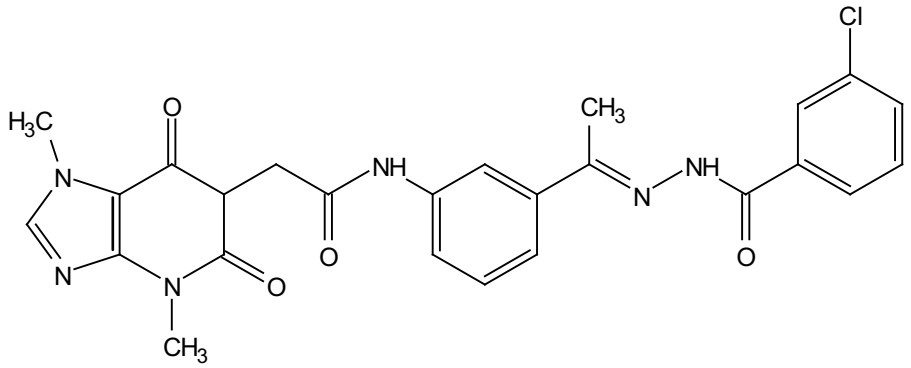
4000
3500
3000
2500
2000
1500
1000
500
0

180 175 170 165 160 155 150 145 140 135 130 125 120 115 110 105 100
f1



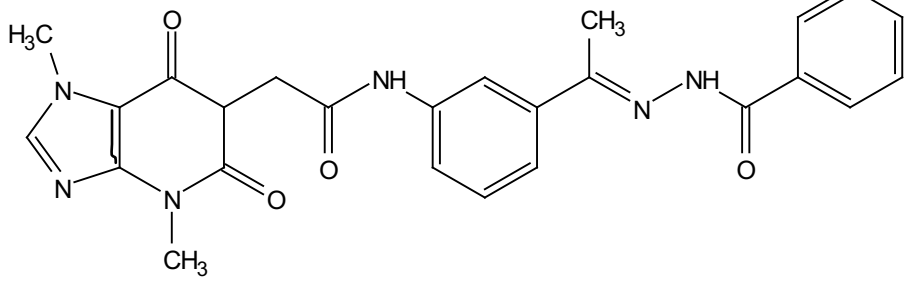
C:\Users\Public\Documents\Bruker\OPUS_8.1.29\DATA\MEAS\XN-6.0	XN-6	Instrument type and / or accessory	2/28/2023

1H NMR of 16



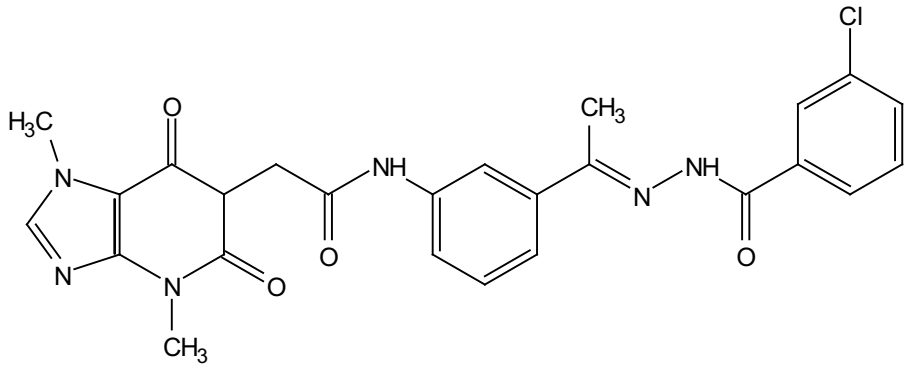
16

1H NMR of 16



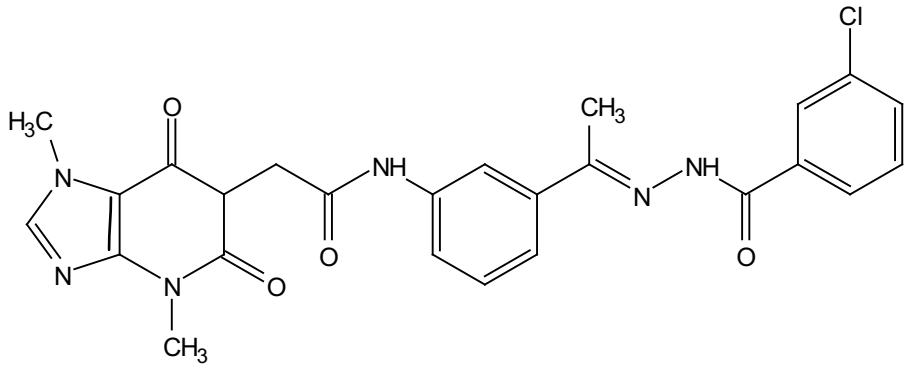
16

1H NMR of 16



16

13C NMR of 16



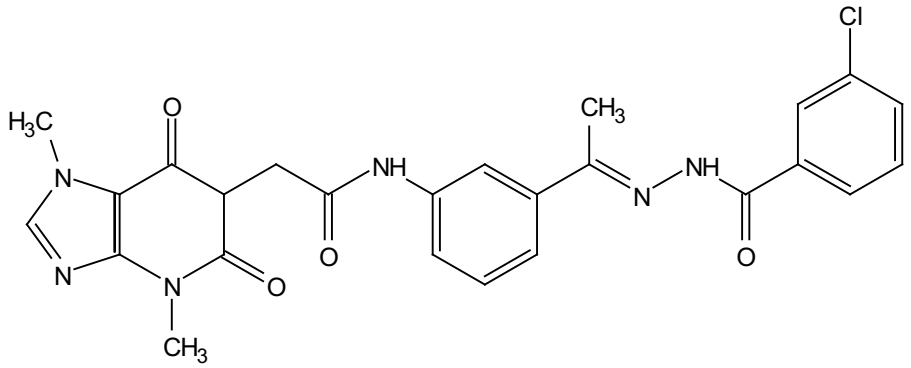
16

166.41
154.65
151.38
148.96
143.69
139.34
139.03
136.51
133.57
131.78
130.77
129.33
128.11
127.15
122.19
120.71
117.48
107.06
43.89
40.57 DMSO
40.36 DMSO
40.15 DMSO
39.94 DMSO
39.73 DMSO
39.52 DMSO
39.32 DMSO
33.67
29.92
15.22

9000
8000
7000
6000
5000
4000
3000
2000
1000
0

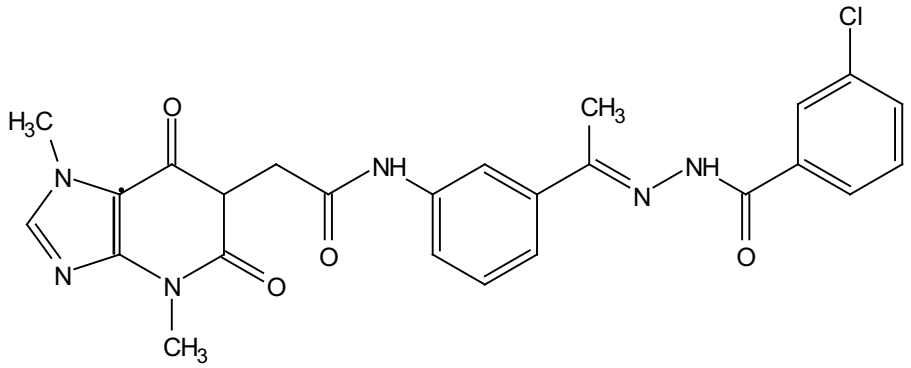
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1

13C NMR of 16



16

13C NMR of 16

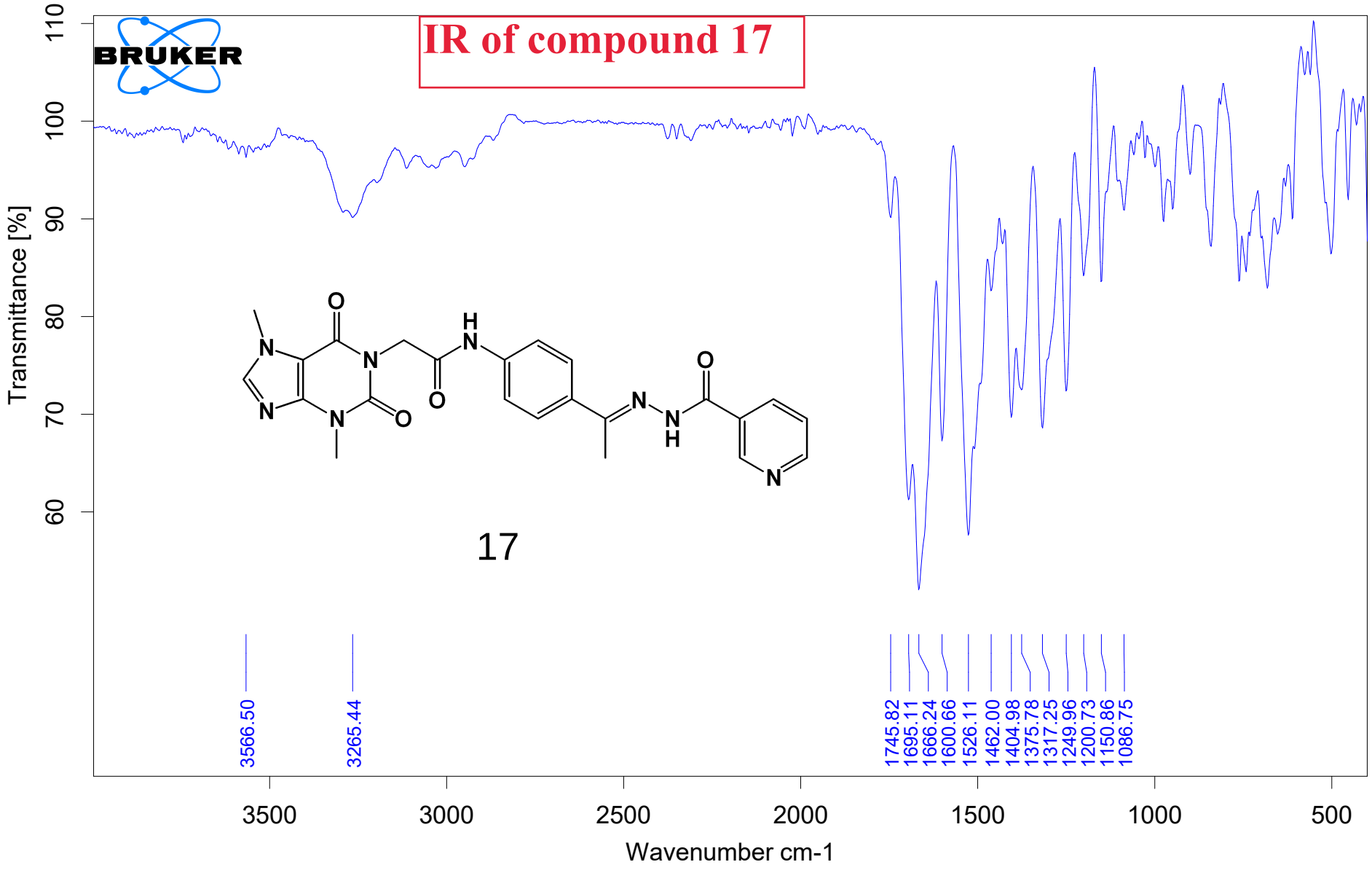


166.41
154.65
151.38
148.96
143.69
139.34
139.03
136.51
133.57
131.78
130.77
129.33
128.11
127.15
122.19
120.71
117.48
107.06

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

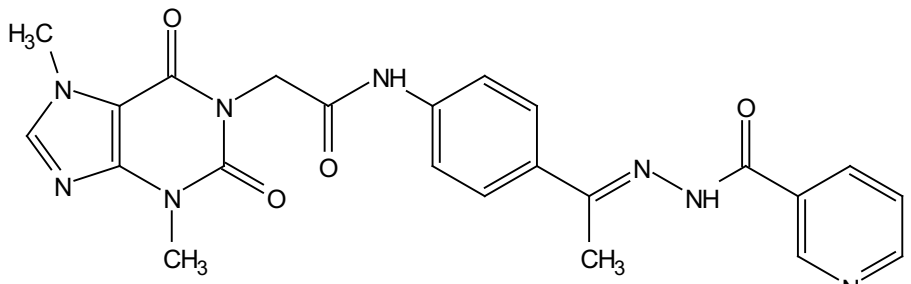
95 190 185 180 175 170 165 160 155 150 145 140 135 130 125 120 115 110 105 100 95

f1



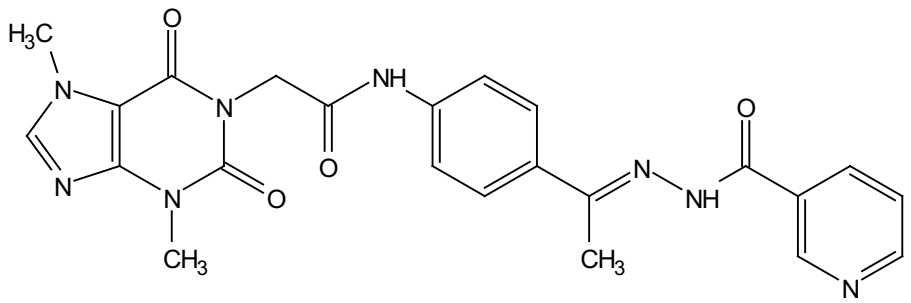
C:\Users\Public\Documents\Bruker\OPUS_8.1.29\DATA\MEAS\XN-3.25	XN-3	Instrument type and / or accessory	2/28/2023

1H NMR of 17



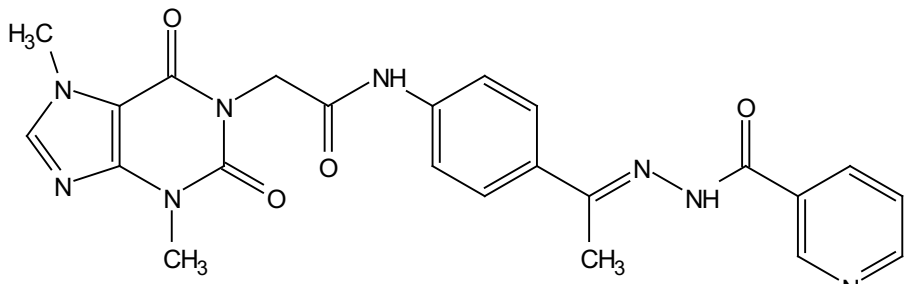
17

1H NMR of 17



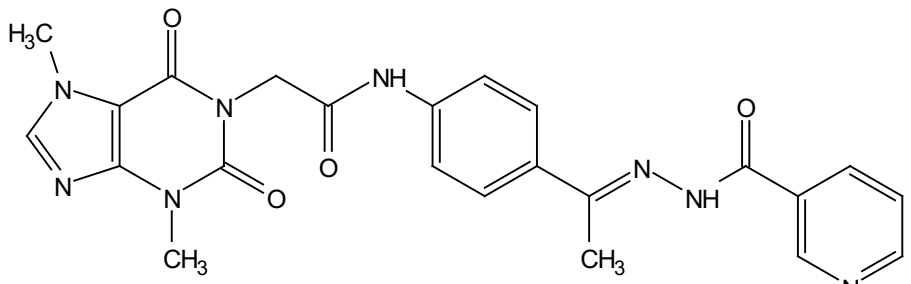
17

1H NMR of 17

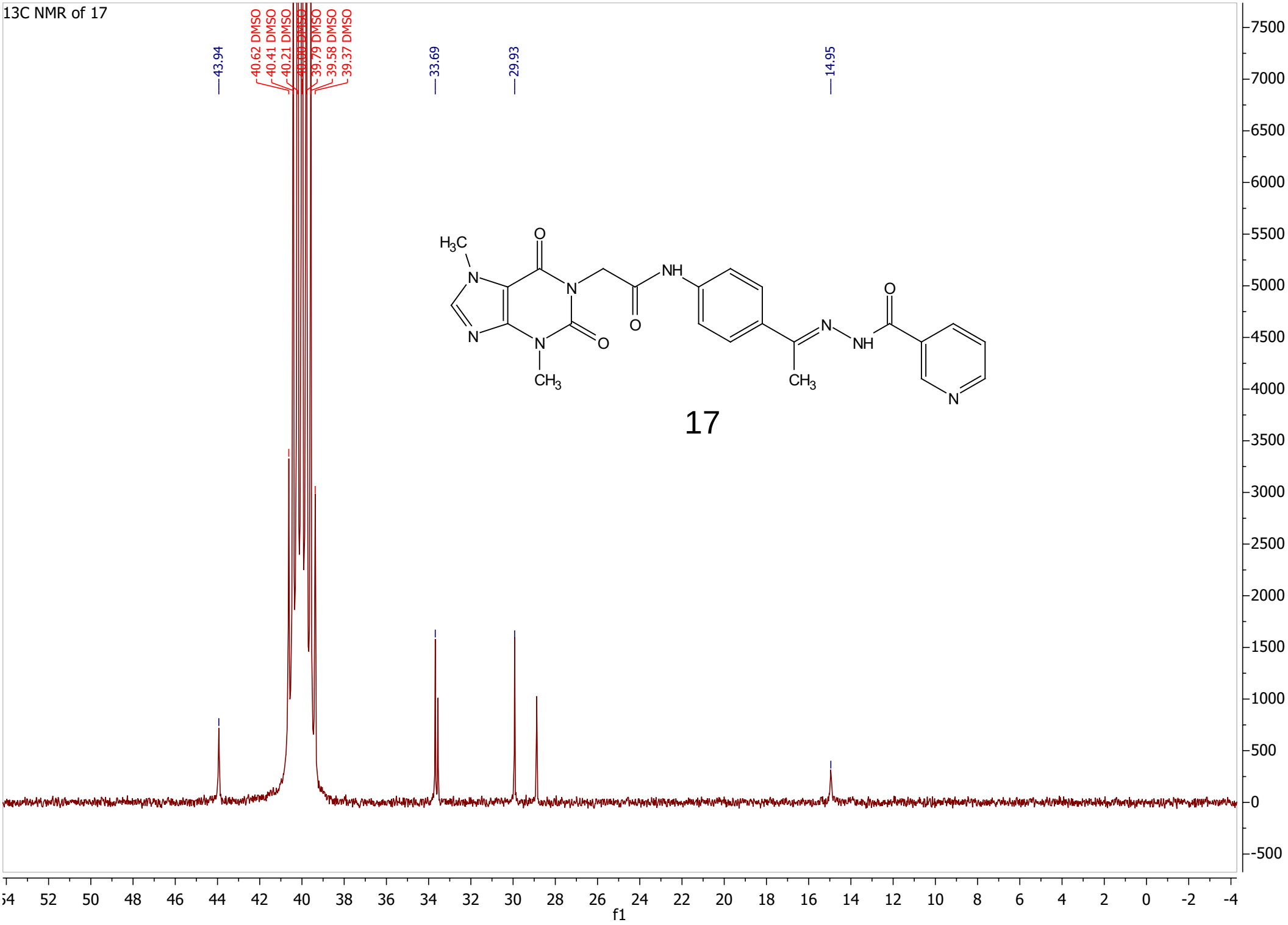


17

13C NMR of 17

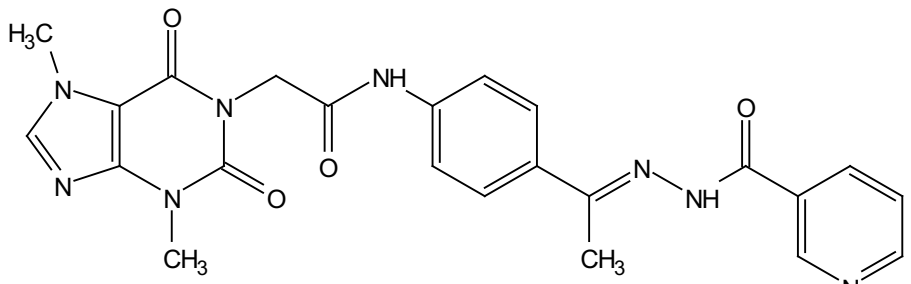


17

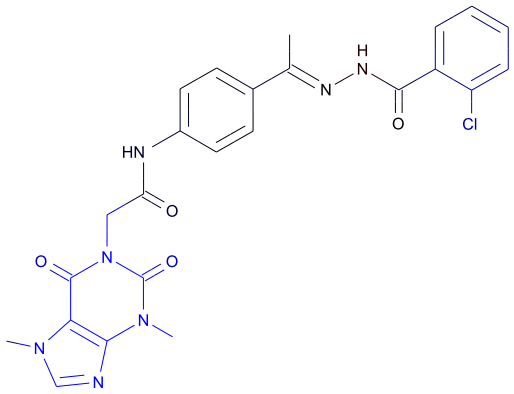

13C NMR of 17
43.94
40.62 DMSO
40.41 DMSO
40.21 DMSO
39.79 DMSO
39.58 DMSO
39.37 DMSO
33.69
29.93
14.95
H3C
CH3
CH3
NH
NH
O
N
f1

17

13C NMR of 17



17



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.182

Enrichment: 0.326

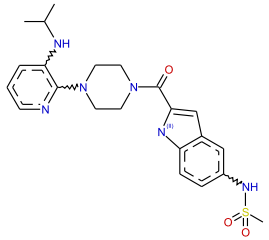
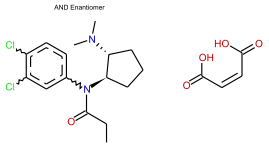
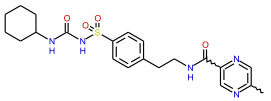
Bayesian Score: -14.6

Mahalanobis Distance: 13.4

Mahalanobis Distance p-value: 4.74e-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	Delavirdine	67450-45-7	GLIPIZIDE
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.567	0.576	0.635
Reference	Contrera, J.F., Matthews, E.J., Kruhlak, N.L., and Benz, R.D., Regulatory Toxicology and Pharmacology 2005, 313-323.	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	PDR 1994

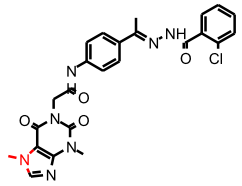
Model Applicability

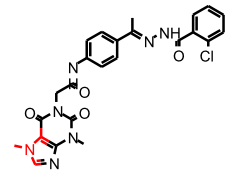
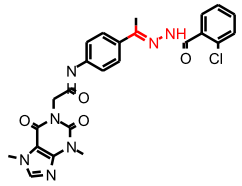
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

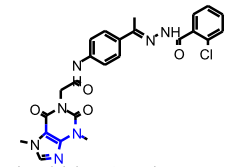
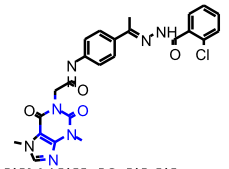
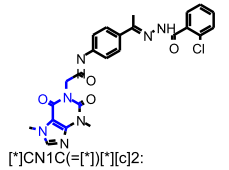
Feature Contribution

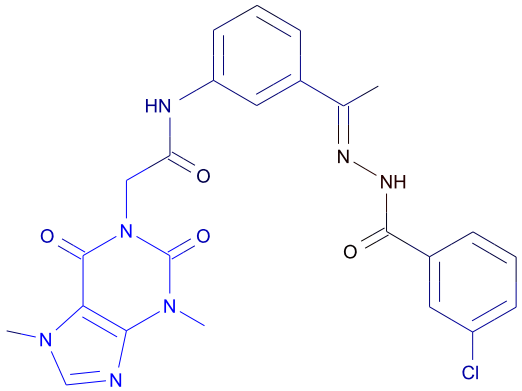
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	136358998	 [*]:n(:[*])C	0.455	142 out of 157

SCFP_12	-443505090	 [*][c]1:[*]:[*]:[cH]: n:1C	0.343	176 out of 218
SCFP_12	-1326021460	 [*]NN=C([*])[*]	0.328	44 out of 55

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1205795299	 [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]	-1.22	2 out of 16
SCFP_12	1731225349	 [*]N1[*][c]2:[*]:[*]: n:[c]:2N(C)C1=O	-1.19	0 out of 4
SCFP_12	1445006032	 [*]CN1C(=[*])[*][c]2: [*]:[*]:n([*]):[c]:2 C1=O	-1.19	0 out of 4



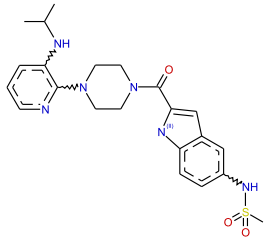
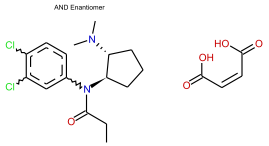
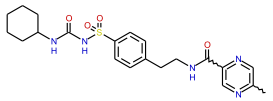
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Mutagen
Probability: 0.124
Enrichment: 0.223
Bayesian Score: -16.4
Mahalanobis Distance: 13.4
Mahalanobis Distance p-value: 4.74e-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	Delavirdine	67450-45-7	GLIPIZIDE
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.567	0.577	0.635
Reference	Contrera, J.F., Matthews, E.J., Kruhlak, N.L., and Benz, R.D., Regulatory Toxicology and Pharmacology 2005, 313-323.	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	PDR 1994

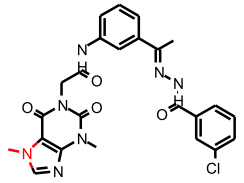
Model Applicability

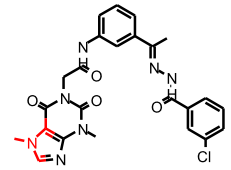
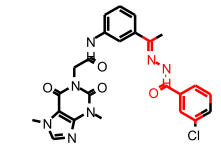
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

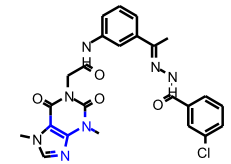
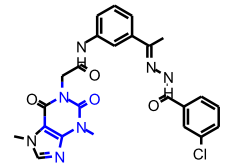
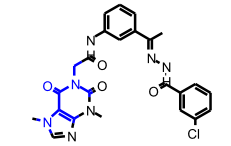
Feature Contribution

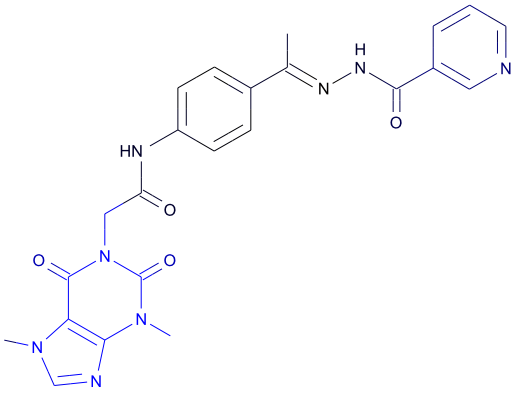
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	136358998	 [*]:n(:[*])C	0.455	142 out of 157

SCFP_12	-443505090	 [*][c]1:[*]:[*]:[cH]: n:1C	0.343	176 out of 218
SCFP_12	-319622735	 [*]C(=NNC(=O)[c]1:[cH]:[cH]:[*]:[c]([*]): [cH]:1)[*]	0.337	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1205795299	 [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]	-1.22	2 out of 16
SCFP_12	1731225349	 [*]N1[*][c]2:[*]:[*]: n:[c]:2N(C)C1=O	-1.19	0 out of 4
SCFP_12	1445006032	 [*]CN1C(=[*])[*][c]2: [*]:[*]:n([*]):[c]:2 C1=O	-1.19	0 out of 4



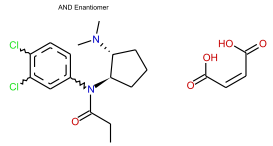
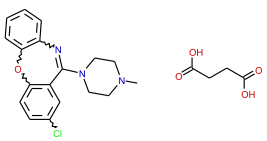
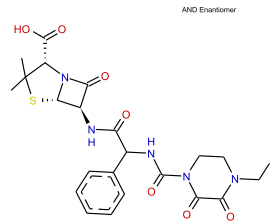
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: Non-Mutagen
Probability: 0.212
Enrichment: 0.38
Bayesian Score: -13.8
Mahalanobis Distance: 13.2
Mahalanobis Distance p-value: 1.65e-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	67450-45-7	27833-64-3	61477-96-1
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.620	0.625	0.636
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

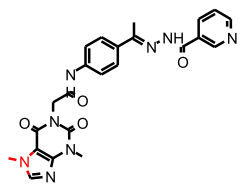
Model Applicability

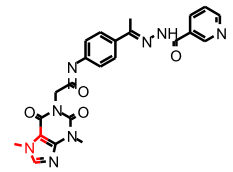
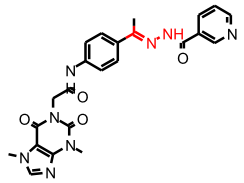
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

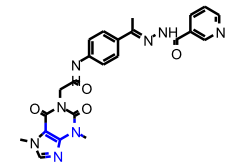
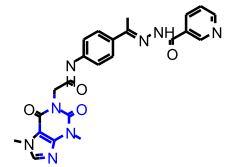
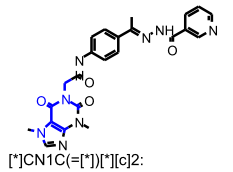
Feature Contribution

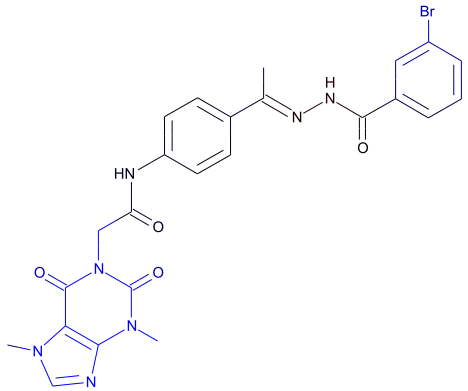
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	136358998	 [*]:n(:[*])C	0.455	142 out of 157

SCFP_12	-443505090	 [*][c]1:[*]:[*]:[cH]: n:1C	0.343	176 out of 218
SCFP_12	-1326021460	 [*]NN=C([*])[*]	0.328	44 out of 55

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1205795299	 [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]	-1.22	2 out of 16
SCFP_12	1731225349	 [*]N1[*][c]2:[*]:[*]: n:[c]:2N(C)C1=O	-1.19	0 out of 4
SCFP_12	1445006032	 [*]CN1C(=[*])[*][c]2: [*]:[*]:n([*]):[c]:2 C1=O	-1.19	0 out of 4



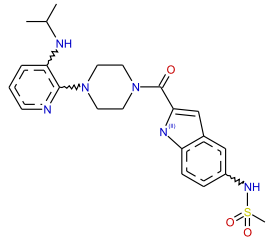
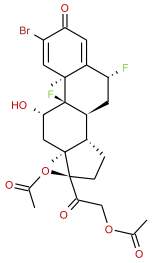
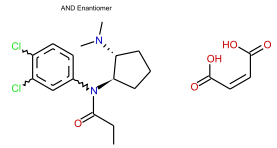
$C_{24}H_{22}BrN_7O_4$
Molecular Weight: 552.37997
ALogP: 2.118
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Mutagen
Probability: 0.178
Enrichment: 0.318
Bayesian Score: -14.7
Mahalanobis Distance: 15.3
Mahalanobis Distance p-value: 1.39e-013

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

Structural Similar Compounds

Name	Delavirdine	57781-14-3	67450-45-7
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.618	0.625	0.639
Reference	Contrera, J.F., Matthews, E.J., Kruhlak, N.L., and Benz, R.D., Regulatory Toxicology and Pharmacology 2005, 313-323.	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

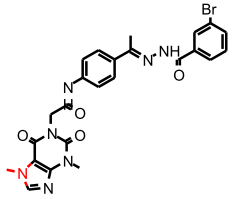
Model Applicability

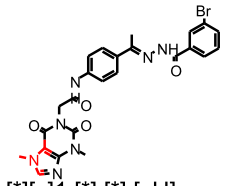
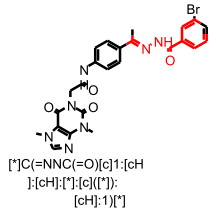
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

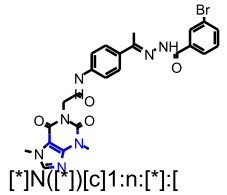
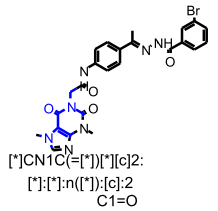
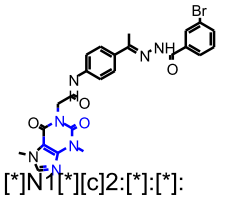
Feature Contribution

Top features for positive contribution

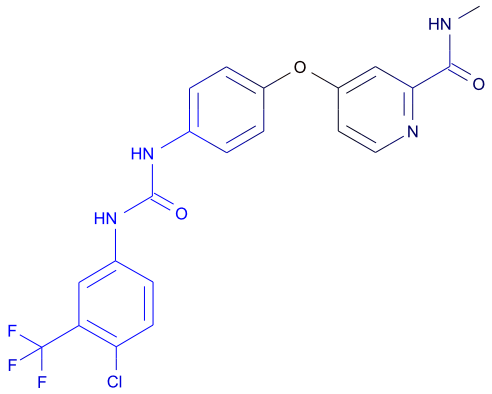
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	136358998	 [*]:n(:[*])C	0.455	142 out of 157

SCFP_12	-443505090		0.343	176 out of 218
SCFP_12	-319622735		0.337	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1205795299		-1.22	2 out of 16
SCFP_12	1445006032		-1.19	0 out of 4
SCFP_12	1731225349		-1.19	0 out of 4

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

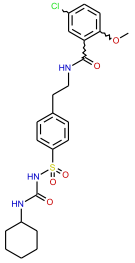
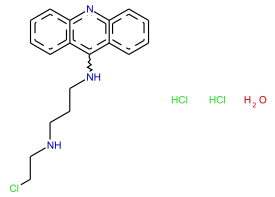
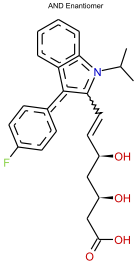
Rotatable Bonds: 6

Acceptors: 4

Donors: 3

TOPKAT_Ames_Mutagenicity

Structural Similar Compounds

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

Model Prediction

Prediction: Non-Mutagen

Probability: 0.0531

Enrichment: 0.0951

Bayesian Score: -19.7

Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 2.73e-006

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

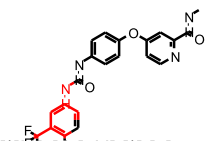
Model Applicability

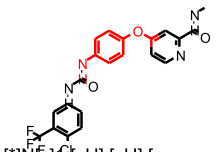
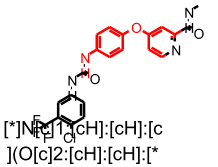
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

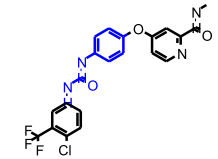
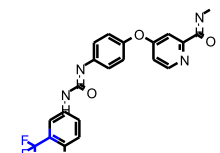
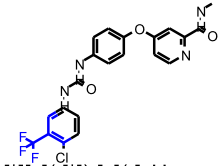
Feature Contribution

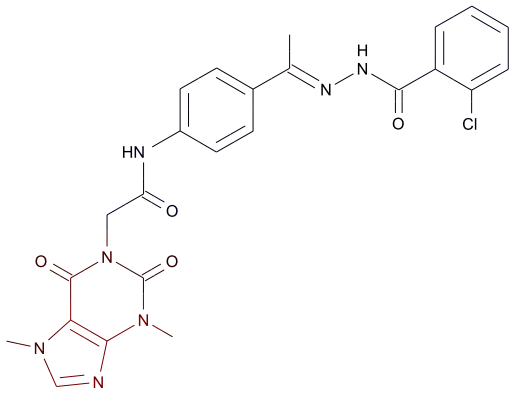
Top features for positive contribution

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

SCFP_12	1208843554	 [*]N[c]1:[cH]:[cH]:[c](O[c](:[*]):[*]):[cH]:[cH]:1	0.337	6 out of 7
SCFP_12	-1943080297	 [*]N[c]1:[cH]:[cH]:[c](O[c]2:[cH]:[cH]:[*]:[c]([*]):[cH]:2):[cH]:[cH]:1	0.304	5 out of 6

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	816802409	 [*]NC(=O)N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	-1.82	0 out of 9
SCFP_12	-300280774	 [*]:[c](:[*])C(F)(F)F	-1.51	3 out of 30
SCFP_12	-1903175541	 [*][c](:[*]):[c](:[cH]:[*])C(F)(F)F	-1.51	3 out of 30



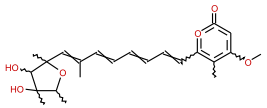
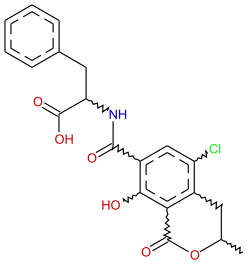
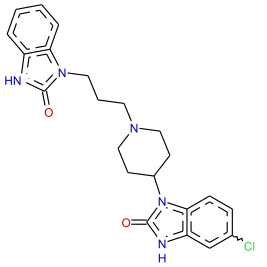
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.583
Enrichment: 1.11
Bayesian Score: 0.933
Mahalanobis Distance: 9.26
Mahalanobis Distance p-value: 0.11

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

Structural Similar Compounds

Name	Citreoviridin	Ochratoxin a	Domperidone
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.647	0.684	0.694
Reference	Food Chem Toxicol 24(12):1315-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	Yakuri to Chiryo 8:4125-4136; 1980

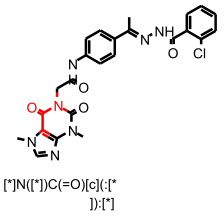
Model Applicability

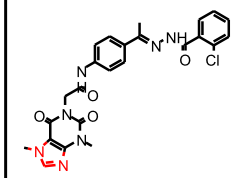
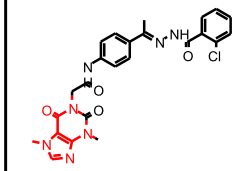
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

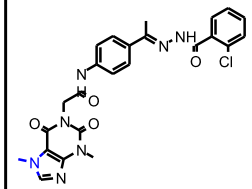
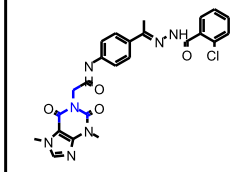
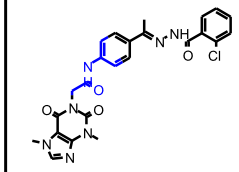
Feature Contribution

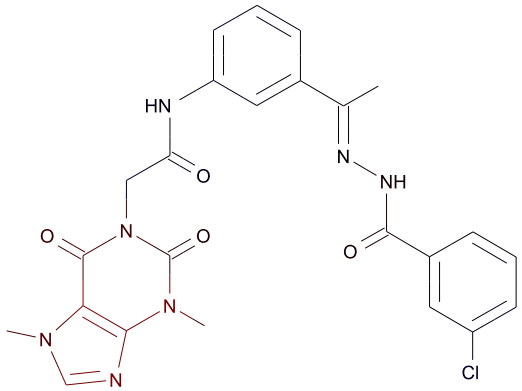
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1257084377	 [*]N([*])C(=O)[c](:[*]):[*]	0.362	14 out of 18

SCFP_6	-1181430618	 [*]n1:[*]:[*]:n:[cH]: 1	0.298	6 out of 8
SCFP_6	1443356060	 [*]N1[*]N([*])[c]2:n: [cH]:n(C):[c]:2C1=O	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	136358998	 [*]:n(:[*])C	-0.55	2 out of 8
SCFP_6	399659969	 [*]CN(C(=[*])[*])C(=[*])[*]	-0.526	3 out of 11
SCFP_6	2097618059	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	-0.422	0 out of 1



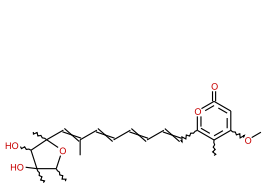
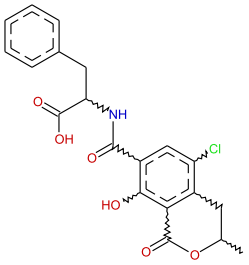
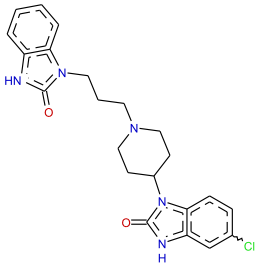
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.588
Enrichment: 1.12
Bayesian Score: 1.06
Mahalanobis Distance: 8.81
Mahalanobis Distance p-value: 0.24

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

Structural Similar Compounds

Name	Citreoviridin	Ochratoxin a	Domperidone
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.648	0.681	0.692
Reference	Food Chem Toxicol 24(12):1315-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	Yakuri to Chiryo 8:4125-4136; 1980

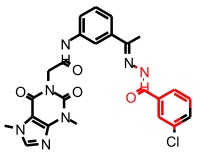
Model Applicability

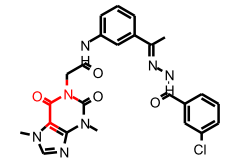
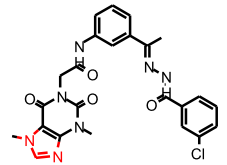
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

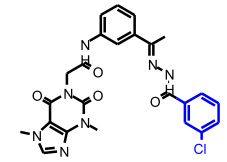
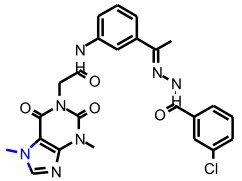
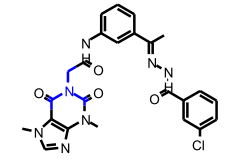
Feature Contribution

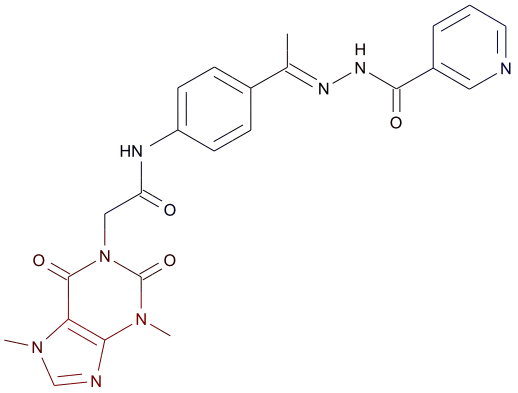
Top features for positive contribution

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

SCFP_6	1257084377	 [*]N([*])C(=O)[c](:[*]):[*]	0.362	14 out of 18
SCFP_6	-1181430618	 [*]n1:[*]:[*]:n:[cH]:1	0.298	6 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	52043406	 [*][c]1:[cH]:[cH]:[cH]:[c](Cl):[cH]:1	-0.718	0 out of 2
SCFP_6	136358998	 [*]:n(:[*])C	-0.55	2 out of 8
SCFP_6	399659969	 [*]CN(C(=[*])[*])C(=[*])[*]	-0.526	3 out of 11



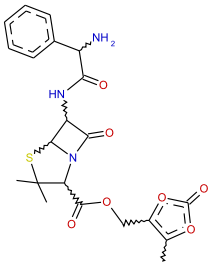
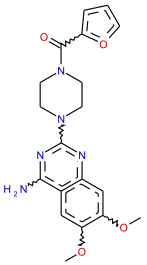
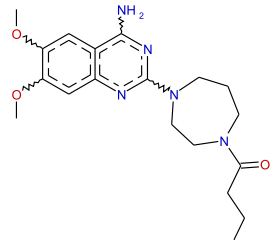
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.599
Enrichment: 1.14
Bayesian Score: 1.34
Mahalanobis Distance: 8.98
Mahalanobis Distance p-value: 0.183

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

Structural Similar Compounds

Name	Lenampicillin .HCl (Free base form)	Prazosin .HCl (Free base form)	Bunazosin .HCl (Free base form)
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.650	0.736	0.743
Reference	Chemotherapy 32:130-145; 1984	Oyo Yakuri 17:57-62; 1979	Kiso to Rinsho 17:914-924; 1983

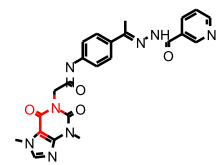
Model Applicability

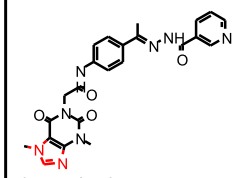
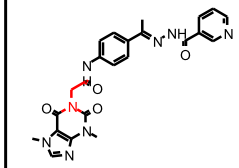
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

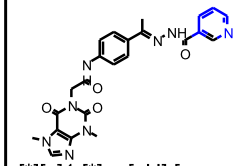
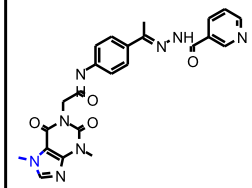
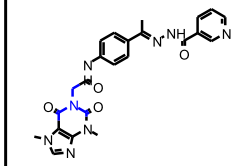
Feature Contribution

Top features for positive contribution

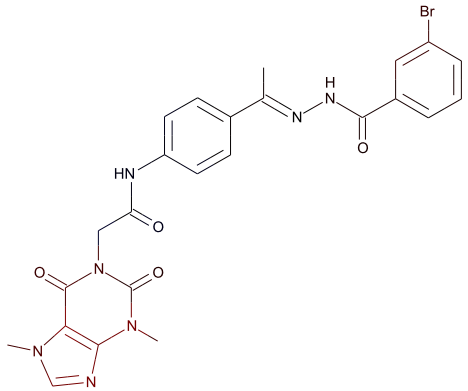
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1257084377	 [*]N([*])C(=O)[c](:[*])[*]	0.362	14 out of 18

SCFP_6	-1181430618	 [*]n1:[*]:[*]:n:[cH]:1	0.298	6 out of 8
SCFP_6	-587539325	 [*]N([*])CC(=[*])[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.646	2 out of 9
SCFP_6	136358998	 [*]:n(:[*])C	-0.55	2 out of 8
SCFP_6	399659969	 [*]CN(C(=[*])[*])C(=[*])[*]	-0.526	3 out of 11

TOPKAT_Developmental_Toxicity_Potential



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Toxic

Probability: 0.655

Enrichment: 1.25

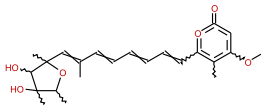
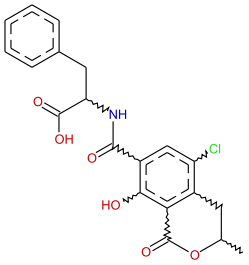
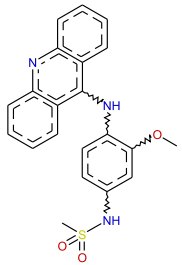
Bayesian Score: 2.76

Mahalanobis Distance: 8.15

Mahalanobis Distance p-value: 0.535

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

Structural Similar Compounds

Name	Citreoviridin	Ochratoxin a	Amsacrine
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.695	0.732	0.734
Reference	Food Chem Toxicol 24(12):1315-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	Fundam Appl Toxicol 7(2):214-20; 1986

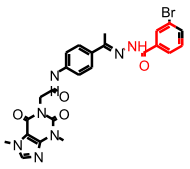
Model Applicability

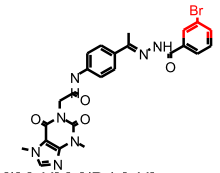
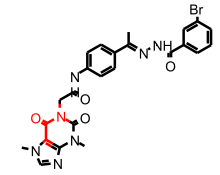
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

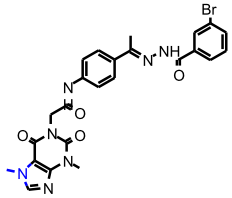
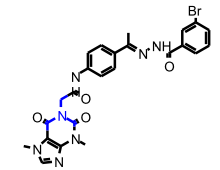
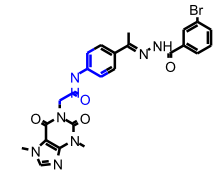
Feature Contribution

Top features for positive contribution

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

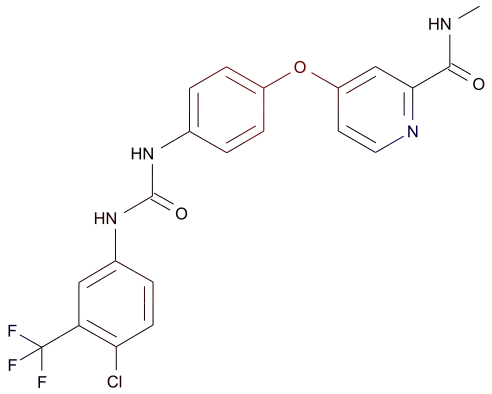
SCFP_6	-1505150337	 [*]:[cH]:[c](Br):[cH]:[*]	0.369	5 out of 6
SCFP_6	1257084377	 [*]N([*])C(=O)[c](:[*]):[*]	0.362	14 out of 18

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	136358998	 [*]:n(:[*])C	-0.55	2 out of 8
SCFP_6	399659969	 [*]CN(C(=[*])[*])C(=[*])[*]	-0.526	3 out of 11
SCFP_6	2097618059	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	-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

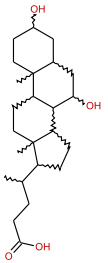
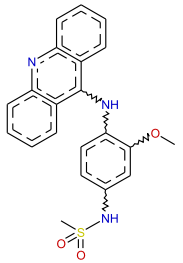
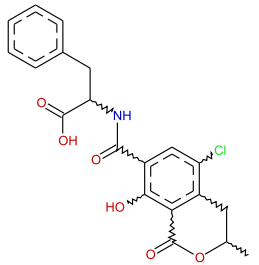
Bayesian Score: 1.15

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

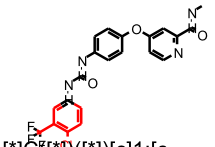
Model Applicability

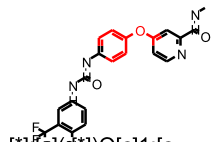
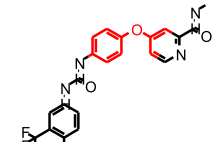
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

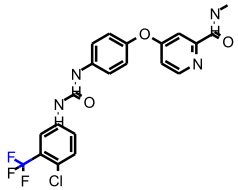
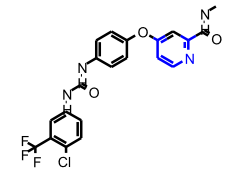
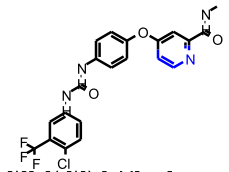
Feature Contribution

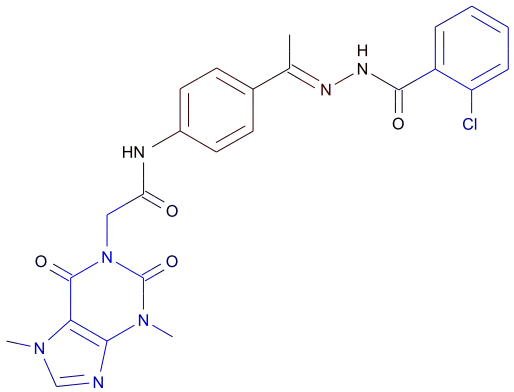
Top features for positive contribution

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

Top Features for negative contribution

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



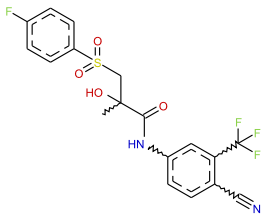
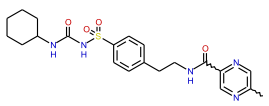
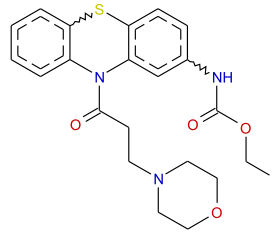
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.221
Enrichment: 0.69
Bayesian Score: -9.76
Mahalanobis Distance: 12.9
Mahalanobis Distance p-value: 0.000676

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Glipizide	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.640	0.662	0.674
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

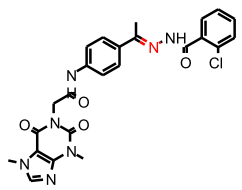
Model Applicability

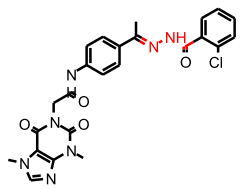
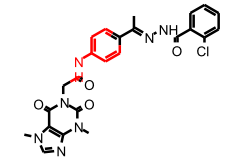
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

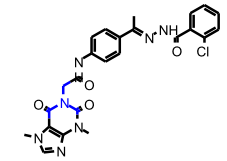
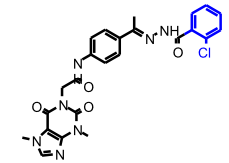
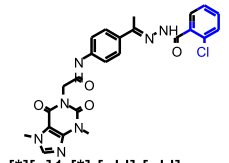
Feature Contribution

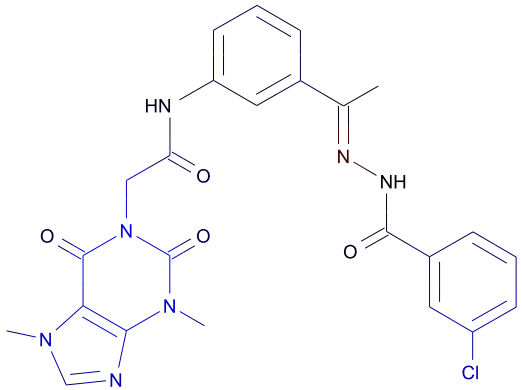
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14

ECFP_6	544048674	 [*]C(=[*])NN=[*]	0.617	2 out of 2
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-661097313	 [*]CN(C(=[*])[*])C(=[*])[*]	-1.55	0 out of 12
ECFP_6	-1625071872	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[c]:1Cl	-0.935	0 out of 5
ECFP_6	1641317964	 [*][c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.789	1 out of 11



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.228

Enrichment: 0.712

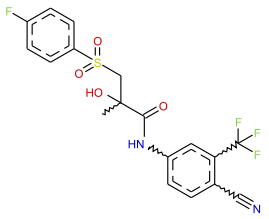
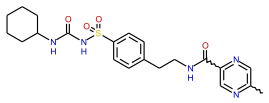
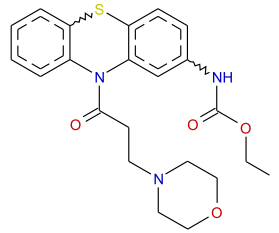
Bayesian Score: -11.4

Mahalanobis Distance: 19.4

Mahalanobis Distance p-value: 1.19e-016

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Glipizide	Moricizine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.637	0.663	0.674
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

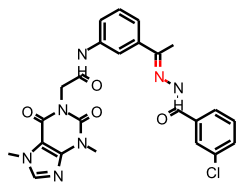
Model Applicability

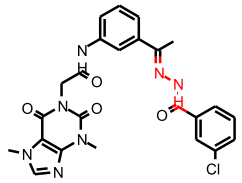
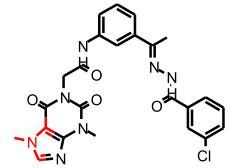
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

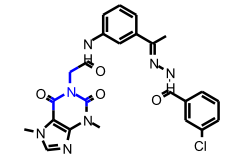
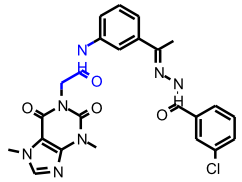
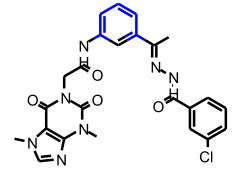
Feature Contribution

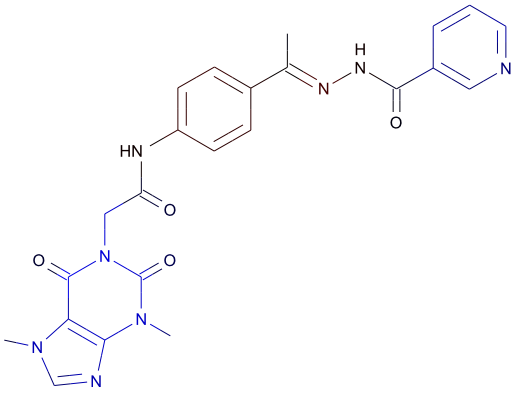
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14

ECFP_6	544048674	 [*]C(=[*])NN=[*]	0.617	2 out of 2
ECFP_6	-407983022	 [*][c]1:[*]:[*]:[cH]: n:1C	0.442	2 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-661097313	 [*]CN(C(=[*])[*])C(=[*])[*]	-1.55	0 out of 12
ECFP_6	1731843802	 [*]CC(=O)N[*]	-0.657	0 out of 3
ECFP_6	2007300961	 [*][c]1:[*]:[c]([*]): [cH]:[cH]:[cH]:1	-0.652	5 out of 34



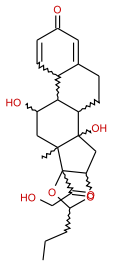
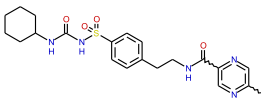
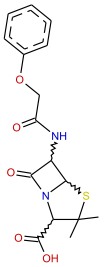
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.214
Enrichment: 0.667
Bayesian Score: -8.9
Mahalanobis Distance: 12.9
Mahalanobis Distance p-value: 0.000644

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

Structural Similar Compounds

Name	Budesonide	Glipizide	Penicillin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.685	0.697	0.711
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

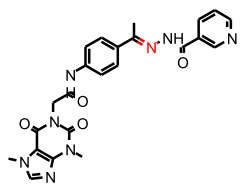
Model Applicability

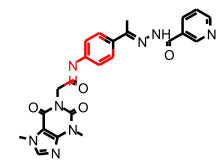
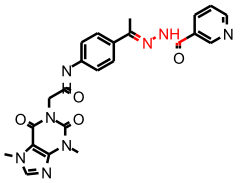
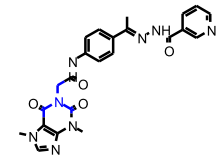
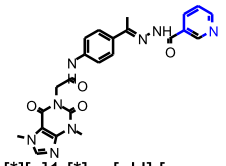
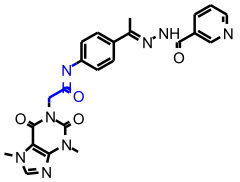
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

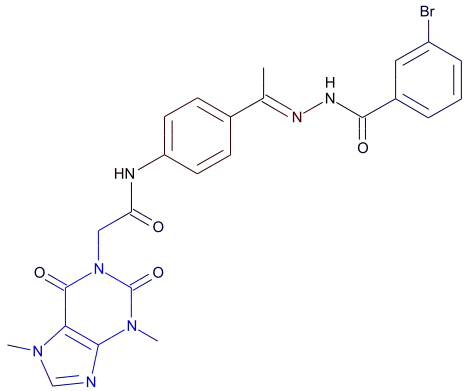
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14

ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	544048674	 [*]C(=[*])NN=[*]	0.617	2 out of 2
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_6	-661097313	 [*]CN(C(=[*])[*])C(=[*])[*]	-1.55	0 out of 12
ECFP_6	2013347047	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.805	0 out of 4
ECFP_6	1731843802	 [*]CC(=O)N[*]	-0.657	0 out of 3



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.212

Enrichment: 0.66

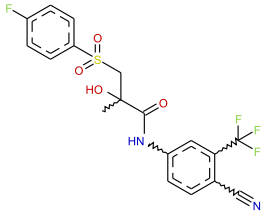
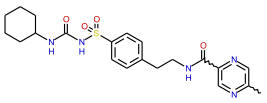
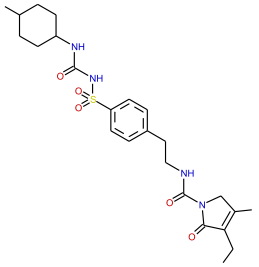
Bayesian Score: -8.58

Mahalanobis Distance: 13.9

Mahalanobis Distance p-value: 2.08e-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	Glipizide	Glimepride
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.696	0.713	0.718
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

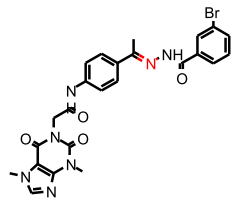
Model Applicability

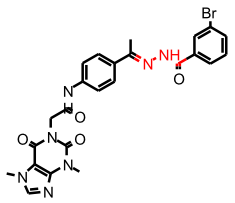
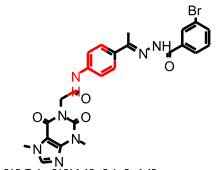
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]
4. Unknown ECFP_2 feature: -177935549: [*]:[cH]:[c](Br):[cH]:[*]

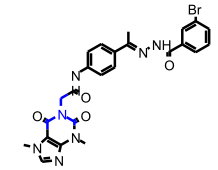
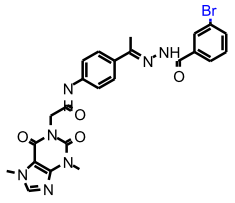
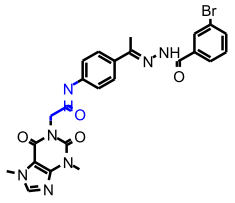
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14

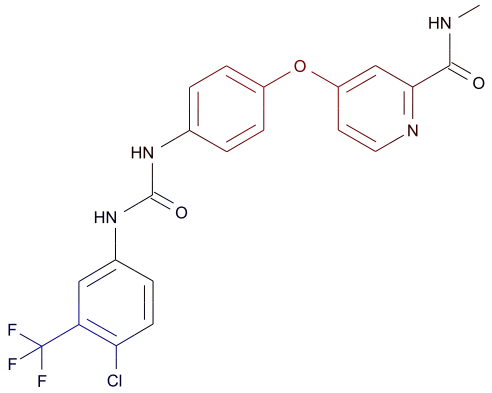
ECFP_6	544048674	 [*]C(=[*])NN=[*]	0.617	2 out of 2
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-661097313	 [*]CN(C(=[*])[*])C(=[*])[*]	-1.55	0 out of 12
ECFP_6	-302078100	 [*]Br	-0.805	0 out of 4
ECFP_6	1731843802	 [*]CC(=O)N[*]	-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

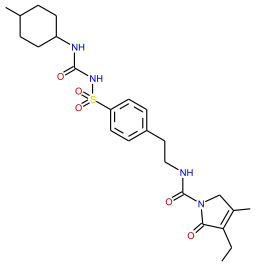
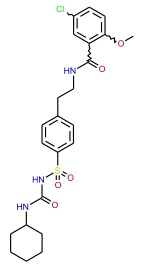
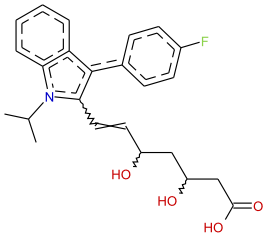
Bayesian Score: -0.321

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

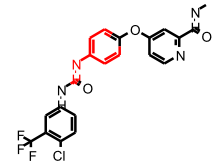
Model Applicability

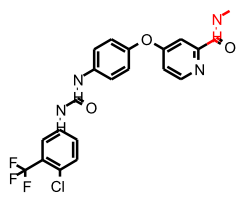
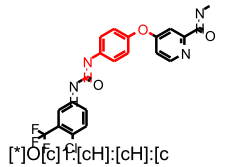
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

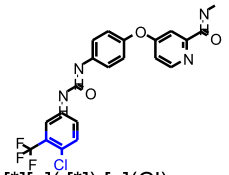
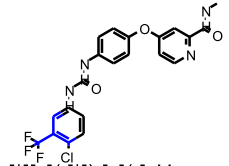
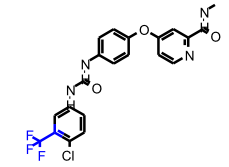
Feature Contribution

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

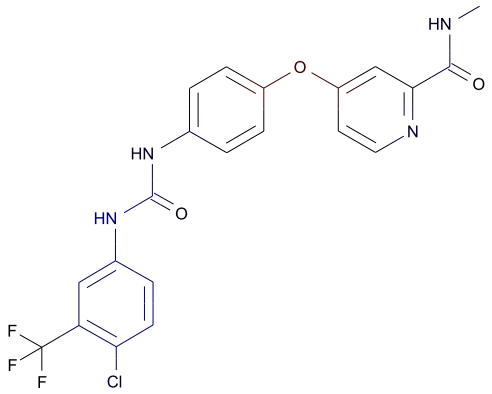
ECFP_6	1338334141	 [*]C(=[*])NC	0.442	2 out of 3
ECFP_6	-335167981	 [*]O[c]1:[cH]:[cH]:[c](NC(=[*])[*]):[cH]: [cH]:1	0.424	1 out of 1

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	1336678434	 [*][c](:[*]):[c](:[cH]:[*])C([*])([*])[*]	-0.657	0 out of 3
ECFP_6	-1952889961	 [*]:[c](:[*])C(F)(F)F	-0.657	0 out of 3

Sorafenib

TOPKAT_Mouse_Female_FDA_Single_vs_Multiple



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.283

Enrichment: 0.691

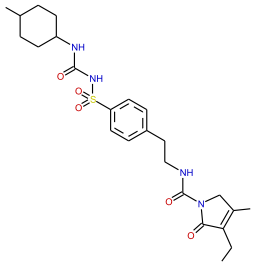
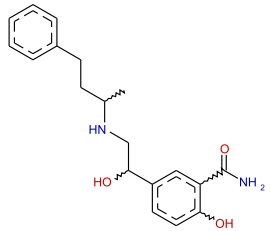
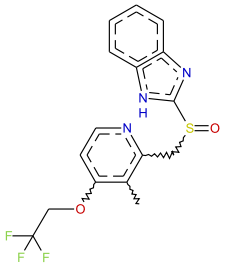
Bayesian Score: -3.89

Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.00221

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

Structural Similar Compounds

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

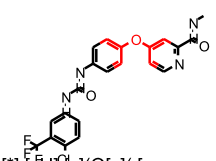
Model Applicability

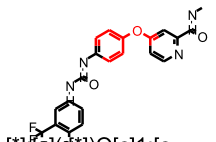
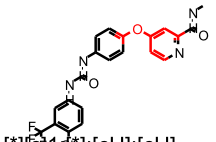
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

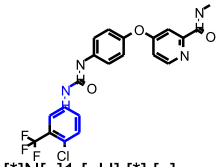
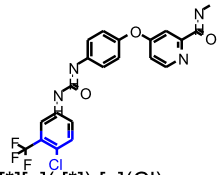
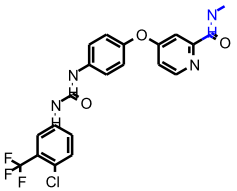
Feature Contribution

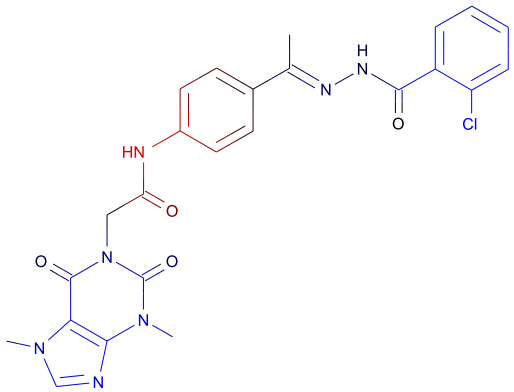
Top features for positive contribution

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

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

Top Features for negative contribution

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



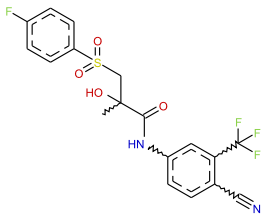
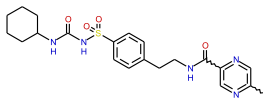
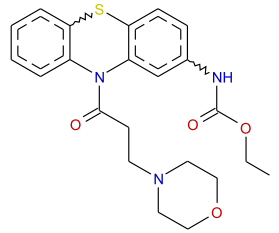
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.137
Enrichment: 0.466
Bayesian Score: -10.5
Mahalanobis Distance: 17
Mahalanobis Distance p-value: 5.83e-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	Bicalutamide	Glipizide	Moricizine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.594	0.654	0.672
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

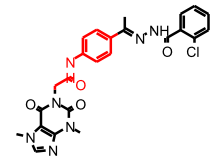
Model Applicability

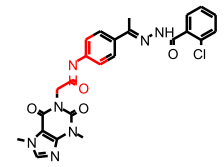
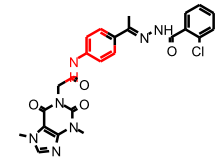
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

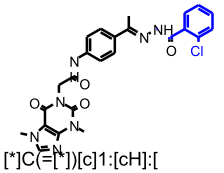
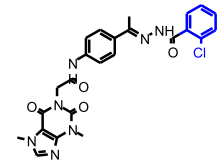
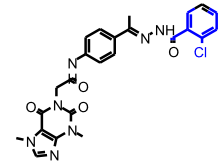
Feature Contribution

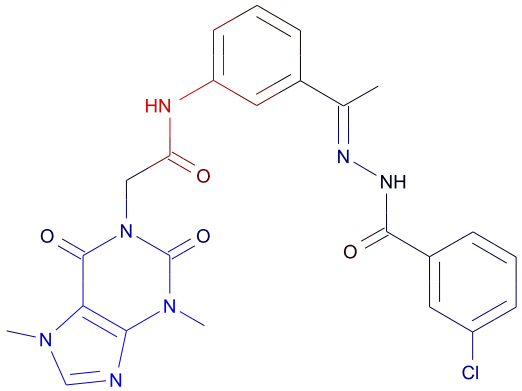
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-451043714	 [*]CC(=O)N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	0.676	2 out of 2

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-543363217	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1Cl	-1.29	0 out of 10
FCFP_6	-2095752315	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[c]:1Cl	-1.13	0 out of 8
FCFP_6	2104062943	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	-1.01	1 out of 17



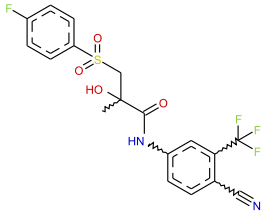
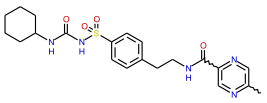
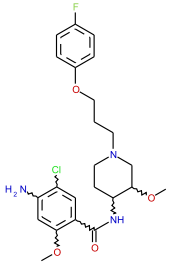
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.164
Enrichment: 0.556
Bayesian Score: -7.66
Mahalanobis Distance: 16.3
Mahalanobis Distance p-value: 1.77e-010

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

Structural Similar Compounds

Name	Bicalutamide	Glipizide	Cisapride
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.587	0.656	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

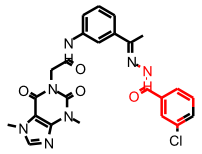
Model Applicability

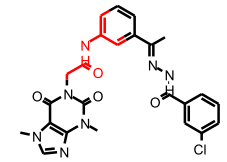
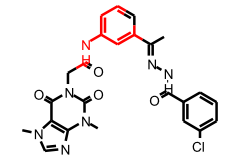
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

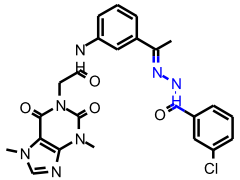
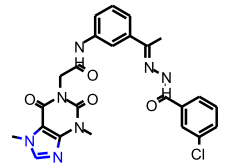
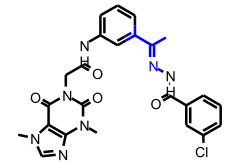
Feature Contribution

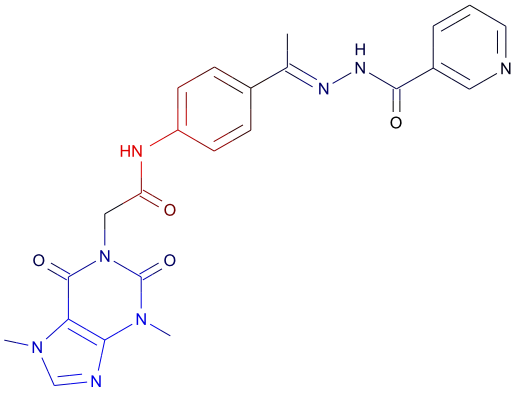
Top features for positive contribution

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

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885520711	 [*]C(=[*])NN=[*]	-0.839	0 out of 5
FCFP_6	-124685461	 [*]n1:[*]:[*]:n:[cH]:1	-0.731	1 out of 12
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21



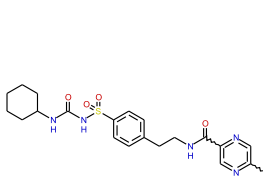
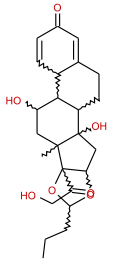
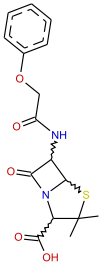
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.201
Enrichment: 0.684
Bayesian Score: -4.9
Mahalanobis Distance: 14.9
Mahalanobis Distance p-value: 1.25e-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	Glipizide	Budesonide	Penicillin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.688	0.692	0.717
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

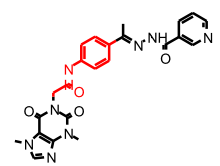
Model Applicability

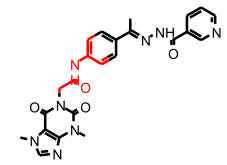
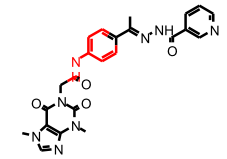
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

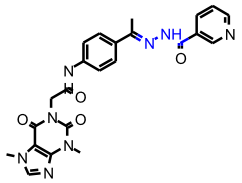
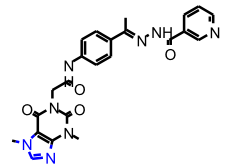
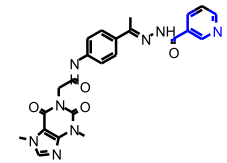
Feature Contribution

Top features for positive contribution

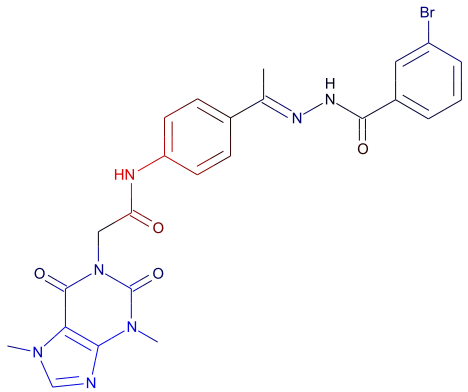
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-451043714	 [*]CC(=O)N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	0.676	2 out of 2

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885520711	 [*]C(=[*])NN=[*]	-0.839	0 out of 5
FCFP_6	-124685461	 [*]n1:[*]:[*]:n:[cH]: 1	-0.731	1 out of 12
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.167

Enrichment: 0.567

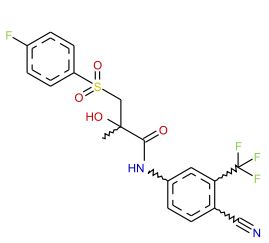
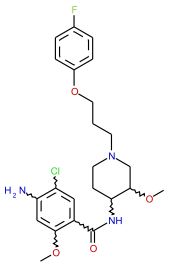
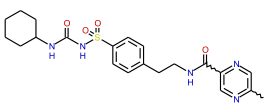
Bayesian Score: -7.38

Mahalanobis Distance: 16.7

Mahalanobis Distance p-value: 3.44e-011

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

Structural Similar Compounds

Name	Bicalutamide	Cisapride	Glipizide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.650	0.697	0.706
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

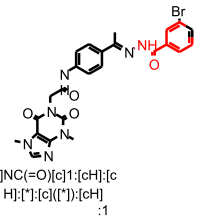
Model Applicability

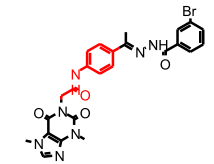
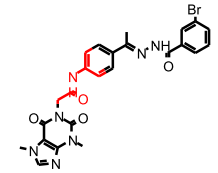
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

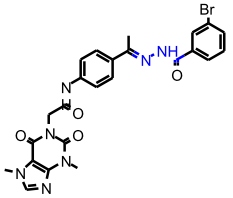
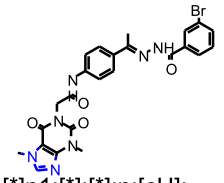
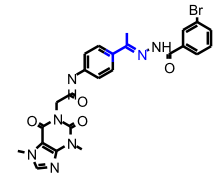
Feature Contribution

Top features for positive contribution

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

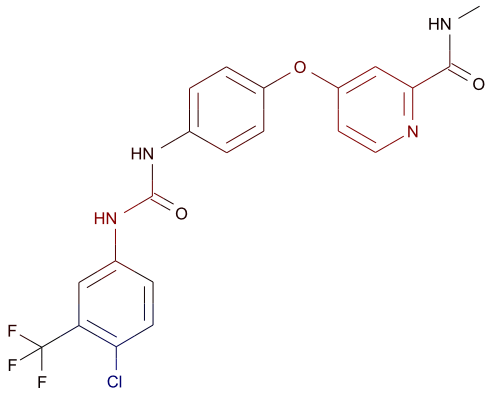
FCFP_6	-451043714	 [*]CC(=O)N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	0.676	2 out of 2
FCFP_6	1175665944	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	0.655	7 out of 12

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885520711	 [*]C(=[*])NN=[*]	-0.839	0 out of 5
FCFP_6	-124685461	 [*]n1:[*]:[*]:n:[cH]:1	-0.731	1 out of 12
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21

Sorafenib

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.444

Enrichment: 1.51

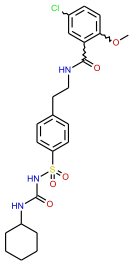
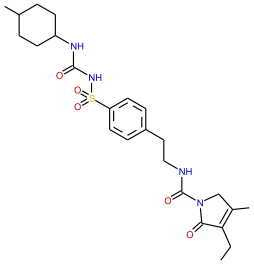
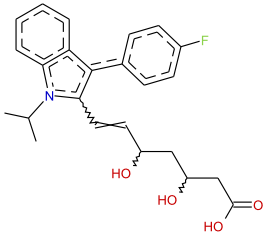
Bayesian Score: 4.21

Mahalanobis Distance: 20.3

Mahalanobis Distance p-value: 1.28e-019

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

Structural Similar Compounds

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

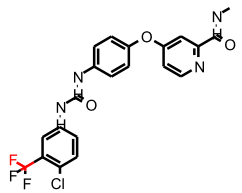
Model Applicability

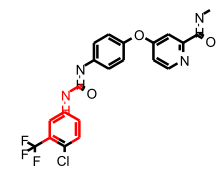
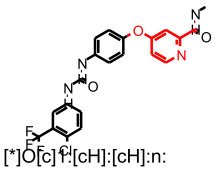
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

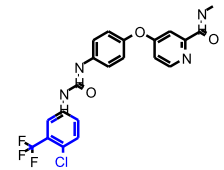
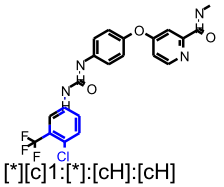
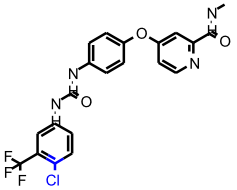
Feature Contribution

Top features for positive contribution

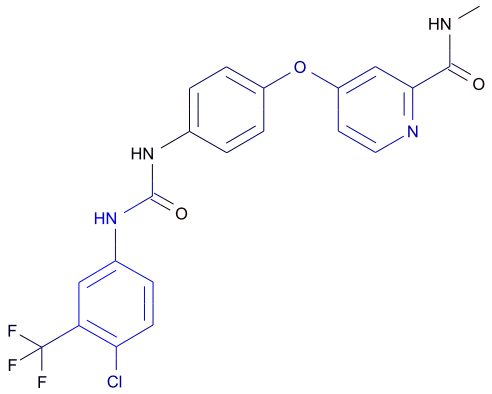
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	71953198	 [*]C([*])([*])F	0.612	12 out of 23

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

Top Features for negative contribution

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

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.139

Enrichment: 0.461

Bayesian Score: -14.7

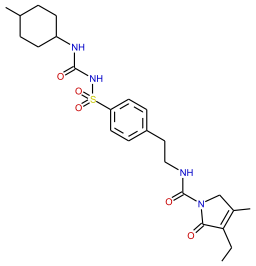
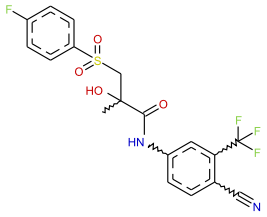
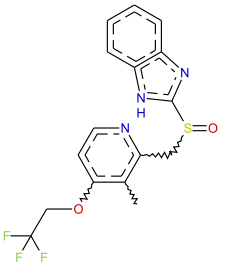
Mahalanobis Distance: 21.3

Mahalanobis Distance p-value: 4.93e-011

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

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple

Structural Similar Compounds

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

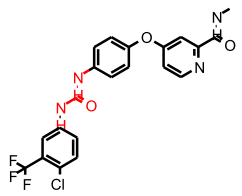
Model Applicability

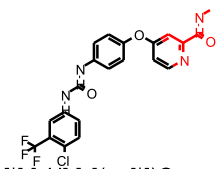
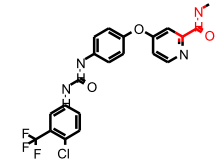
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

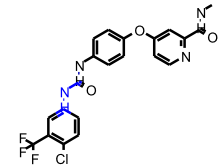
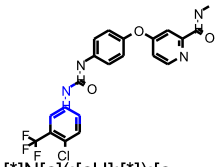
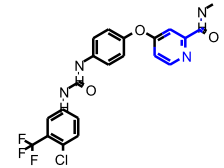
Feature Contribution

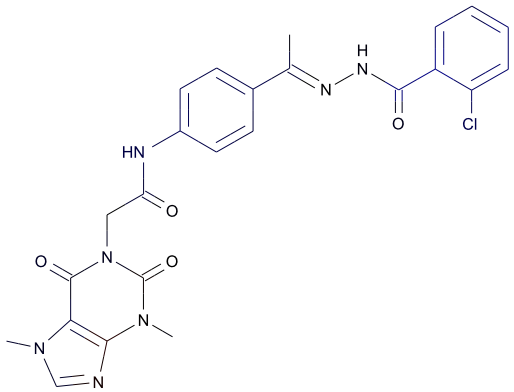
Top features for positive contribution

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

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

Top Features for negative contribution

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



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.575

Enrichment: 0.834

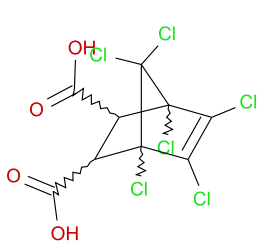
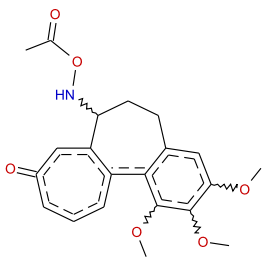
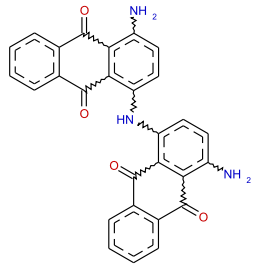
Bayesian Score: -5.32

Mahalanobis Distance: 11.2

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.

Structural Similar Compounds

Name	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	COLCHICINE	4;4'-DIAMINO-1;1'-DIANTHRIMIDE
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild
Distance	0.788	0.806	0.809
Reference	28ZPAK-;92;72	AJOPAA 31;837;48	28ZPAK-;125;72

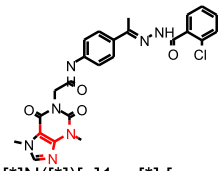
Model Applicability

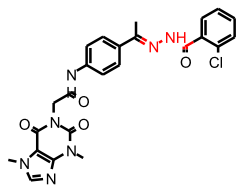
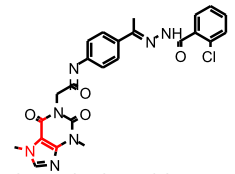
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
4. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]

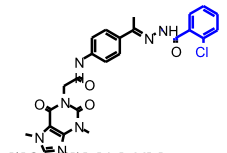
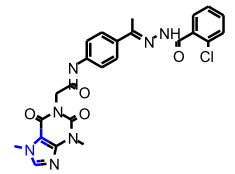
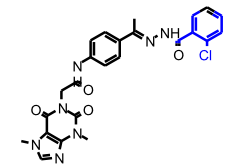
Feature Contribution

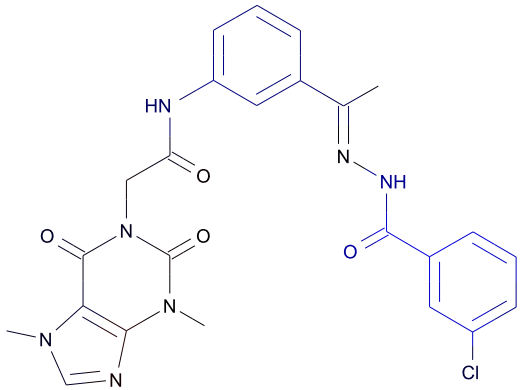
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1410049896	 [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]	0.256	2 out of 2

FCFP_10	-885520711	 [*]C(=[*])NN=[*]	0.256	2 out of 2
FCFP_10	-1539132615	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	0.224	11 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-543363217	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[c]:1C I	-1.01	2 out of 11
FCFP_10	-306856457	 [*][c]1:[*]:[*]:[cH]:n:1C	-0.842	0 out of 2
FCFP_10	2104062943	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	-0.745	7 out of 24



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.306

Enrichment: 0.445

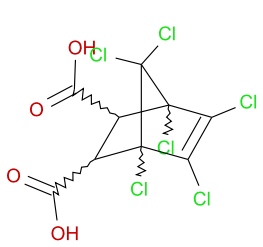
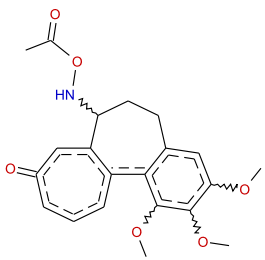
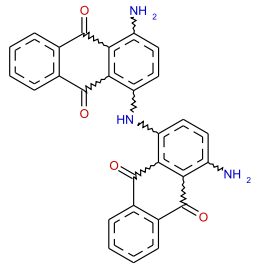
Bayesian Score: -8.13

Mahalanobis Distance: 11.2

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.

Structural Similar Compounds

Name	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-	COLCHICINE	4;4'-DIAMINO-1;1'-DIANTHRIMIDE
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Moderate_Severe	Moderate_Severe	Mild
Distance	0.788	0.806	0.813
Reference	28ZPAK-;92;72	AJOPAA 31;837;48	28ZPAK-;125;72

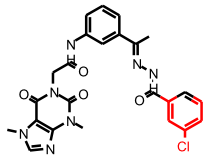
Model Applicability

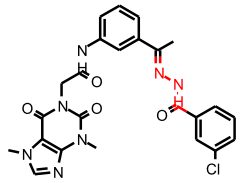
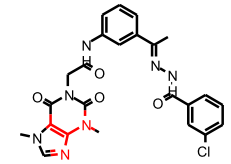
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]
4. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

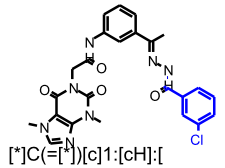
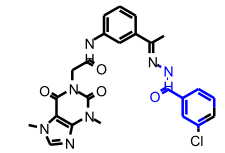
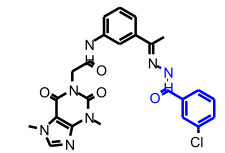
Feature Contribution

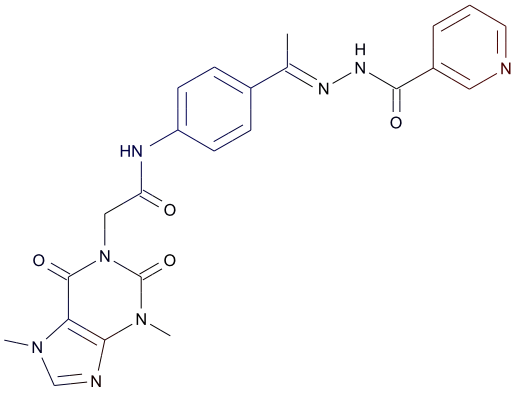
Top features for positive contribution

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

FCFP_10	-885520711	 [*]C(=[*])NN=[*]	0.256	2 out of 2
FCFP_10	-1410049896	 [*]N([*])[c]1:n:[*]:[*]:[*]:[c]:1[*]	0.256	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1700637232	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[c](Cl):[cH]:1	-1.34	1 out of 10
FCFP_10	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	-1.29	0 out of 4
FCFP_10	-1205069278	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[c]([*]):[cH]:1	-1.29	0 out of 4



$C_{23}H_{22}N_8O_4$

Molecular Weight: 474.47198

ALogP: 0.219

Rotatable Bonds: 6

Acceptors: 7

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.783

Enrichment: 1.14

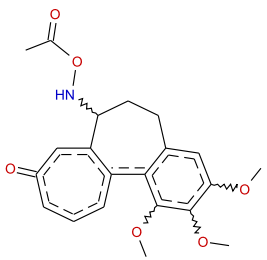
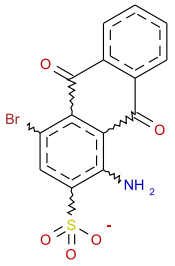
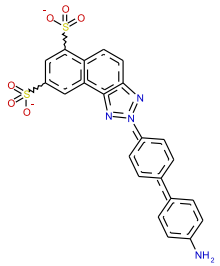
Bayesian Score: -1.59

Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.0193

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

Structural Similar Compounds

Name	COLCHICINE	2-Anthracenesulfonic acid; 9;10-dihydro-1-amino-4-bromo-9;10-dioxo-; sodium	2H-Naphtho(1;2-d)triazole-6;8-disulfonic acid;
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.840	0.851	0.854
Reference	AJOPAA 31;837;48	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1062;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1065;86

Model Applicability

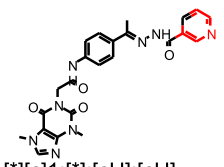
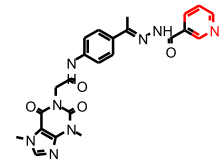
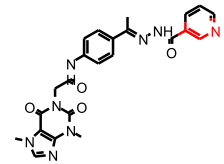
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
4. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]

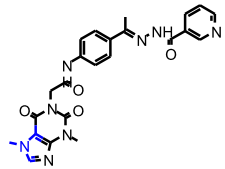
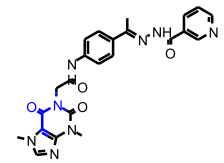
Feature Contribution

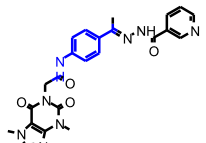
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set

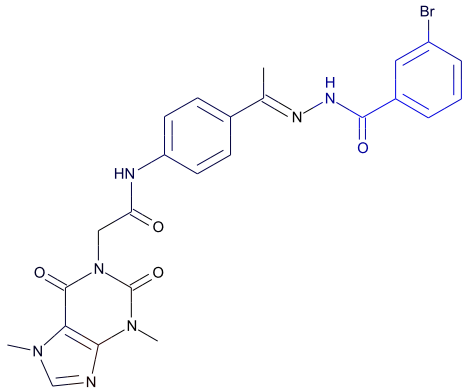
FCFP_10	547884906	 [*][c]1:[*]:[cH]:[cH] :n:[cH]:1	0.317	4 out of 4
FCFP_10	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.285	10 out of 11
FCFP_10	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.259	14 out of 16

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-306856457	 [*][c]1:[*]:[*]:[cH]: n:1C	-0.842	0 out of 2
FCFP_10	-1549163031	 [*]N([*])C(=O)[c](:[*]):[*]	-0.657	5 out of 16

FCFP_10	-790336137	 [*]C(=[*])N[c]1:[cH]:[cH]:[c](:[cH]:[cH]:1)C(=[*])[*]	-0.507	0 out of 1

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.375

Enrichment: 0.544

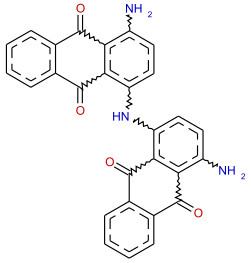
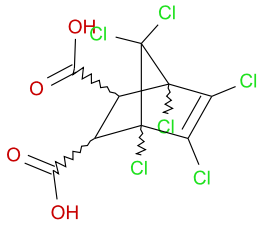
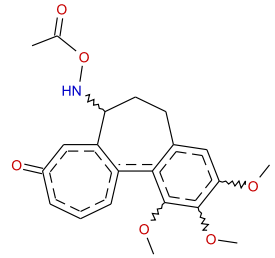
Bayesian Score: -7.44

Mahalanobis Distance: 11.7

Mahalanobis Distance p-value: 0.000241

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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-	COLCHICINE
Structure			
Actual Endpoint	Mild	Moderate_Severe	Moderate_Severe
Predicted Endpoint	Mild	Moderate_Severe	Moderate_Severe
Distance	0.835	0.843	0.858
Reference	28ZPAK-;125;72	28ZPAK-;92;72	AJOPAA 31;837;48

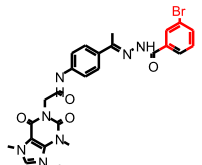
Model Applicability

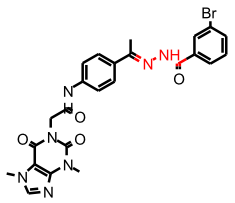
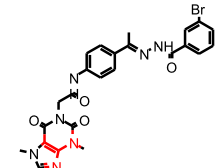
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
4. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]

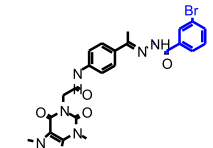
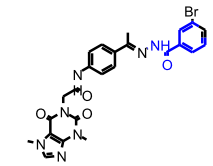
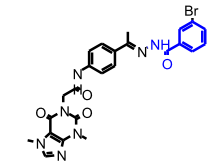
Feature Contribution

Top features for positive contribution

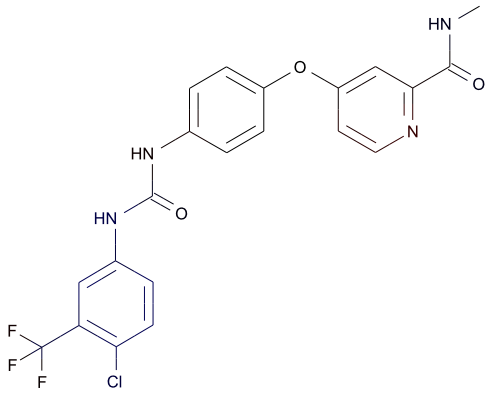
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-745491832	 [*][c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	0.304	29 out of 32

FCFP_10	-885520711	 [*]C(=[*])NN=[*]	0.256	2 out of 2
FCFP_10	-1410049896	 [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]	0.256	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1700637232	 [*]C(=[N])[c]1:[cH]:[cH]:[cH]:[c](Cl):[cH]:1	-1.34	1 out of 10
FCFP_10	-581879738	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:[c]([*]):[cH]:1	-1.29	0 out of 4
FCFP_10	-1205069278	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[c]([*]):[cH]:1	-1.29	0 out of 4

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.776

Enrichment: 1.13

Bayesian Score: -1.8

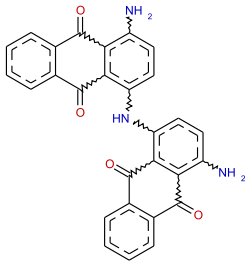
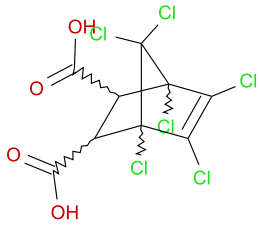
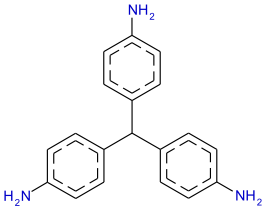
Mahalanobis Distance: 8.95

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.

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe

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

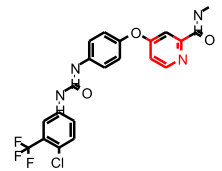
Model Applicability

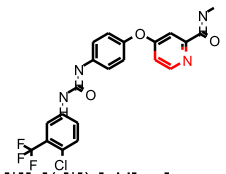
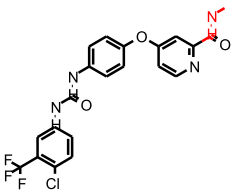
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

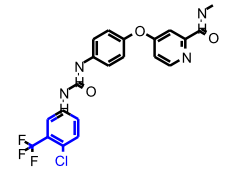
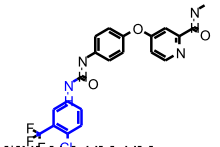
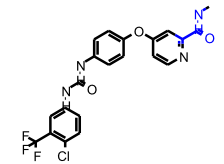
Feature Contribution

Top features for positive contribution

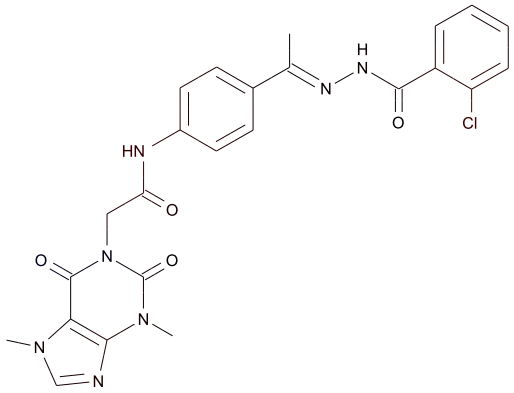
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.285	10 out of 11

FCFP_10	-124655670	 [*][c](:[*]):[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

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

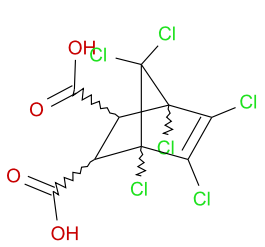
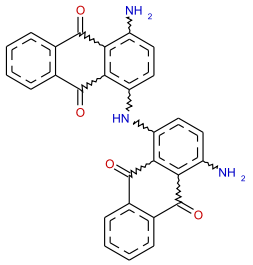
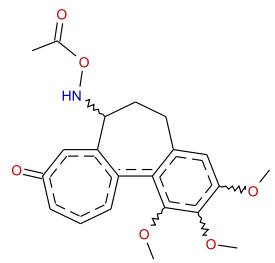
Bayesian Score: 2.82

Mahalanobis Distance: 9.61

Mahalanobis Distance p-value: 0.21

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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-	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	COLCHICINE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.773	0.795	0.800
Reference	28ZPAK-;92;72	28ZPAK-;125;72	AJOPAA 31;837;48

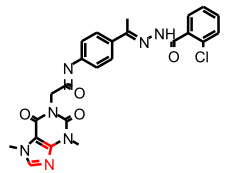
Model Applicability

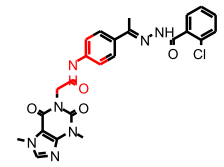
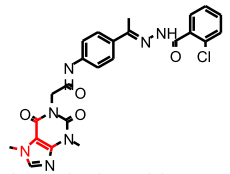
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
4. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]

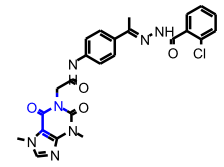
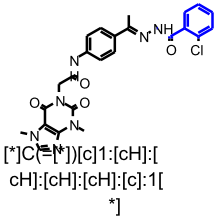
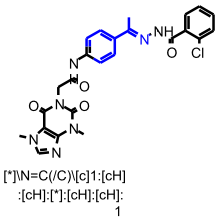
Feature Contribution

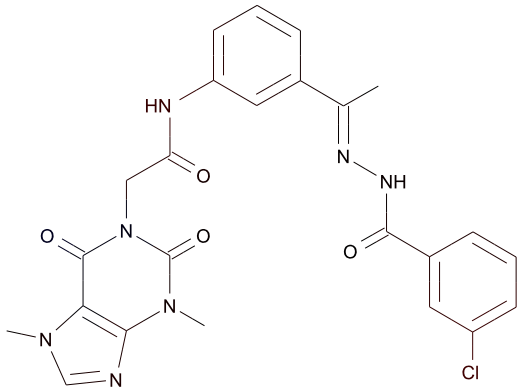
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c]1:[*]:[*]:[cH]:n:1	0.208	44 out of 44

FCFP_12	1175665944		0.198	14 out of 14
FCFP_12	-1539132615		0.197	13 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1549163031		-0.623	16 out of 38
FCFP_12	-1698724694		-0.0964	107 out of 146
FCFP_12	-581162801		0	7 out of 9



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

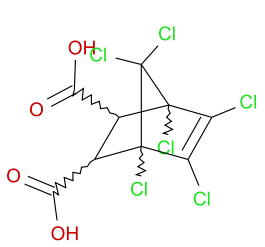
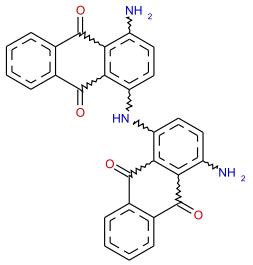
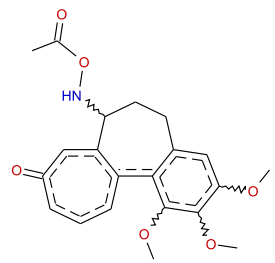
Bayesian Score: 3.04

Mahalanobis Distance: 9.61

Mahalanobis Distance p-value: 0.21

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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-	4;4'-DIAMINO-1;1'-DIANTHRIMIDE	COLCHICINE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.773	0.799	0.800
Reference	28ZPAK-;92;72	28ZPAK-;125;72	AJOPAA 31;837;48

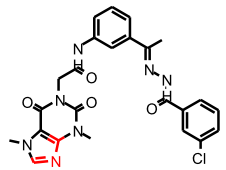
Model Applicability

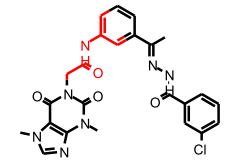
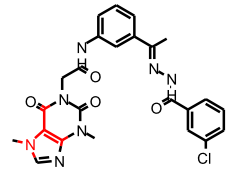
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]
4. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

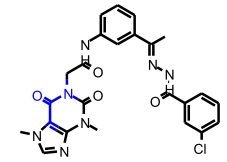
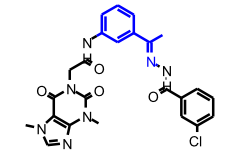
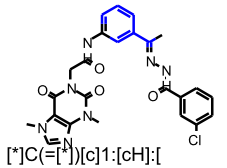
Feature Contribution

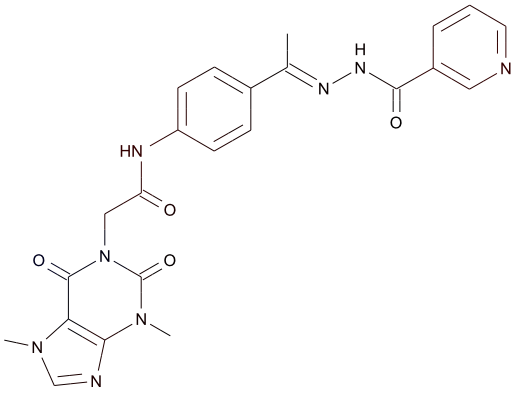
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c]1:[*]:[*]:[cH]:n:1	0.208	44 out of 44

FCFP_12	1175665944	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	0.198	14 out of 14
FCFP_12	-1539132615	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	0.197	13 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1549163031	 [*]N([*])C(=O)[c](:[*]):[*]	-0.623	16 out of 38
FCFP_12	975909016	 [*]\N=C(/C)\[c]1:[cH]:[cH]:[cH]:[c]([*]):[cH]:1	-0.0639	6 out of 8
FCFP_12	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0	264 out of 323



$C_{23}H_{22}N_8O_4$

Molecular Weight: 474.47198

ALogP: 0.219

Rotatable Bonds: 6

Acceptors: 7

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

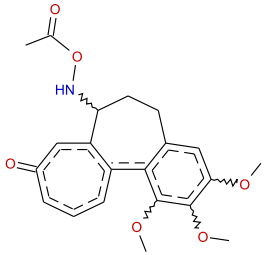
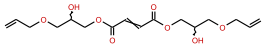
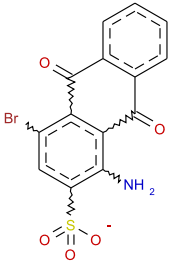
Bayesian Score: 2.92

Mahalanobis Distance: 9.56

Mahalanobis Distance p-value: 0.232

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

Structural Similar Compounds

Name	COLCHICINE	Fumaric acid; bis(3-allyloxy-2-hydroxypropyl) ester	2-Anthracenesulfonic acid; 9;10-dihydro-1-amino-4-bromo-9;10-dioxo-; sodium
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.824	0.833	0.844
Reference	AJOPAA 31;837;48	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;646;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1062;86

Model Applicability

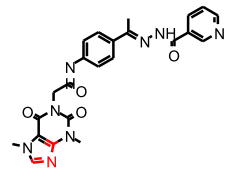
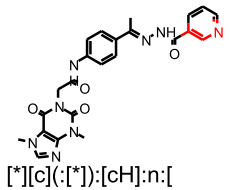
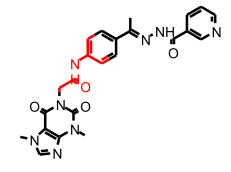
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
4. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]

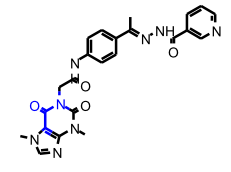
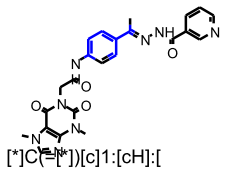
Feature Contribution

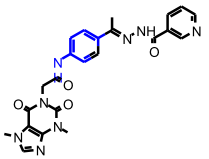
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

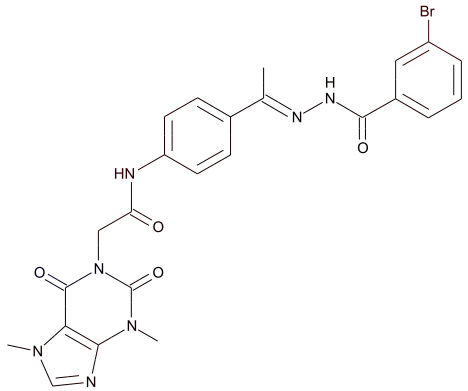
FCFP_12	1747237384	 [*][c]1:[*]:[*]:[cH]:n:1	0.208	44 out of 44
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16
FCFP_12	1175665944	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	0.198	14 out of 14

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1549163031	 [*]N([*])C(=O)[c](:[*]):[*]	-0.623	16 out of 38
FCFP_12	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0	264 out of 323

FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0	102 out of 121

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

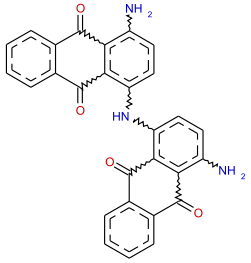
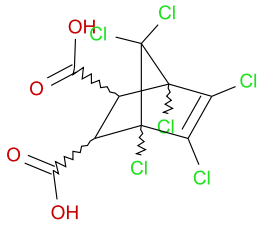
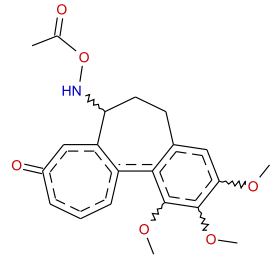
Bayesian Score: 3.05

Mahalanobis Distance: 10.4

Mahalanobis Distance p-value: 0.0308

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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-	COLCHICINE
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.821	0.826	0.848
Reference	28ZPAK-;125;72	28ZPAK-;92;72	AJOPAA 31;837;48

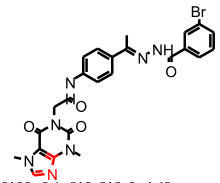
Model Applicability

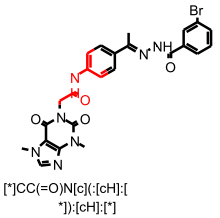
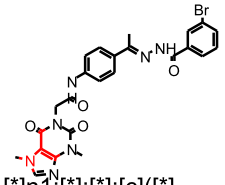
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
4. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]

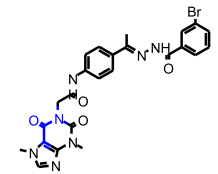
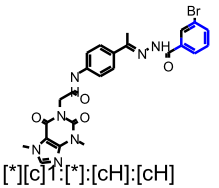
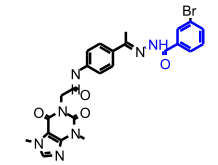
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c]1:[*]:[*]:[cH]:n:1	0.208	44 out of 44

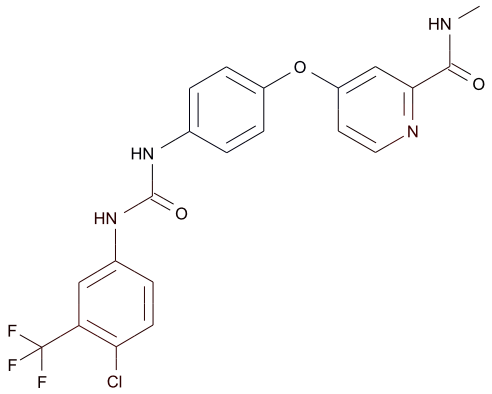
FCFP_12	1175665944	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	0.198	14 out of 14
FCFP_12	-1539132615	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	0.197	13 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1549163031	 [*]N([*])C(=O)[c](:[*]):[*]	-0.623	16 out of 38
FCFP_12	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	-1205069278	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[c]([*]):[cH]:1	0	4 out of 5

Sorafenib

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

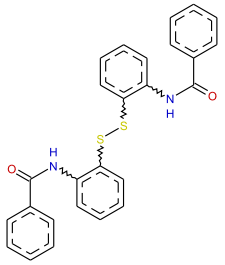
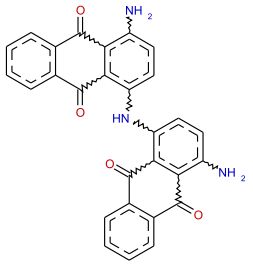
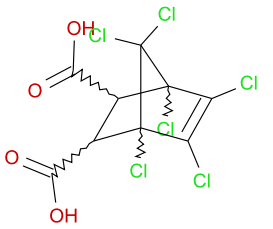
Bayesian Score: 3.04

Mahalanobis Distance: 6.28

Mahalanobis Distance p-value: 1

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

Structural Similar Compounds

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

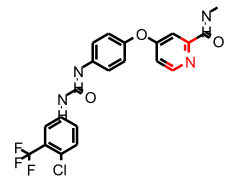
Model Applicability

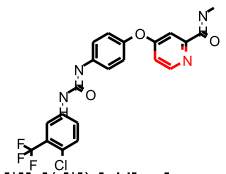
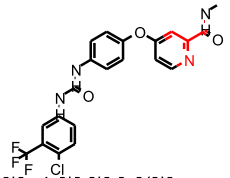
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

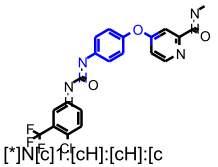
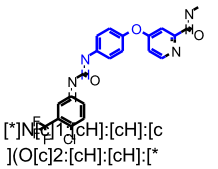
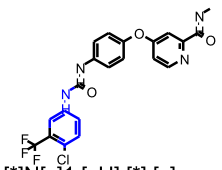
Feature Contribution

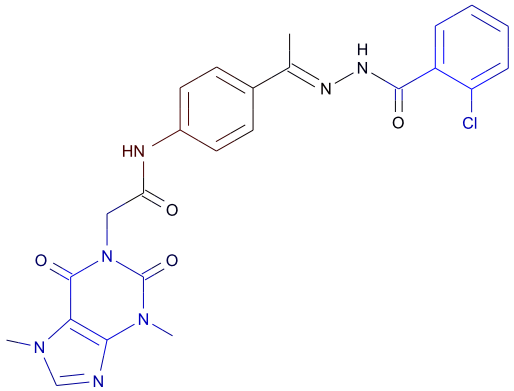
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c]1:[*]:[*]:[cH]:n:1	0.208	44 out of 44

FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1539132615	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	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](:[*]):[*]):[c H]:[cH]:1	-0.268	1 out of 2
FCFP_12	702861189	 [*]N[c]1:[cH]:[cH]:[c](O[c]2:[cH]:[cH]:[*]:[c]([*]):[cH]:2):[cH]:[cH]:1	-0.268	1 out of 2
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121



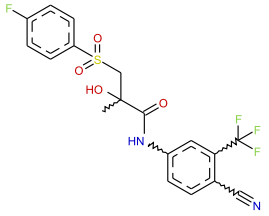
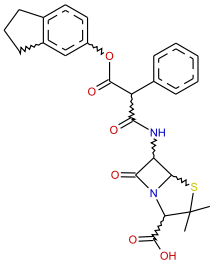
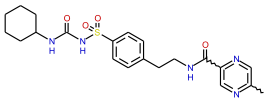
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.174
Enrichment: 0.539
Bayesian Score: -12
Mahalanobis Distance: 12.6
Mahalanobis Distance p-value: 0.000729

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Carbenicillin	Glipizide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.658	0.665	0.678
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

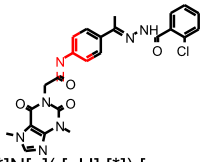
Model Applicability

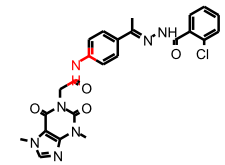
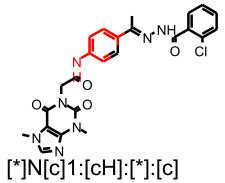
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]

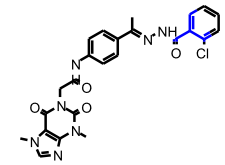
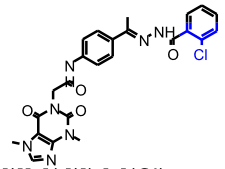
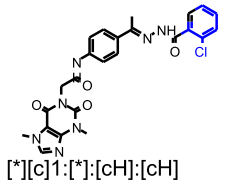
Feature Contribution

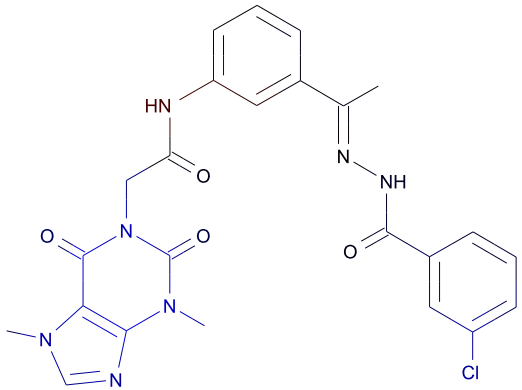
Top features for positive contribution

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	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1337040050	 [*]C(=[*])[c](:[cH]:[*]):[c]([*]):[*]	-1.84	0 out of 17
ECFP_12	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-1.11	2 out of 26
ECFP_12	1641317964	 [*][c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.929	1 out of 13



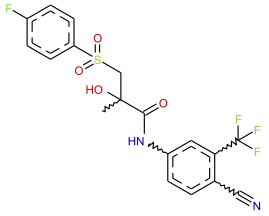
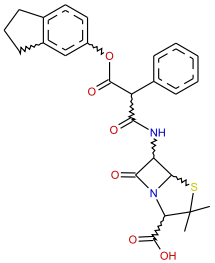
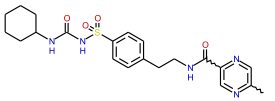
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.187
Enrichment: 0.582
Bayesian Score: -8.96
Mahalanobis Distance: 18.3
Mahalanobis Distance p-value: 1.54e-016

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Carbenicillin	Glipizide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.656	0.665	0.678
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

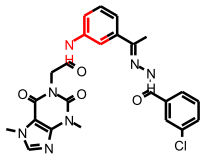
Model Applicability

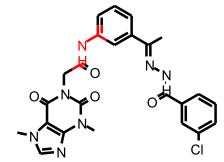
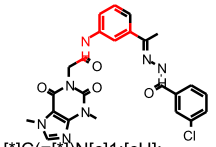
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]

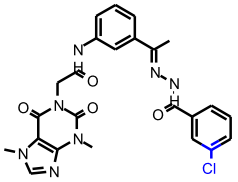
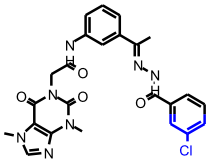
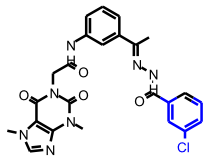
Feature Contribution

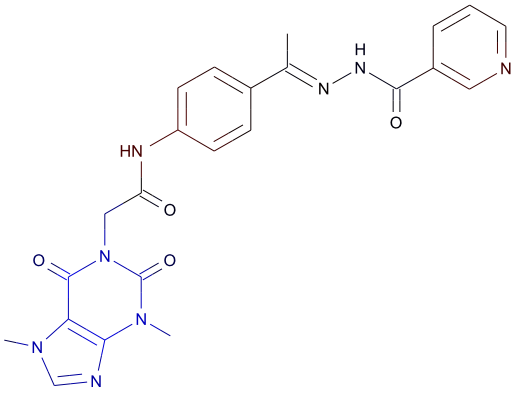
Top features for positive contribution

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	99947387	 [*]:[c](:[*])Cl	-0.817	8 out of 62
ECFP_12	-176494269	 [*]:[cH]:[c](Cl):[cH]:[*]	-0.714	5 out of 36
ECFP_12	577592657	 [*][c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	-0.567	4 out of 25



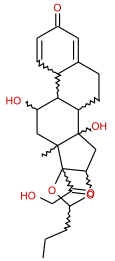
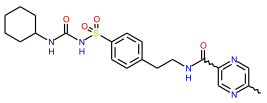
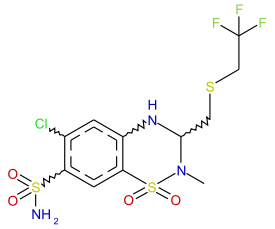
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.246
Enrichment: 0.764
Bayesian Score: -3.06
Mahalanobis Distance: 12.9
Mahalanobis Distance p-value: 0.000226

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

Structural Similar Compounds

Name	Budesonide	Glipizide	Polythiazide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.694	0.717	0.721
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

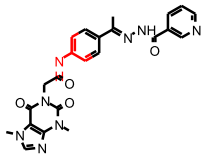
Model Applicability

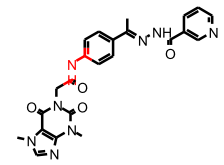
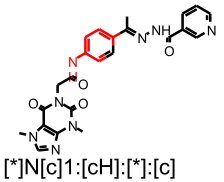
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]

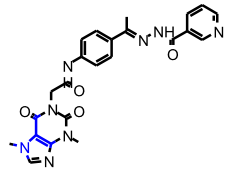
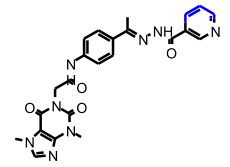
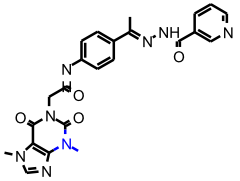
Feature Contribution

Top features for positive contribution

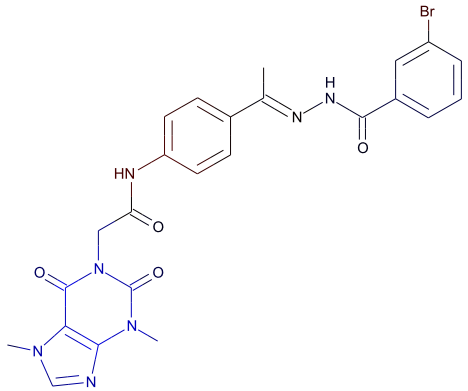
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10

ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]): [*]	0.46	9 out of 17
ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0.454	5 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-813242890	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	-0.485	0 out of 2
ECFP_12	1997021792	 [*]:[cH]:[cH]:[cH]:[*]	-0.296	36 out of 156
ECFP_12	866343404	 [*]N([*])C	-0.281	4 out of 18

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.223

Enrichment: 0.691

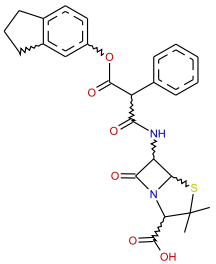
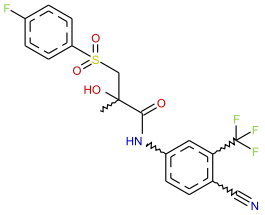
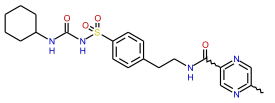
Bayesian Score: -4.88

Mahalanobis Distance: 13.6

Mahalanobis Distance p-value: 1.11e-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	Carbenicillin	Bicalutamide	Glipizide
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.684	0.717	0.730
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

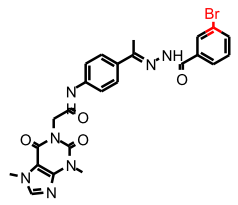
Model Applicability

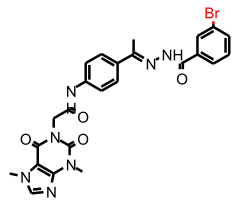
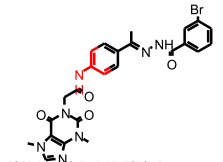
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: -177935549: [*]:[cH]:[c](Br):[cH]:[*]

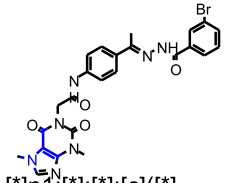
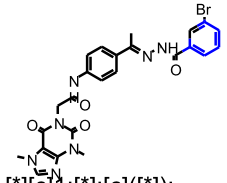
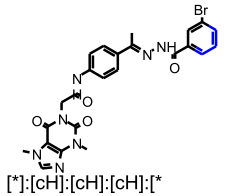
Feature Contribution

Top features for positive contribution

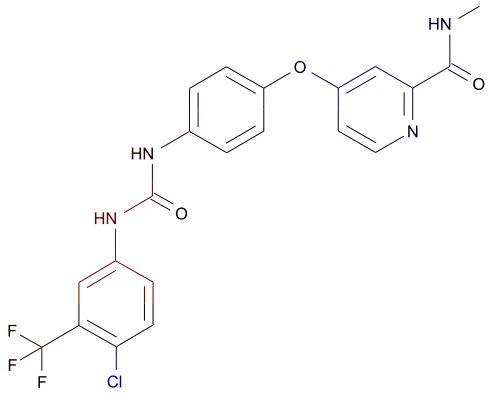
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	459826767	 [*]:[c](:[*])Br	0.613	2 out of 2

ECFP_12	-302078100	 [*]Br	0.575	3 out of 4
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-813242890	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	-0.485	0 out of 2
ECFP_12	2007300961	 [*][c]1:[*]:[c]([*]):[cH]:[cH]:[cH]:1	-0.426	7 out of 36
ECFP_12	1997021792	 [*]:[cH]:[cH]:[cH]:[*]	-0.296	36 out of 156

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.236

Enrichment: 0.734

Bayesian Score: -3.76

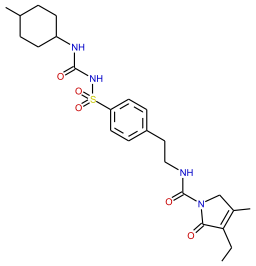
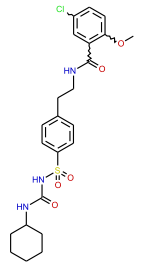
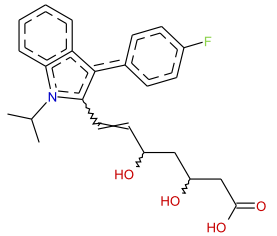
Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 0.00229

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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen

Structural Similar Compounds

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

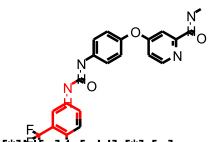
Model Applicability

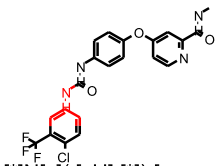
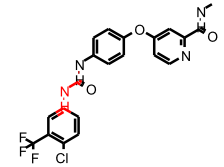
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

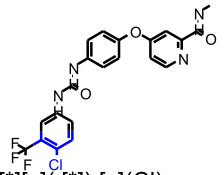
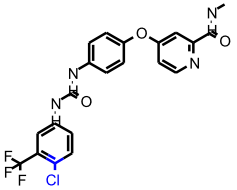
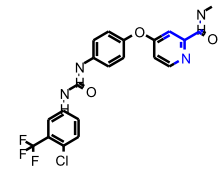
Feature Contribution

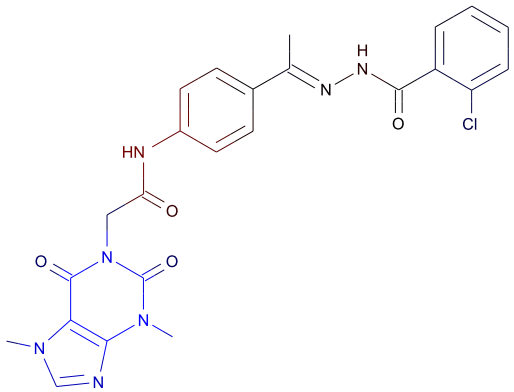
Top features for positive contribution

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

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

Top Features for negative contribution

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



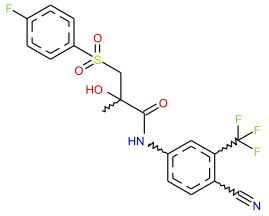
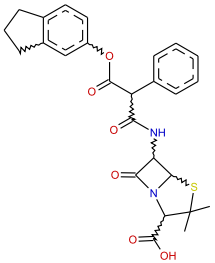
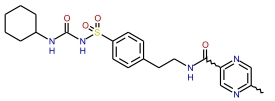
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.199
Enrichment: 0.594
Bayesian Score: -7.44
Mahalanobis Distance: 15.3
Mahalanobis Distance p-value: 1.43e-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	Carbenicillin	Glipizide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.637	0.657	0.658
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

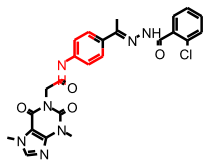
Model Applicability

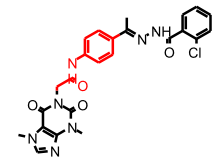
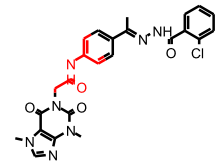
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

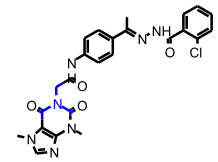
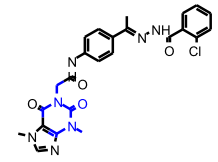
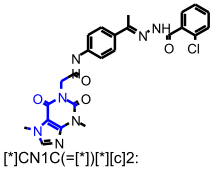
Feature Contribution

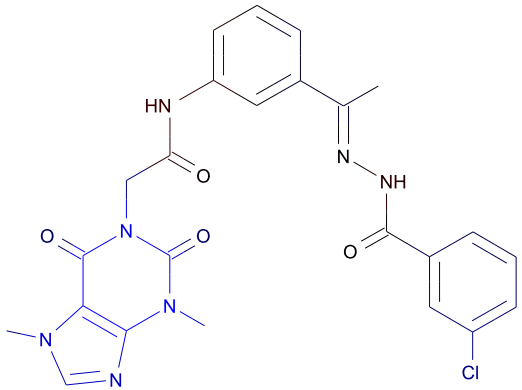
Top features for positive contribution

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

SCFP_6	814408713	 [*]CC(=O)N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	0.603	2 out of 2
SCFP_6	2097618059	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	0.437	7 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	399659969	 [*]CN(C(=[*])[*])C(=[*])[*]	-0.578	1 out of 8
SCFP_6	-2121589288	 [*]CN1C(=[*])[*]:[c](:[*])N(C)C1=O	-0.496	0 out of 2
SCFP_6	1445006032	 [*]CN1C(=[*])[*][c]2:[*]:[*]:n([*]):[c]:2C1=O	-0.496	0 out of 2



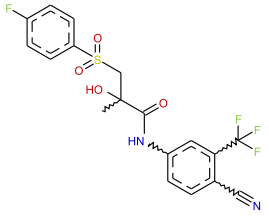
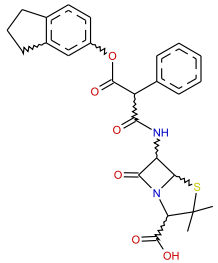
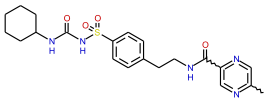
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.17
Enrichment: 0.508
Bayesian Score: -9.69
Mahalanobis Distance: 15.7
Mahalanobis Distance p-value: 2.82e-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	Carbenicillin	Glipizide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.638	0.657	0.659
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

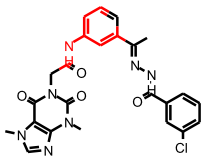
Model Applicability

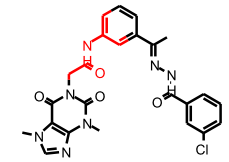
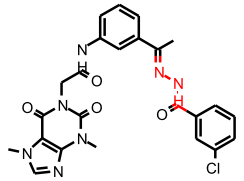
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

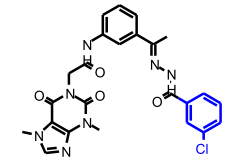
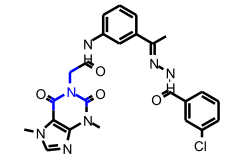
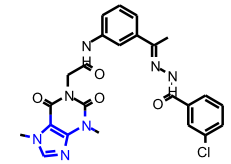
Feature Contribution

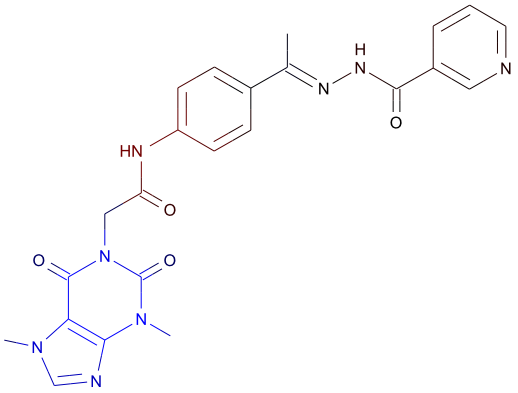
Top features for positive contribution

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

SCFP_6	2097618059	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	0.437	7 out of 13
SCFP_6	698322229	 [*]C(=[*])NN=[*]	0.415	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	52043406	 [*][c]1:[cH]:[cH]:[cH]:[c](Cl):[cH]:1	-0.674	0 out of 3
SCFP_6	399659969	 [*]CN(C(=[*])[*])C(=[*])[*]	-0.578	1 out of 8
SCFP_6	2144809592	 [*]N([*])[c]1:n:[cH]:n([*]):[c]:1[*]	-0.496	0 out of 2



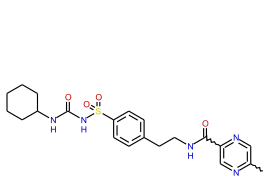
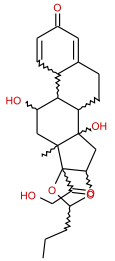
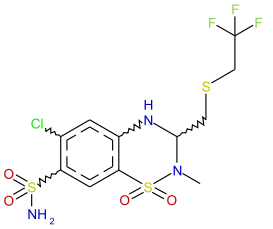
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.259
Enrichment: 0.775
Bayesian Score: -3.96
Mahalanobis Distance: 16.6
Mahalanobis Distance p-value: 2.55e-010

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

Structural Similar Compounds

Name	Glipizide	Budesonide	Polythiazide
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.694	0.696	0.713
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

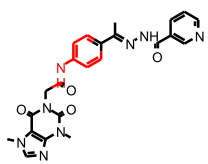
Model Applicability

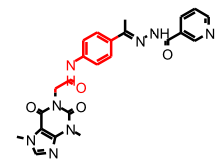
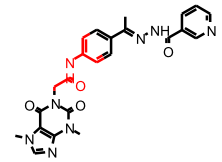
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

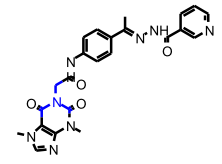
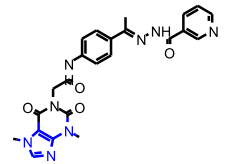
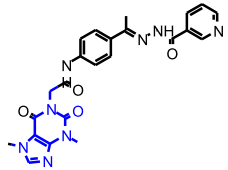
Feature Contribution

Top features for positive contribution

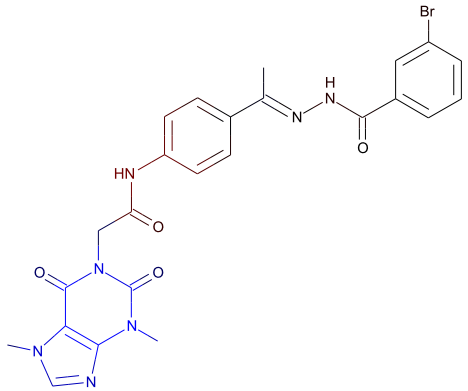
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.615	5 out of 7

SCFP_6	814408713	 [*]CC(=O)N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	0.603	2 out of 2
SCFP_6	2097618059	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	0.437	7 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	399659969	 [*]CN(C(=[*])[*])C(=[*])[*]	-0.578	1 out of 8
SCFP_6	2144809592	 [*]N([*])[c]1:n:[cH]:n([*]):[c]:1[*]	-0.496	0 out of 2
SCFP_6	-1501735365	 [*]CN1C(=O)N(C)[c]2:n:[cH]:n([*]):[c]:2C1=[*]	-0.496	0 out of 2

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.23

Enrichment: 0.687

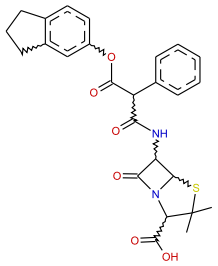
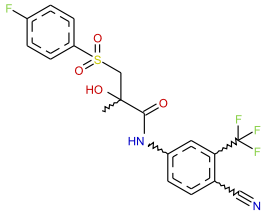
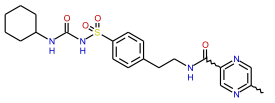
Bayesian Score: -5.5

Mahalanobis Distance: 16.3

Mahalanobis Distance p-value: 1.33e-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	Carbenicillin	Bicalutamide	Glipizide
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.677	0.699	0.712
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

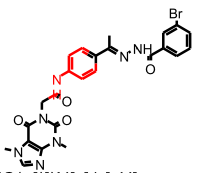
Model Applicability

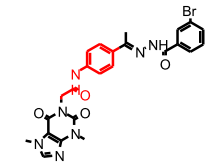
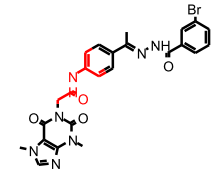
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

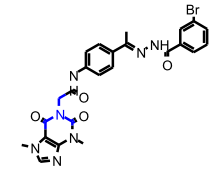
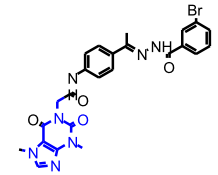
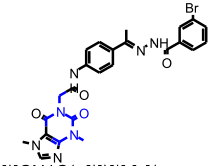
Feature Contribution

Top features for positive contribution

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

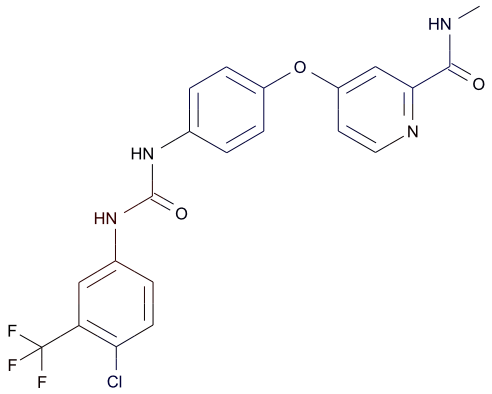
SCFP_6	814408713	 [*]CC(=O)N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	0.603	2 out of 2
SCFP_6	2097618059	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	0.437	7 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	399659969	 [*]CN(C(=[*])[*])C(=[*])[*]	-0.578	1 out of 8
SCFP_6	-1501735365	 [*]CN1C(=O)N(C)[c]2:n:[cH]:n([*]):[c]:2C1=[*]	-0.496	0 out of 2
SCFP_6	-2121589288	 [*]CN1C(=[*])[*]:[c](:[*])N(C)C1=O	-0.496	0 out of 2

Sorafenib

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.293

Enrichment: 0.878

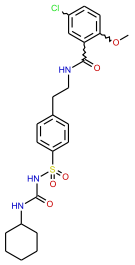
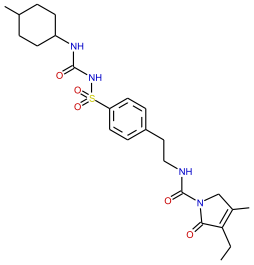
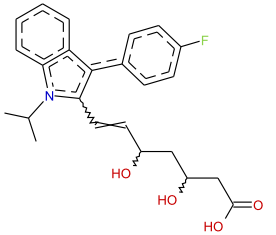
Bayesian Score: -2.4

Mahalanobis Distance: 17.6

Mahalanobis Distance p-value: 1.1e-012

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

Structural Similar Compounds

Name	Glyburide	Glimepride	Fluvastatin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.593	0.600	0.615
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

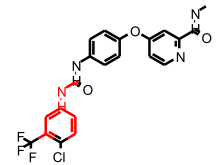
Model Applicability

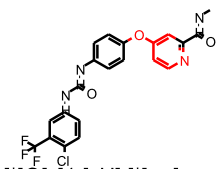
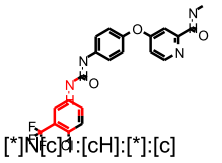
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

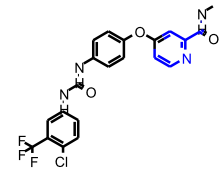
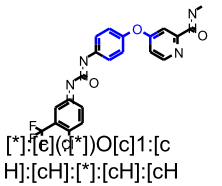
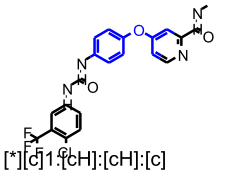
Feature Contribution

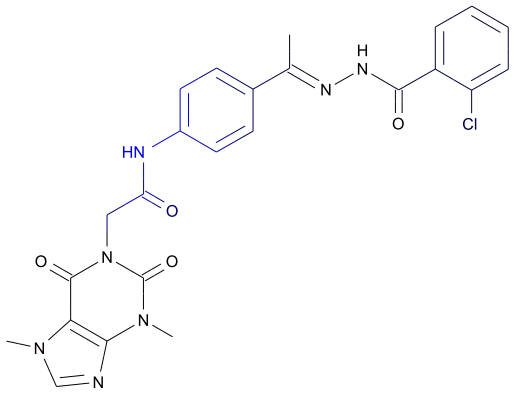
Top features for positive contribution

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

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

Top Features for negative contribution

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



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.0481

Enrichment: 0.0522

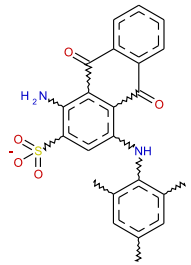
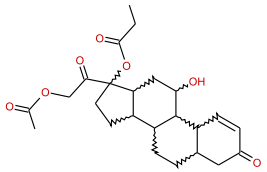
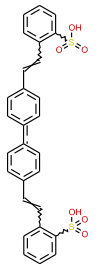
Bayesian Score: -6.29

Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 0.0067

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	0.719	0.787	0.838
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

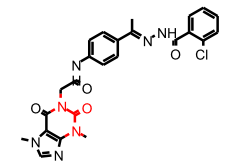
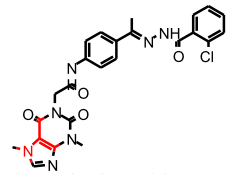
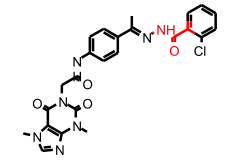
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

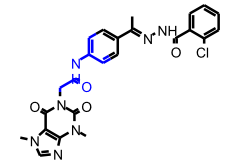
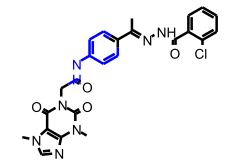
Feature Contribution

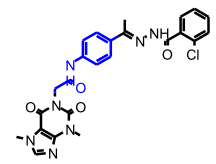
Top features for positive contribution

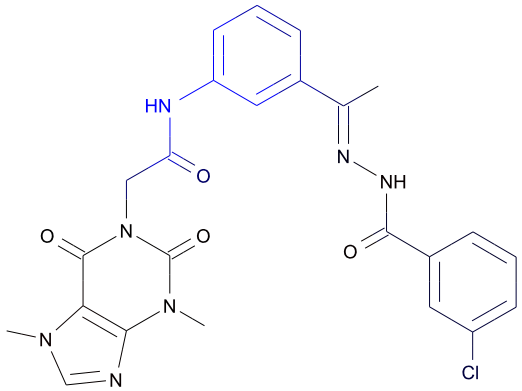
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]N([*])C(=O)N([*])[*]	0.0821	13 out of 13
FCFP_12	-1539132615	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	0.0795	9 out of 9
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.0734	5 out of 5

Top Features for negative contribution

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

FCFP_12	-451043714		-0.65	0 out of 1



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.00599

Enrichment: 0.0065

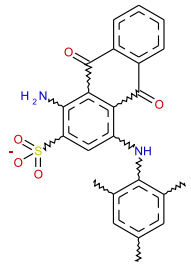
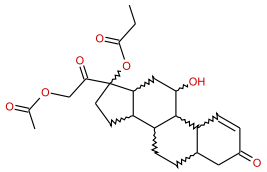
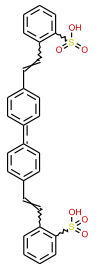
Bayesian Score: -7.3

Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 0.0067

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	0.719	0.787	0.841
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

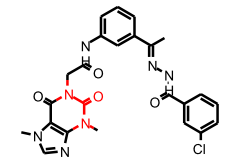
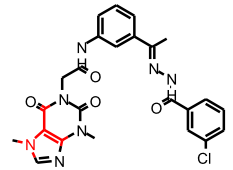
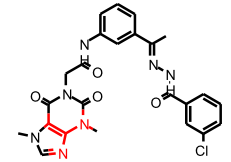
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

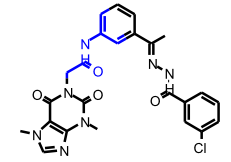
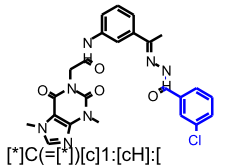
Feature Contribution

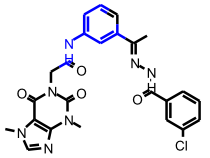
Top features for positive contribution

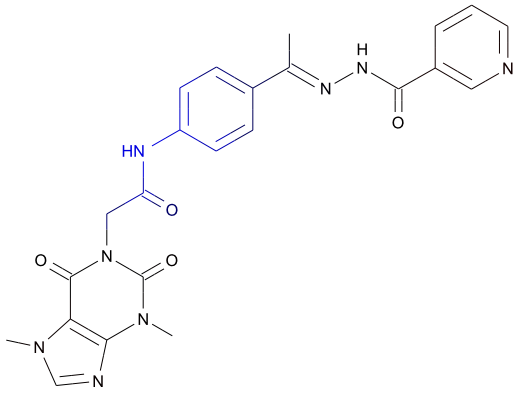
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]N([*])C(=O)N([*])[*]	0.0821	13 out of 13
FCFP_12	-1539132615	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	0.0795	9 out of 9
FCFP_12	-1410049896	 [*]N([*])[c]1:n:[*]:[*]:[c]: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	 [*]CC(=O)N[c](:[cH]:[*]):[cH]:[*]	-1.02	2 out of 8
FCFP_12	-1700637232	 [*]C(=[*])[c]1:[cH]:[cH]:[cH]:[c](Cl):[cH]:1	-0.846	1 out of 4

FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12



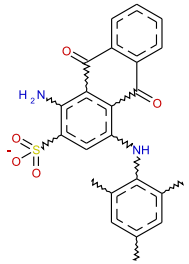
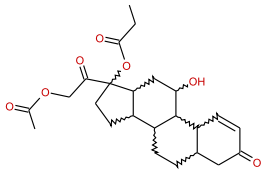
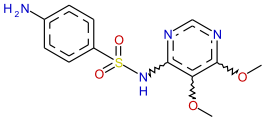
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.427
Enrichment: 0.464
Bayesian Score: -4.74
Mahalanobis Distance: 10.7
Mahalanobis Distance p-value: 0.00888

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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)-	Benzenesulfonamide, 4-amino-N-(5,6-dimethoxy-4-pyrimidinyl)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.719	0.825	0.837
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	FCTXAV 14,307,76

Model Applicability

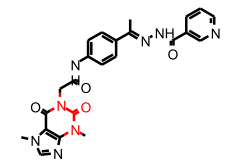
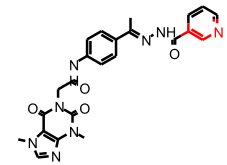
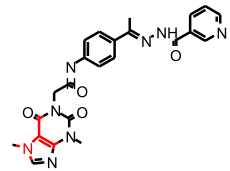
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

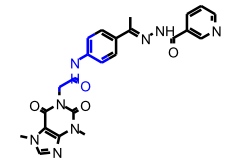
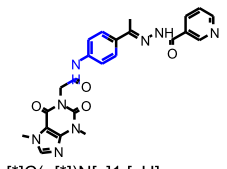
Feature Contribution

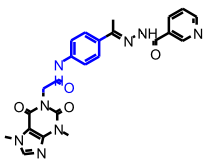
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

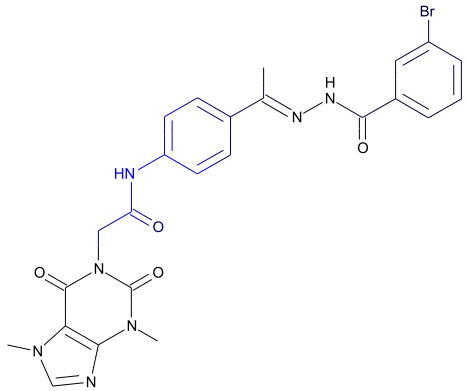
FCFP_12	-1986158408	 [*]N([*])C(=O)N([*])[*]	0.0821	13 out of 13
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1539132615	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	0.0795	9 out of 9

Top Features for negative contribution

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

FCFP_12	-451043714	 [*]CC(=O)N[c]1:[cH]:[cH]:[c]([*]):[cH]:[cH]:1	-0.65	0 out of 1

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.0623

Enrichment: 0.0676

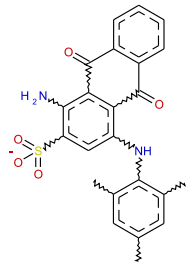
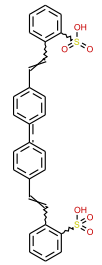
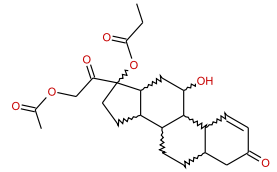
Bayesian Score: -6.15

Mahalanobis Distance: 11.1

Mahalanobis Distance p-value: 0.00167

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	0.788	0.840	0.844
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

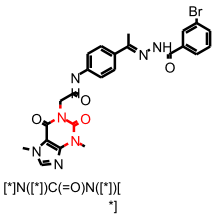
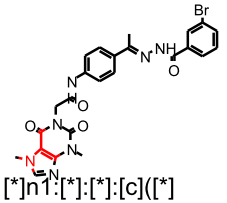
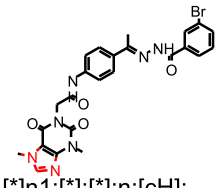
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

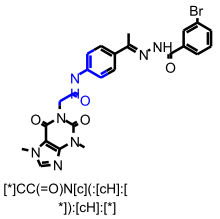
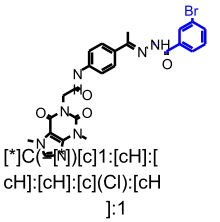
Feature Contribution

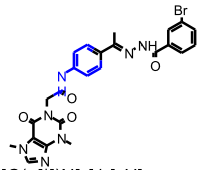
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408		0.0821	13 out of 13
FCFP_12	-1539132615		0.0795	9 out of 9
FCFP_12	-124685461		0.0734	5 out of 5

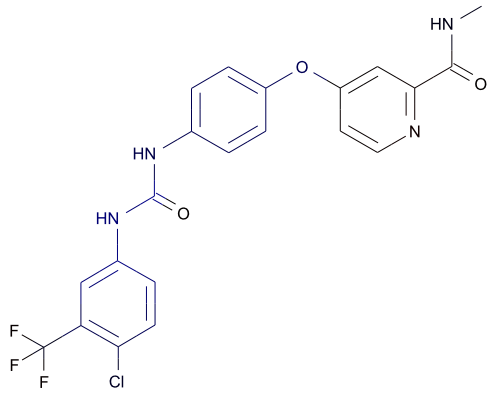
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1175665944		-1.02	2 out of 8
FCFP_12	-1700637232		-0.846	1 out of 4

FCFP_12	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.692	5 out of 12

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

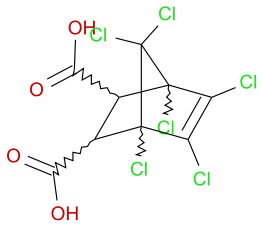
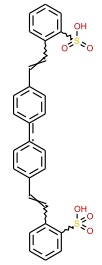
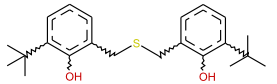
Bayesian Score: -5.23

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

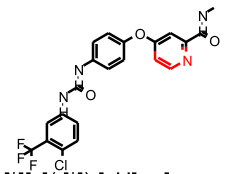
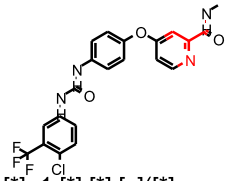
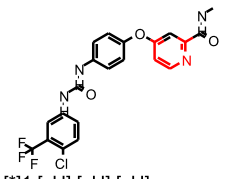
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

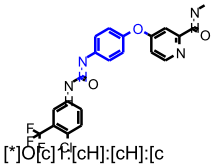
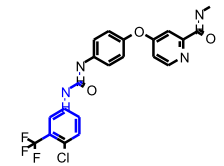
Feature Contribution

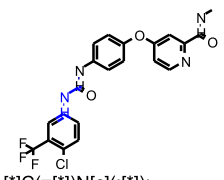
Top features for positive contribution

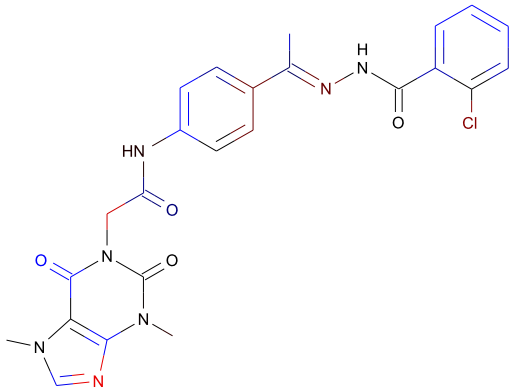
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1539132615	 [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]	0.0795	9 out of 9
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[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

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



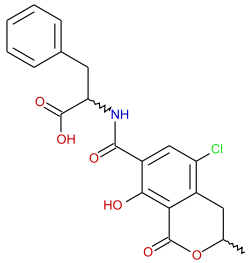
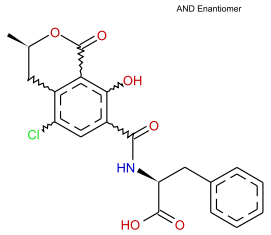
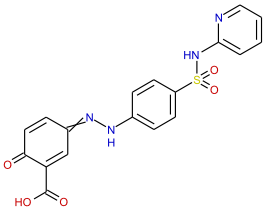
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 18.3
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.6
Mahalanobis Distance p-value: 8.91e-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)	4.79932	4.79932	2.5034
Predicted Endpoint (-log C)	3.6353	3.6353	3.54214
Distance	0.725	0.725	0.762
Reference	CPDB	CPDB	CPDB

Model Applicability

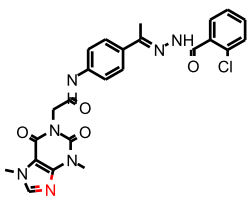
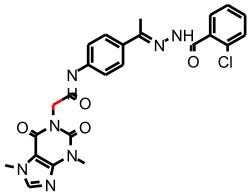
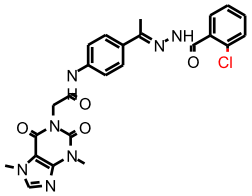
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -960717516: [*]C(=[*])N(C)[c](:[*]):[*]
3. Unknown ECFP_2 feature: -661097313: [*]CN(C(=[*])[*])C(=[*])[*]
4. Unknown ECFP_2 feature: 1135573248: [*]N([*])C(=O)N([*])[*]
5. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]
6. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

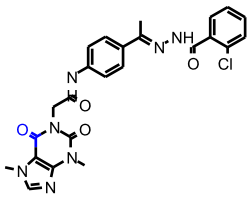
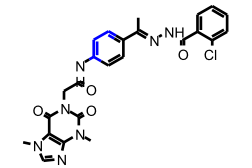
Feature Contribution

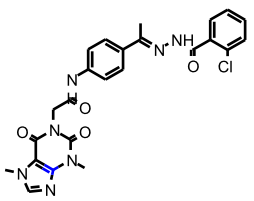
Top features for positive contribution

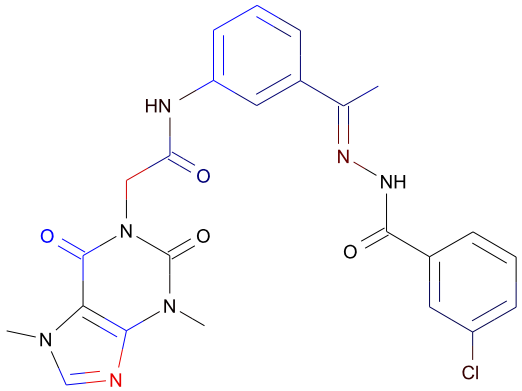
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]:n:[*]	0.229
ECFP_6	1559650422	 [*]C[*]	0.203
ECFP_6	-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



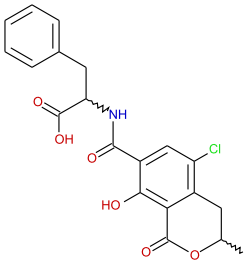
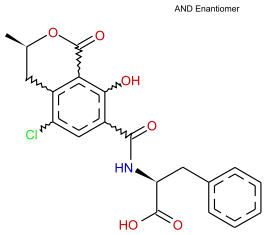
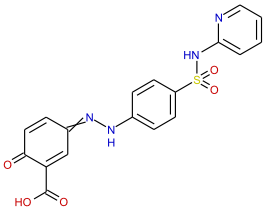
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 24.5
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15.3
Mahalanobis Distance p-value: 7.54e-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)	4.79932	4.79932	2.5034
Predicted Endpoint (-log C)	3.6353	3.6353	3.54214
Distance	0.734	0.734	0.770
Reference	CPDB	CPDB	CPDB

Model Applicability

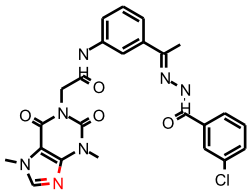
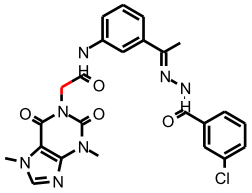
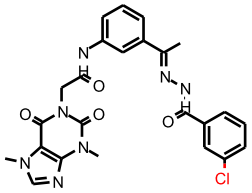
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC20 out of range. Value: 3.7212. Training min, max, SD, explained variance: -4.3384, 3.4394, 1.14, 0.0162.
2. Unknown ECFP_2 feature: -960717516: [*]C(=[*])N(C)[c](:[*]):[*]
3. Unknown ECFP_2 feature: -661097313: [*]CN(C(=[*])[*])C(=[*])[*]
4. Unknown ECFP_2 feature: 1135573248: [*]N([*])C(=O)N([*])[*]
5. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]
6. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

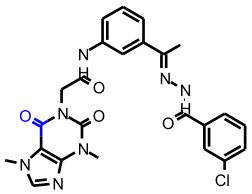
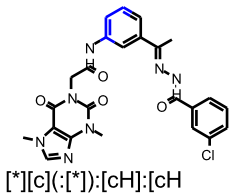
Feature Contribution

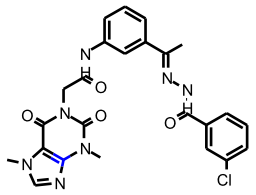
Top features for positive contribution

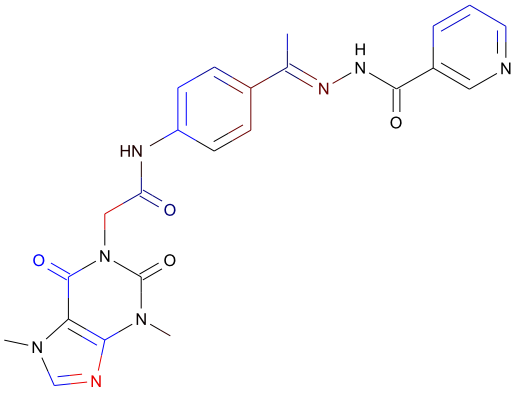
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]:n:[*]	0.229
ECFP_6	1559650422	 [*]C[*]	0.203
ECFP_6	-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



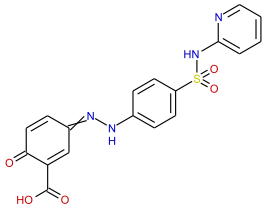
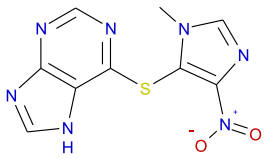
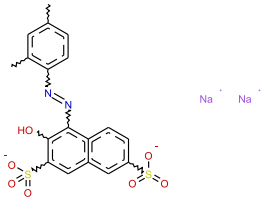
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: 27
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 1.77e-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	Salicylazosulfapyridine	Azathioprine s	420
Structure			
Actual Endpoint (-log C)	2.5034	4.49253	2.78302
Predicted Endpoint (-log C)	3.54214	4.28607	3.31546
Distance	0.738	0.785	0.790
Reference	CPDB	CPDB	CPDB

Model Applicability

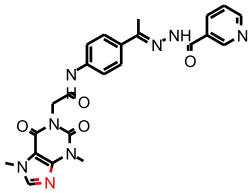
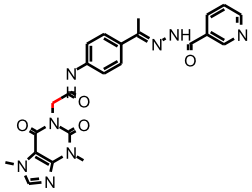
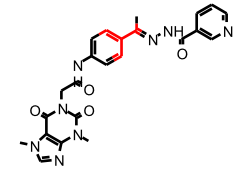
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -960717516: [*]C(=[*])N(C)[c](:[*]):[*]
3. Unknown ECFP_2 feature: -661097313: [*]CN(C(=[*])[*])C(=[*])[*]
4. Unknown ECFP_2 feature: 1135573248: [*]N([*])C(=O)N([*])[*]
5. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]
6. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

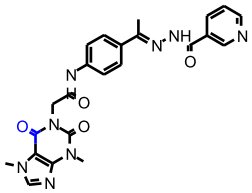
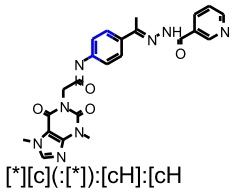
Feature Contribution

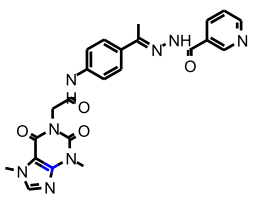
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

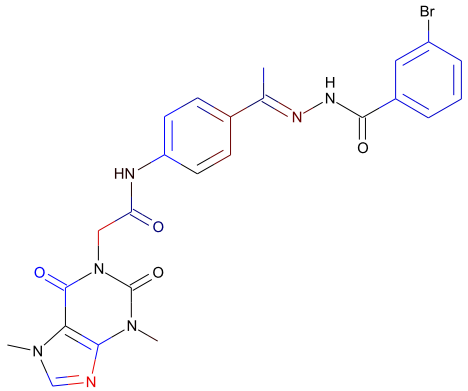
ECFP_6	655739385	 [*]:n:[*]	0.229
ECFP_6	1559650422	 [*]C[*]	0.203
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 24

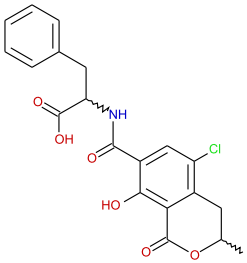
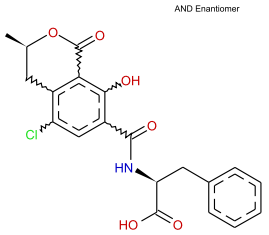
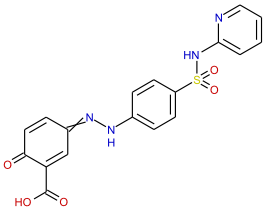
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 14.3

Mahalanobis Distance p-value: 5.61e-011

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

Structural Similar Compounds

Name	Ochratoxin A	542	Salicylazosulfapyridine
Structure			
Actual Endpoint (-log C)	4.79932	4.79932	2.5034
Predicted Endpoint (-log C)	3.6353	3.6353	3.54214
Distance	0.765	0.765	0.791
Reference	CPDB	CPDB	CPDB

Model Applicability

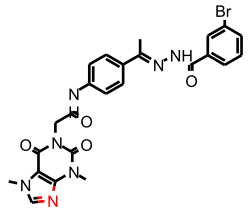
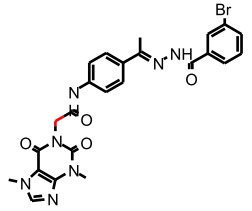
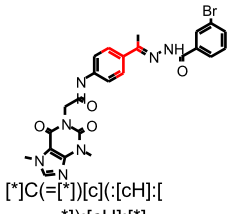
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -960717516: [*]C(=[*])N(C)[c](:[*]):[*]
3. Unknown ECFP_2 feature: -661097313: [*]CN(C(=[*])[*])C(=[*])[*]
4. Unknown ECFP_2 feature: 1135573248: [*]N([*])C(=O)N([*])[*]
5. Unknown ECFP_2 feature: -37698365: [*]N([*])CC(=[*])[*]
6. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown ECFP_2 feature: 560380707: [*]NN=C([*])[*]

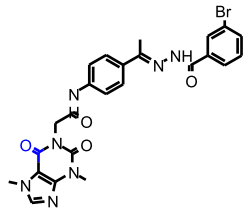
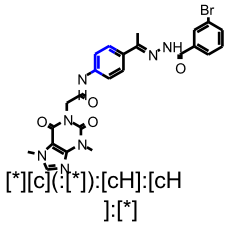
Feature Contribution

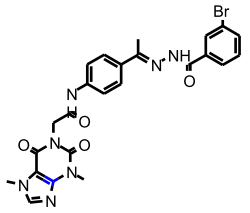
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

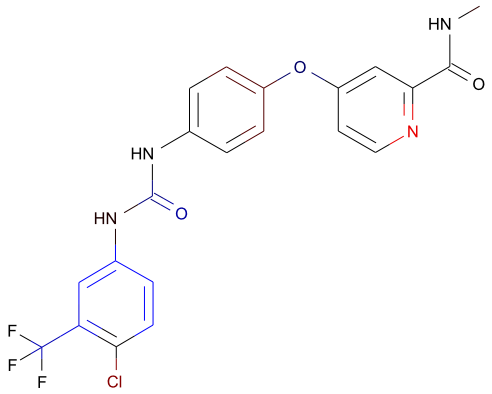
ECFP_6	655739385	 [*]:n:[*]	0.229
ECFP_6	1559650422	 [*]C[*]	0.203
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

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

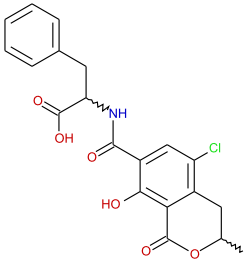
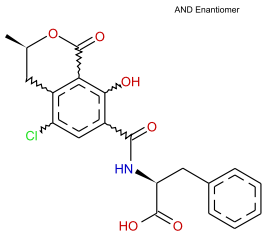
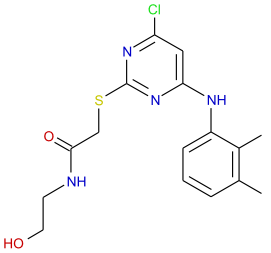
Mahalanobis Distance: 12.4

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

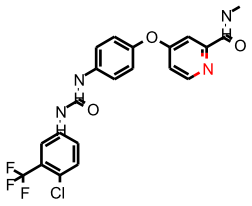
Model Applicability

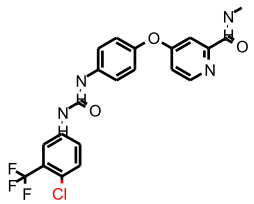
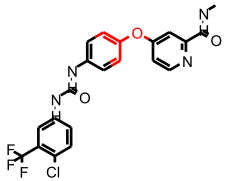
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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:[*]

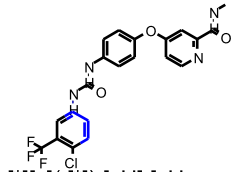
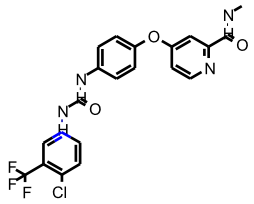
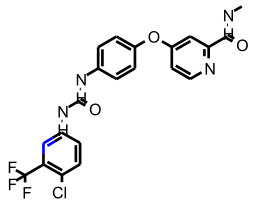
Feature Contribution

Top features for positive contribution

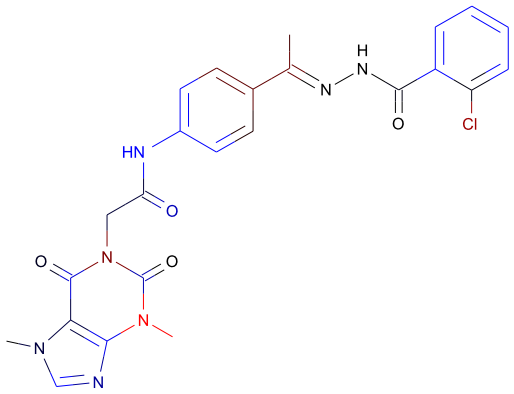
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

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

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 11.6

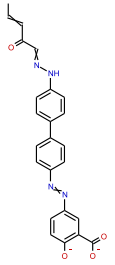
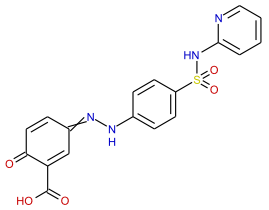
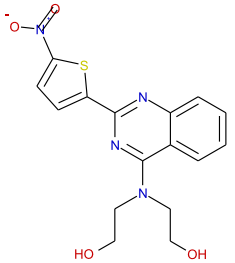
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 16.1

Mahalanobis Distance p-value: 1.12e-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	C.I. direct brown 95	Salicylazosulfapyridine	4-Bis(2-hydroxyethyl)amino-2-(5-nitro-2-thienyl)quinazoline
Structure			
Actual Endpoint (-log C)	5.31387	2.39891	5.05984
Predicted Endpoint (-log C)	4.30266	3.17598	4.23808
Distance	0.745	0.747	0.761
Reference	CPDB	CPDB	CPDB

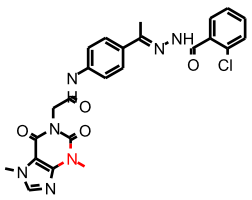
Model Applicability

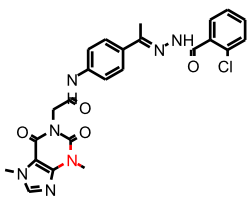
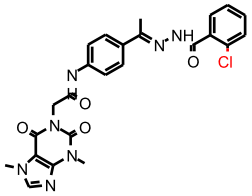
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

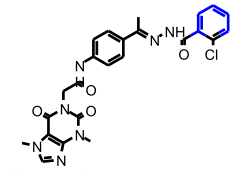
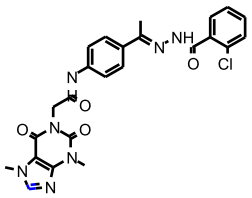
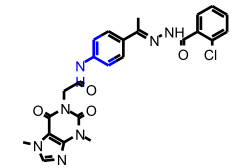
Feature Contribution

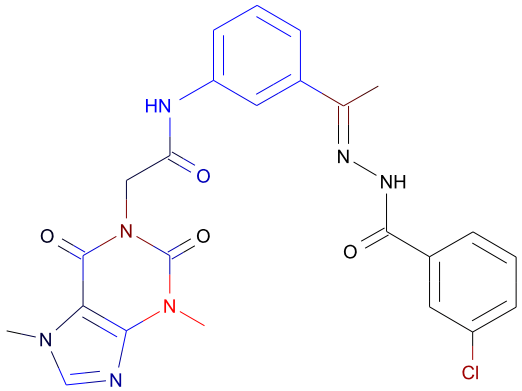
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]N([*])C	0.69

FCFP_6	1	 [*]N([*])[*]	0.234
FCFP_6	32	 [*]Cl	0.154

Top Features for negative contribution

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



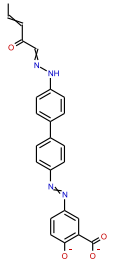
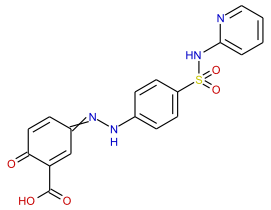
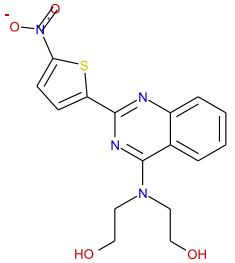
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 11.1
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 16.4
Mahalanobis Distance p-value: 1.12e-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	C.I. direct brown 95	Salicylazosulfapyridine	4-Bis(2-hydroxyethyl)amino-2-(5-nitro-2-thienyl)quinazoline
Structure			
Actual Endpoint (-log C)	5.31387	2.39891	5.05984
Predicted Endpoint (-log C)	4.30266	3.17598	4.23808
Distance	0.747	0.748	0.762
Reference	CPDB	CPDB	CPDB

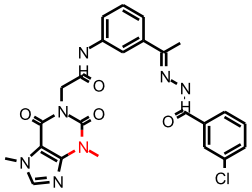
Model Applicability

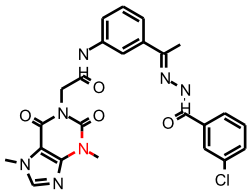
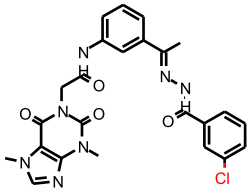
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

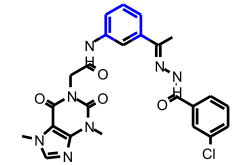
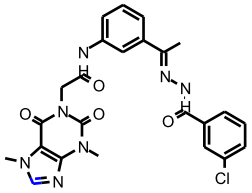
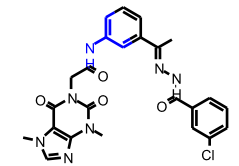
Feature Contribution

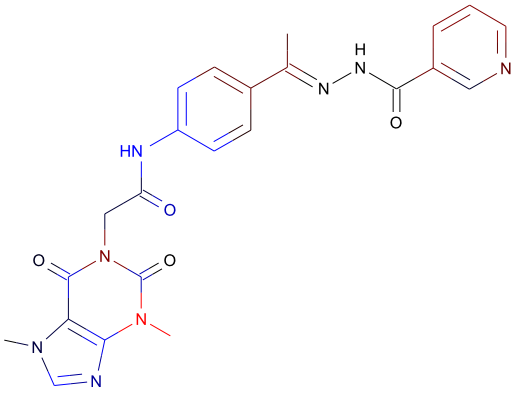
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]N([*])C	0.69

FCFP_6	1	 [*]N([*])[*]	0.234
FCFP_6	32	 [*]Cl	0.154

Top Features for negative contribution

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



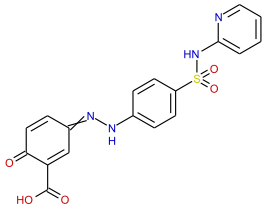
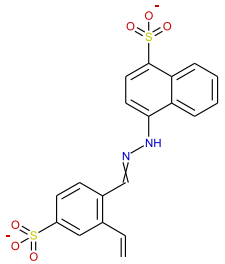
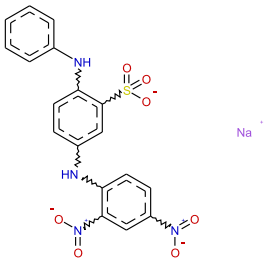
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: 6.24
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 16.4
Mahalanobis Distance p-value: 1.05e-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	Salicylazosulfapyridine	FD & C red no. 2	623
Structure			
Actual Endpoint (-log C)	2.39891	2.46661	2.39985
Predicted Endpoint (-log C)	3.17598	3.60201	3.4177
Distance	0.738	0.742	0.748
Reference	CPDB	CPDB	CPDB

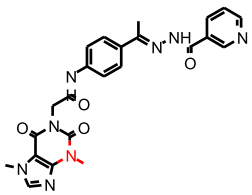
Model Applicability

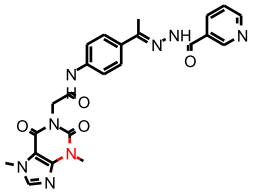
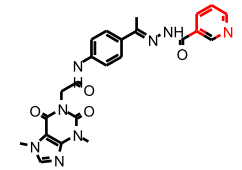
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC25 out of range. Value: 4.4122. Training min, max, SD, explained variance: -3.5991, 4.3957, 1.055, 0.0107.

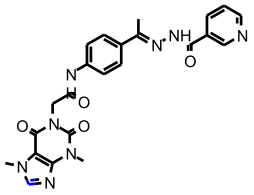
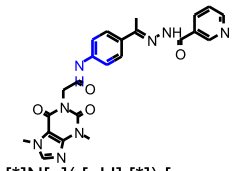
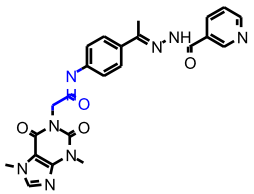
Feature Contribution

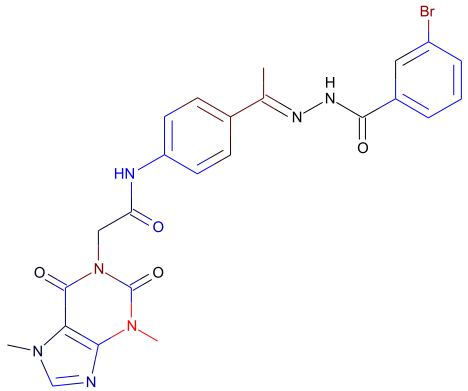
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]N([*])C	0.69

FCFP_6	1	 [*]N([*])[*]	0.234
FCFP_6	730557100	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	0.141

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.323
FCFP_6	566058135	 [*]CC(=O)N[*]	-0.182



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 11.3

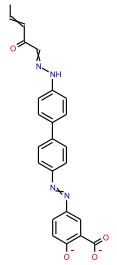
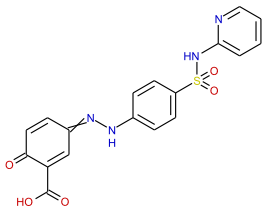
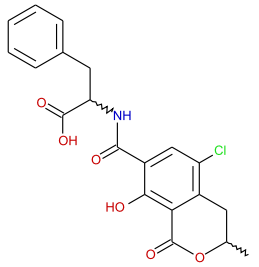
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 16.4

Mahalanobis Distance p-value: 9.63e-016

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

Structural Similar Compounds

Name	C.I. direct brown 95	Salicylazosulfapyridine	Ochratoxin A
Structure			
Actual Endpoint (-log C)	5.31387	2.39891	6.47264
Predicted Endpoint (-log C)	4.30266	3.17598	5.06501
Distance	0.760	0.766	0.782
Reference	CPDB	CPDB	CPDB

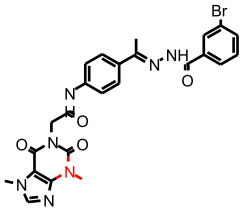
Model Applicability

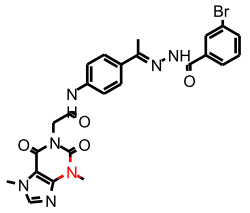
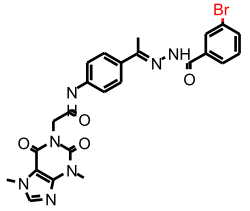
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

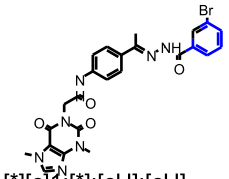
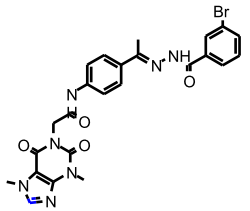
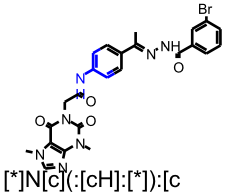
Feature Contribution

Top features for positive contribution

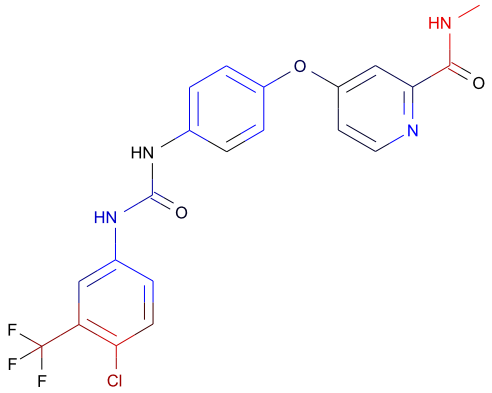
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	136627117	 [*]N([*])C	0.69

FCFP_6	1	 [*]N([*])[*]	0.234
FCFP_6	32	 [*]Cl	0.154

Top Features for negative contribution

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

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

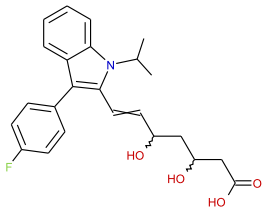
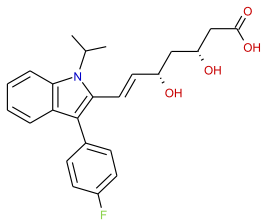
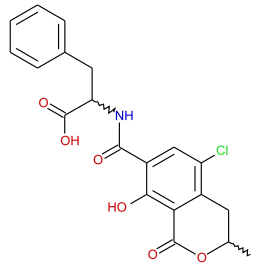
Mahalanobis Distance: 20.4

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

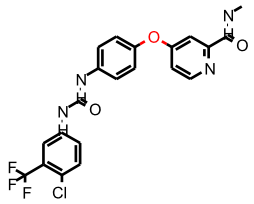
Model Applicability

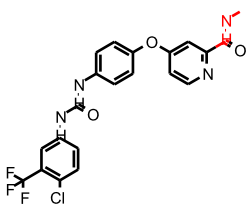
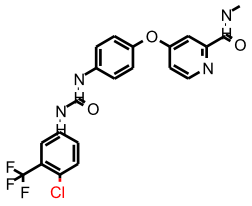
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

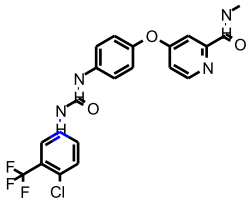
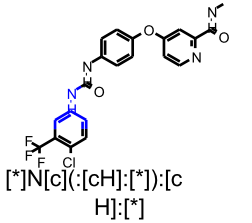
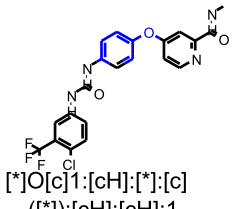
Feature Contribution

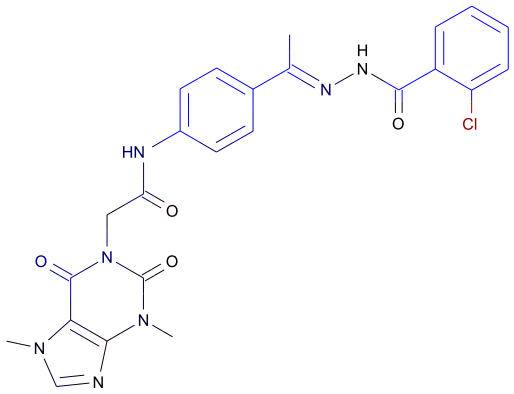
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]N([*])[*]	0.234

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

Top Features for negative contribution

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



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.108

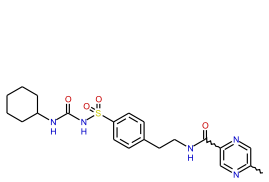
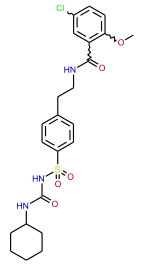
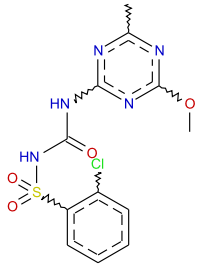
Unit: g/kg_body_weight

Mahalanobis Distance: 31.5

Mahalanobis Distance p-value: 2.64e-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	GLYBURIDE	CHLORSULFURON
Structure			
Actual Endpoint (-log C)	3.94991	4.21661	4.15566
Predicted Endpoint (-log C)	3.95594	4.21035	3.79771
Distance	0.650	0.693	0.747
Reference	NDA-17583	UPJ-26452	EPA COVER SHEET 0027;880301;(1)

Model Applicability

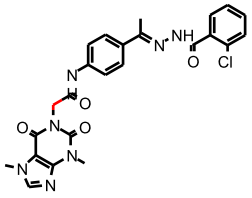
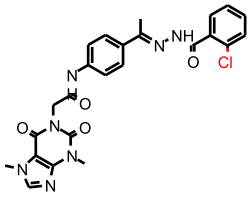
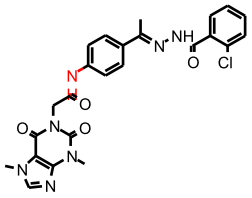
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 672362763: [*]n(:[*]):[*]
3. Unknown ECFP_6 feature: -677309799: [*][c]1:[*]:[*]:[cH]:n:1
4. Unknown ECFP_6 feature: -708878603: [*]n1:[*]:[*]:n:[cH]:1
5. Unknown ECFP_6 feature: -407983022: [*][c]1:[*]:[*]:[cH]:n:1C
6. Unknown ECFP_6 feature: -960717516: [*]C(=[*])N(C)[c](:[*]):[*]
7. Unknown ECFP_6 feature: -509950643: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown ECFP_6 feature: -813242890: [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]
9. Unknown ECFP_6 feature: 1945129186: [*]N([*])C(=O)[c](:[*]):[*]
10. Unknown ECFP_6 feature: -661097313: [*]CN(C(=[*])[*])C(=[*])[*]
11. Unknown ECFP_6 feature: 1135573248: [*]N([*])C(=O)N([*])[*]
12. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
13. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
14. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: 866343404: [*]N([*])C
16. Unknown ECFP_6 feature: 866450950: [*]:n(:[*])C
17. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
18. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
19. Unknown ECFP_6 feature: 560380707: [*]NN=C([*])[*]

20. Unknown ECFP_6 feature: 544048674: [*]C(=[*])NN=[*]
21. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
22. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
23. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

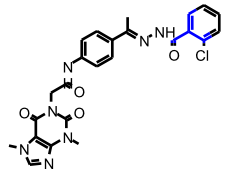
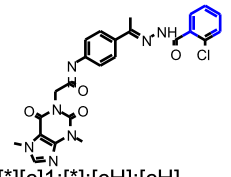
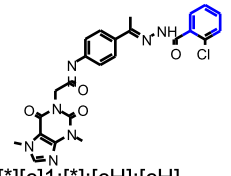
Feature Contribution

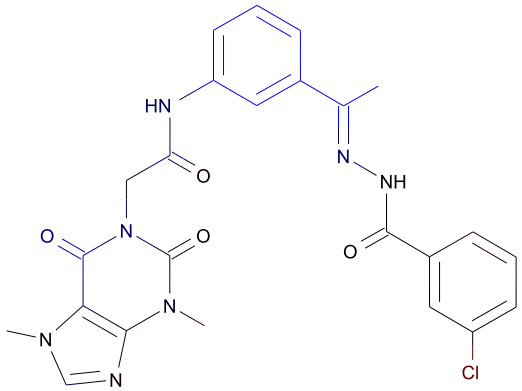
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.129
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	1337040050	 [*]C(=[*])[c](:[cH]:[*]):[c]([*]):[*]	-0.158
FCFP_6	991735244	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.134
ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.133



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.0464

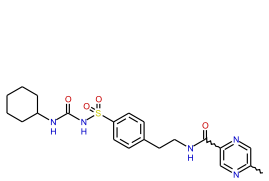
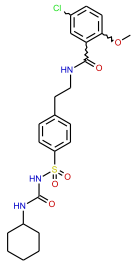
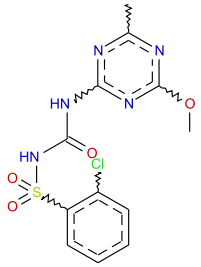
Unit: g/kg_body_weight

Mahalanobis Distance: 31.3

Mahalanobis Distance p-value: 6.01e-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	GLYBURIDE	CHLORSULFURON
Structure			
Actual Endpoint (-log C)	3.94991	4.21661	4.15566
Predicted Endpoint (-log C)	3.95594	4.21035	3.79771
Distance	0.651	0.689	0.760
Reference	NDA-17583	UPJ-26452	EPA COVER SHEET 0027;880301;(1)

Model Applicability

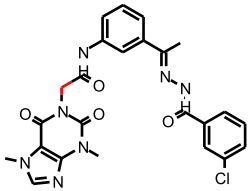
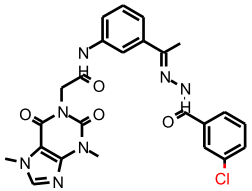
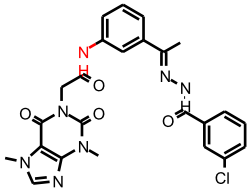
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 672362763: [*]n(:[*]):[*]
3. Unknown ECFP_6 feature: -677309799: [*][c]1:[*]:[*]:[cH]:n:1
4. Unknown ECFP_6 feature: -708878603: [*]n1:[*]:[*]:n:[cH]:1
5. Unknown ECFP_6 feature: -407983022: [*][c]1:[*]:[*]:[cH]:n:1C
6. Unknown ECFP_6 feature: -960717516: [*]C(=[*])N(C)[c](:[*]):[*]
7. Unknown ECFP_6 feature: -509950643: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown ECFP_6 feature: -813242890: [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]
9. Unknown ECFP_6 feature: 1945129186: [*]N([*])C(=O)[c](:[*]):[*]
10. Unknown ECFP_6 feature: -661097313: [*]CN(C(=[*])[*])C(=[*])[*]
11. Unknown ECFP_6 feature: 1135573248: [*]N([*])C(=O)N([*])[*]
12. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
13. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
14. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
16. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
17. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
18. Unknown ECFP_6 feature: 560380707: [*]NN=C([*])[*]
19. Unknown ECFP_6 feature: 544048674: [*]C(=[*])NN=[*]

20. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
21. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
22. Unknown ECFP_6 feature: 866343404: [*]N([*])C
23. Unknown ECFP_6 feature: 866450950: [*]:n(:[*])C
24. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

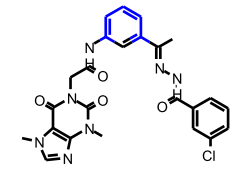
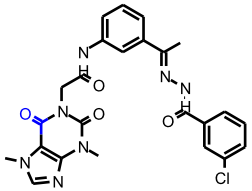
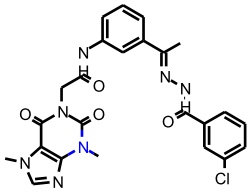
Feature Contribution

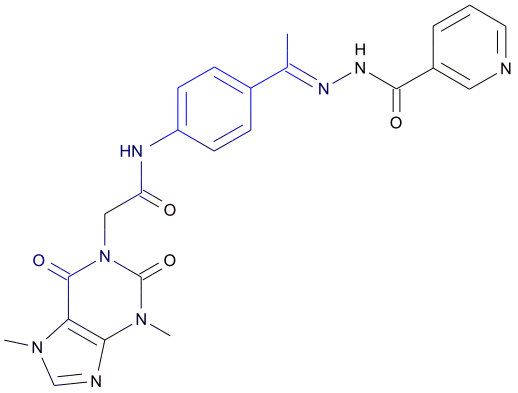
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.129
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

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



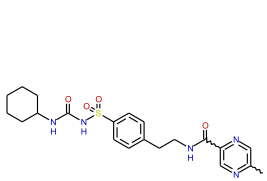
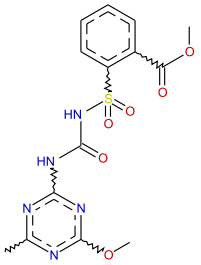
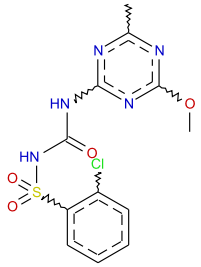
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: 0.0616
Unit: g/kg_body_weight
Mahalanobis Distance: 32.6
Mahalanobis Distance p-value: 2.26e-029

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

Structural Similar Compounds

Name	GLIPIZIDE	ALLY	CHLORSULFURON
Structure			
Actual Endpoint (-log C)	3.94991	3.1834	4.15566
Predicted Endpoint (-log C)	3.95594	3.59541	3.79771
Distance	0.663	0.707	0.762
Reference	NDA-17583	EPA COVER SHEET 0288;891101;(1)	EPA COVER SHEET 0027;880301;(1)

Model Applicability

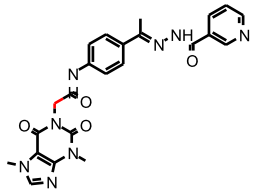
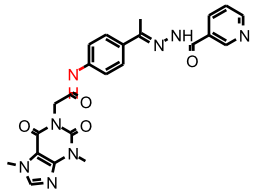
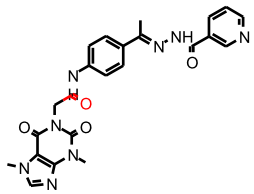
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 672362763: [*]n(:[*]):[*]
3. Unknown ECFP_6 feature: -677309799: [*][c]1:[*]:[*]:[cH]:n:1
4. Unknown ECFP_6 feature: -708878603: [*]n1:[*]:[*]:n:[cH]:1
5. Unknown ECFP_6 feature: -407983022: [*][c]1:[*]:[*]:[cH]:n:1C
6. Unknown ECFP_6 feature: -960717516: [*]C(=[*])N(C)[c](:[*]):[*]
7. Unknown ECFP_6 feature: -509950643: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown ECFP_6 feature: -813242890: [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]
9. Unknown ECFP_6 feature: 1945129186: [*]N([*])C(=O)[c](:[*]):[*]
10. Unknown ECFP_6 feature: -661097313: [*]CN(C(=[*])[*])C(=[*])[*]
11. Unknown ECFP_6 feature: 1135573248: [*]N([*])C(=O)N([*])[*]
12. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
13. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
14. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: 866343404: [*]N([*])C
16. Unknown ECFP_6 feature: 866450950: [*]:n(:[*])C
17. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
18. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
19. Unknown ECFP_6 feature: 560380707: [*]NN=C([*])[*]

20. Unknown ECFP_6 feature: 544048674: [*]C(=[*])NN=[*]
21. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
22. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
23. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]
24. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]
25. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

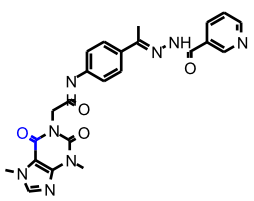
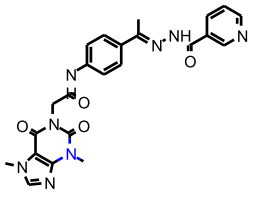
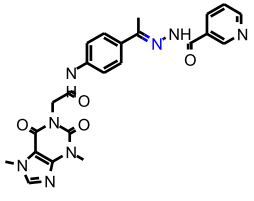
Feature Contribution

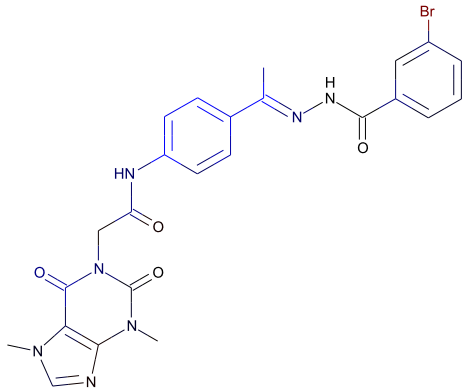
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.129
FCFP_6	3	 [*]N[*]	0.0924
ECFP_6	2099970318	 [*]C(=O)[*]	0.0766

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	2106656448	 [*]C(=O)[*]	-0.11
FCFP_6	1	 [*]N([*])[*]	-0.102
ECFP_6	-1087070950	 [*]N=[*]	-0.102



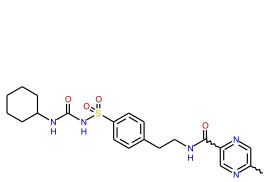
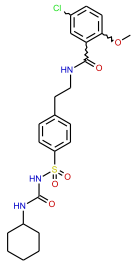
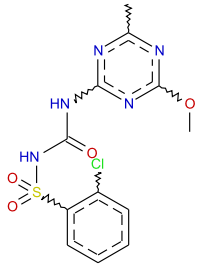
$C_{24}H_{22}BrN_7O_4$
Molecular Weight: 552.37997
ALogP: 2.118
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.0458
Unit: g/kg_body_weight
Mahalanobis Distance: 31.3
Mahalanobis Distance p-value: 5.34e-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	GLYBURIDE	CHLORSULFURON
Structure			
Actual Endpoint (-log C)	3.94991	4.21661	4.15566
Predicted Endpoint (-log C)	3.95594	4.21035	3.79771
Distance	0.675	0.701	0.798
Reference	NDA-17583	UPJ-26452	EPA COVER SHEET 0027;880301;(1)

Model Applicability

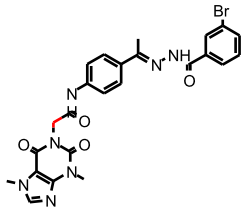
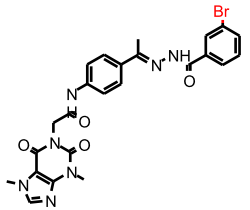
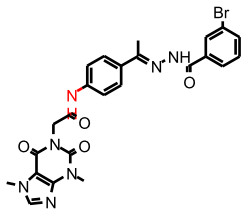
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 672362763: [*]n(:[*]):[*]
3. Unknown ECFP_6 feature: -302078100: [*]Br
4. Unknown ECFP_6 feature: -677309799: [*][c]1:[*]:[*]:[cH]:n:1
5. Unknown ECFP_6 feature: -708878603: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown ECFP_6 feature: -407983022: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown ECFP_6 feature: -960717516: [*]C(=[*])N(C)[c](:[*]):[*]
8. Unknown ECFP_6 feature: -509950643: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
9. Unknown ECFP_6 feature: -813242890: [*]n1:[*]:[*]:[c]([*]):[c]:1C(=[*])[*]
10. Unknown ECFP_6 feature: 1945129186: [*]N([*])C(=O)[c](:[*]):[*]
11. Unknown ECFP_6 feature: -661097313: [*]CN(C(=[*])[*])C(=[*])[*]
12. Unknown ECFP_6 feature: 1135573248: [*]N([*])C(=O)N([*])[*]
13. Unknown ECFP_6 feature: -37698365: [*]N([*])CC(=[*])[*]
14. Unknown ECFP_6 feature: 1731843802: [*]CC(=O)N[*]
15. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
16. Unknown ECFP_6 feature: 866343404: [*]N([*])C
17. Unknown ECFP_6 feature: 866450950: [*]:n(:[*])C
18. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
19. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]

20. Unknown ECFP_6 feature: 560380707: [*]NN=C([*])[*]
21. Unknown ECFP_6 feature: 544048674: [*]C(=[*])NN=[*]
22. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
23. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
24. Unknown ECFP_6 feature: -177935549: [*]:[cH]:[c](Br):[cH]:[*]
25. Unknown ECFP_6 feature: 459826767: [*]:[c](:[*])Br

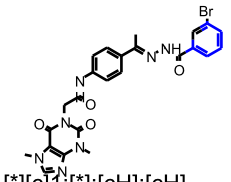
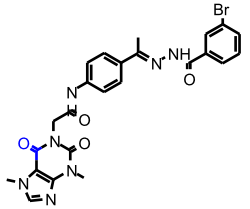
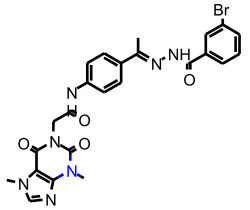
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	1559650422	 [*]C[*]	0.129
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

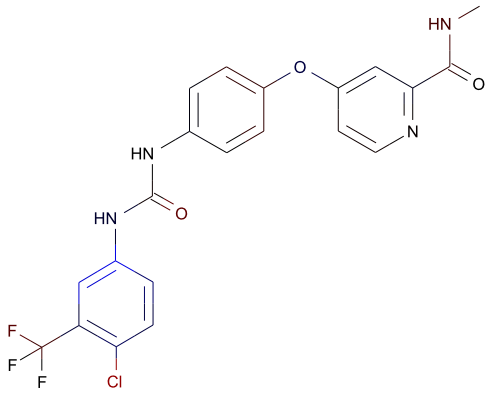
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

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

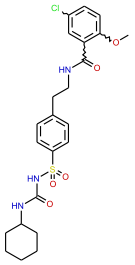
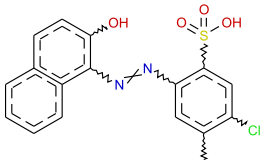
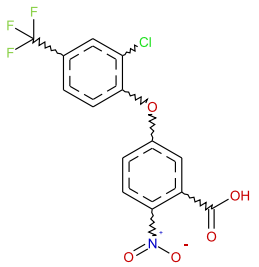
Unit: g/kg_body_weight

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)

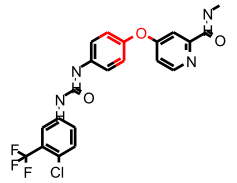
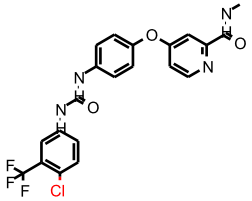
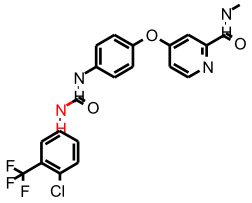
Model Applicability

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

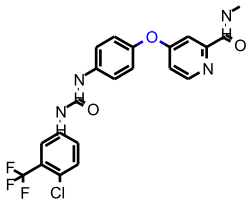
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -1046436026: [*]F
3. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
4. Unknown ECFP_6 feature: 226796801: [*]C([*])([*])F
5. Unknown ECFP_6 feature: 1305253718: [*]:[c](:[*])O[c](:[*]):[*]
6. Unknown ECFP_6 feature: -677309799: [*][c]1:[*]:[*]:[cH]:n:1
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

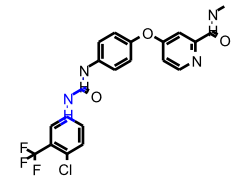
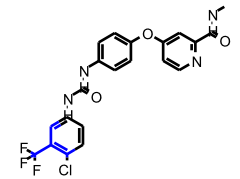
Feature Contribution

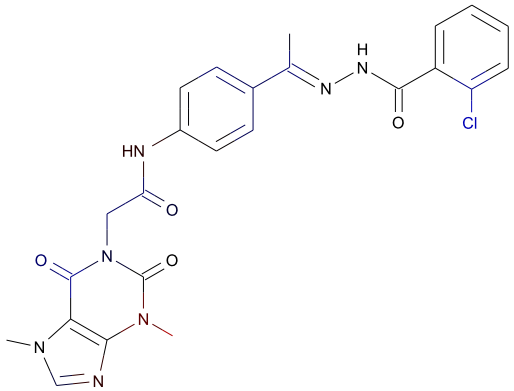
Top features for positive contribution

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	 [*]N([*])[*]	-0.102

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



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.0382

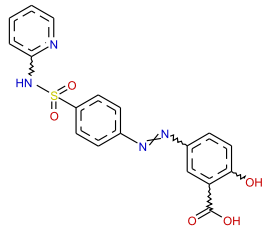
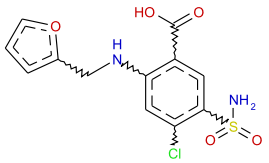
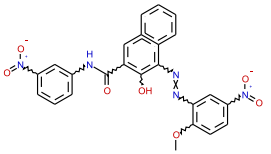
Unit: g/kg_body_weight

Mahalanobis Distance: 10.8

Mahalanobis Distance p-value: 2.21e-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	SALICYLAZOSULFAPYRIDINE	FUROSEMIDE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	3.375	4.04236	2.30052
Predicted Endpoint (-log C)	2.80292	2.8614	3.55333
Distance	0.726	0.729	0.896
Reference	NCI/NTP TR-457	NCI/NTP TR-356	NCI/NTP TR-411

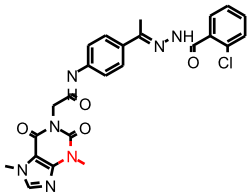
Model Applicability

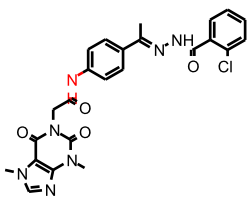
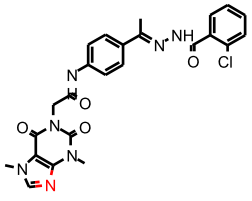
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
4. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

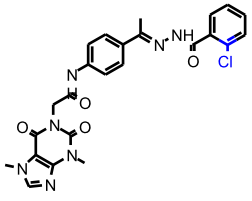
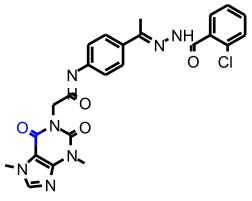
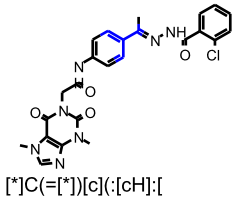
Feature Contribution

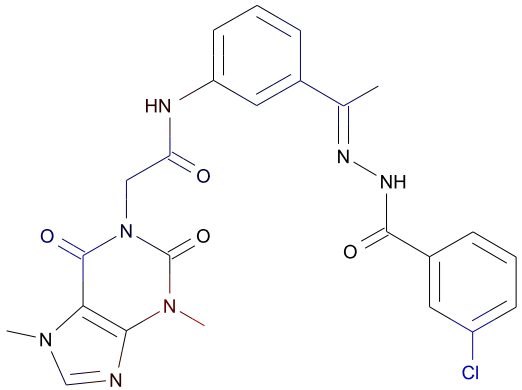
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]N([*])C	0.173

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	17	 [*]:n:[*]	0.0441

Top Features for negative contribution

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



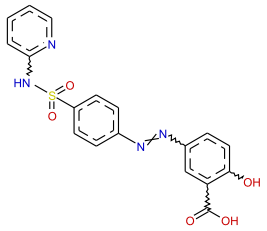
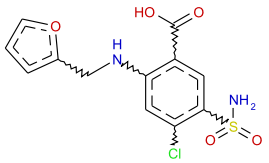
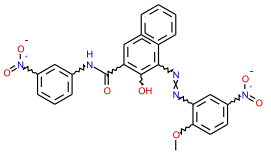
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.0382
Unit: g/kg_body_weight
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 2.21e-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	SALICYLAZOSULFAPYRIDINE	FUROSEMIDE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	3.375	4.04236	2.30052
Predicted Endpoint (-log C)	2.80292	2.8614	3.55333
Distance	0.726	0.729	0.896
Reference	NCI/NTP TR-457	NCI/NTP TR-356	NCI/NTP TR-411

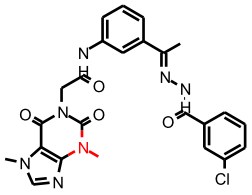
Model Applicability

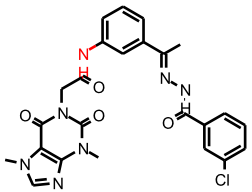
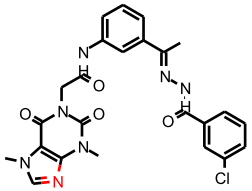
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
4. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

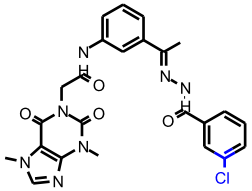
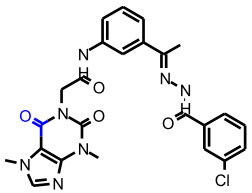
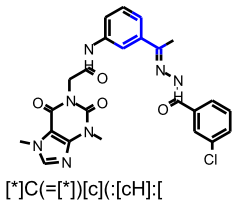
Feature Contribution

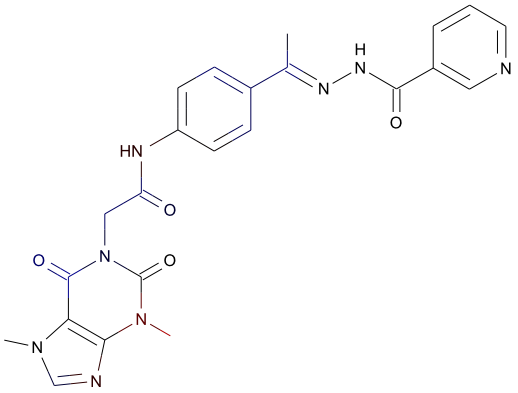
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]N([*])C	0.173

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	17	 [*]:n:[*]	0.0441

Top Features for negative contribution

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



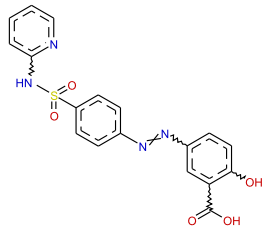
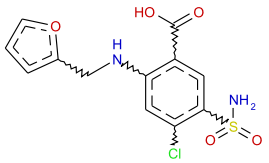
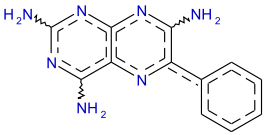
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: 0.0285
Unit: g/kg_body_weight
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 9.89e-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	SALICYLAZOSULFAPYRIDINE	FUROSEMIDE	TRIAMTERENE
Structure			
Actual Endpoint (-log C)	3.375	4.04236	4.00564
Predicted Endpoint (-log C)	2.80292	2.8614	3.1992
Distance	0.725	0.745	0.931
Reference	NCI/NTP TR-457	NCI/NTP TR-356	NCI/NTP TR-420

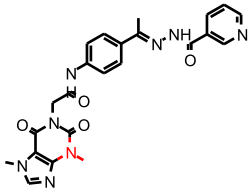
Model Applicability

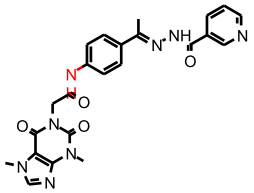
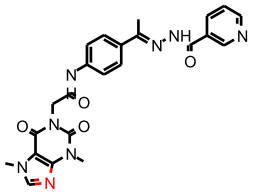
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
4. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

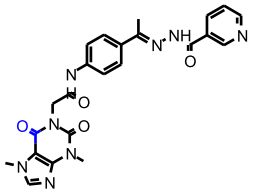
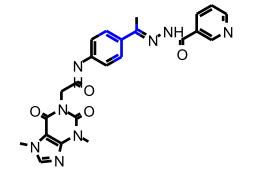
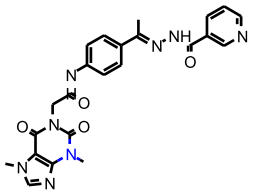
Feature Contribution

Top features for positive contribution

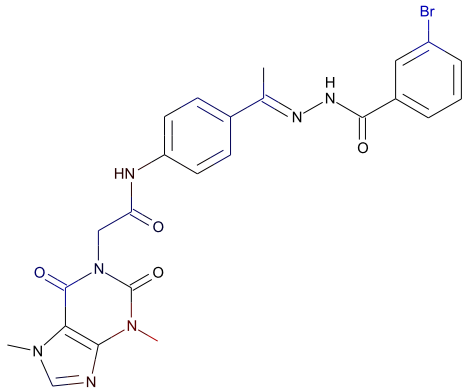
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]N([*])C	0.173

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	17	 [*]:n:[*]	0.0441

Top Features for negative contribution

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

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 0.0248

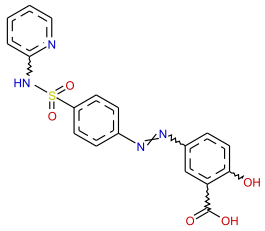
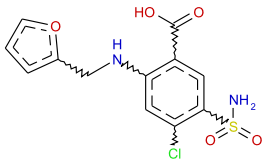
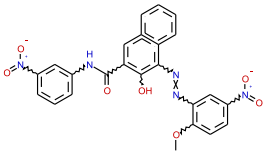
Unit: g/kg_body_weight

Mahalanobis Distance: 11.2

Mahalanobis Distance p-value: 4.39e-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	SALICYLAZOSULFAPYRIDINE	FUROSEMIDE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	3.375	4.04236	2.30052
Predicted Endpoint (-log C)	2.80292	2.8614	3.55333
Distance	0.767	0.784	0.901
Reference	NCI/NTP TR-457	NCI/NTP TR-356	NCI/NTP TR-411

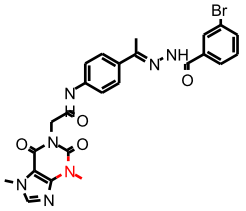
Model Applicability

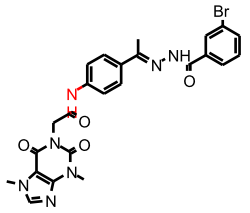
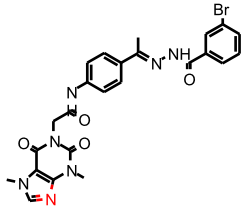
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
3. Unknown FCFP_2 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
4. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

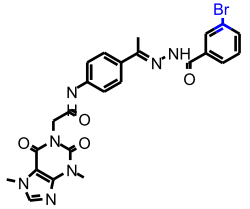
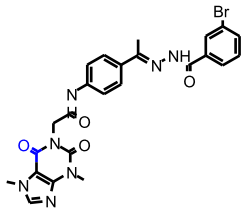
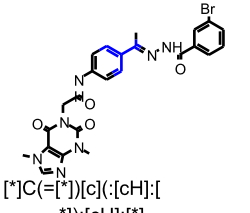
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136627117	 [*]N([*])C	0.173

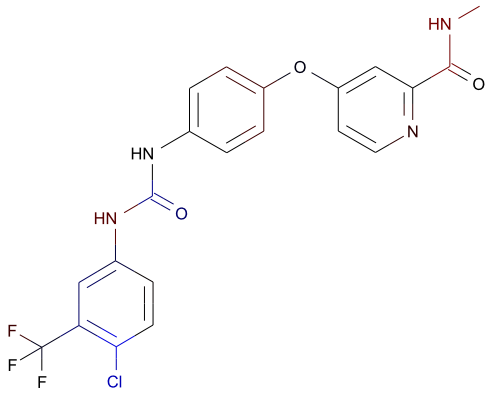
FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	17	 [*]:n:[*]	0.0441

Top Features for negative contribution

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

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.0885

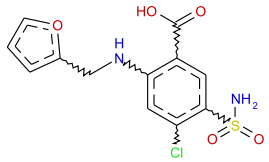
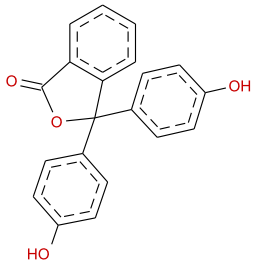
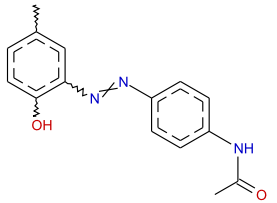
Unit: g/kg_body_weight

Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 1.76e-009

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

Structural Similar Compounds

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

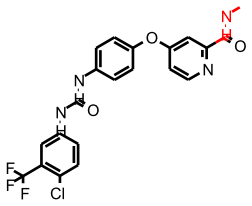
Model Applicability

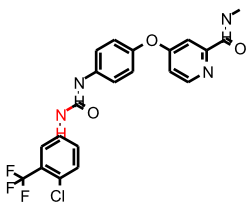
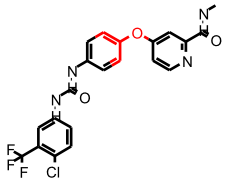
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

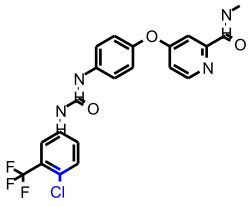
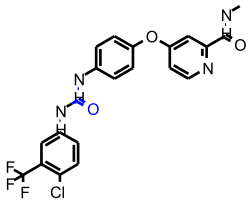
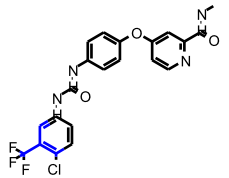
Feature Contribution

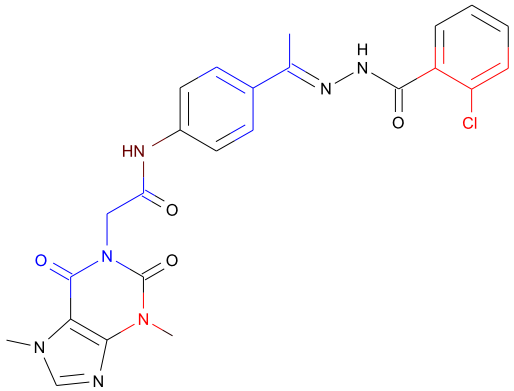
Top features for positive contribution

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

FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	332760439	 [*]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



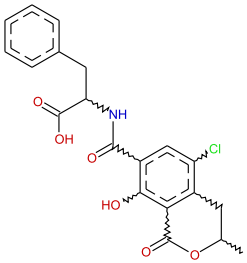
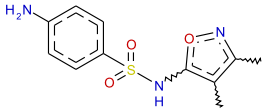
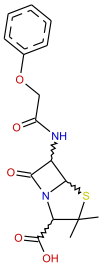
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.0156
Unit: g/kg_body_weight
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 1.47e-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	0.777	1.004	1.010
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

Model Applicability

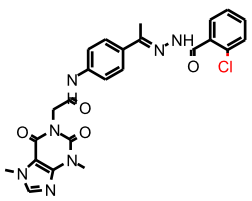
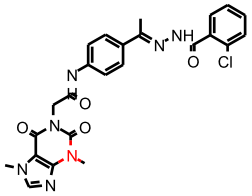
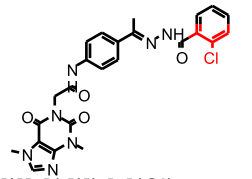
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 507.93. 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: -4.6919. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
4. OPS PC7 out of range. Value: -3.1347. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
5. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown FCFP_2 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown FCFP_2 feature: -1410049896: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown FCFP_2 feature: -1986158408: [*]N([*])C(=O)N([*])[*]
9. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
10. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
11. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]
12. Unknown FCFP_2 feature: -885520711: [*]C(=[*])NN=[*]

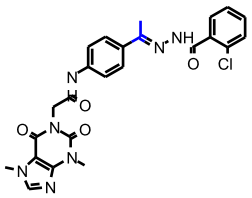
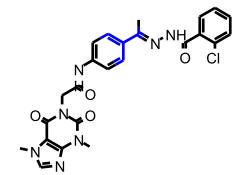
Feature Contribution

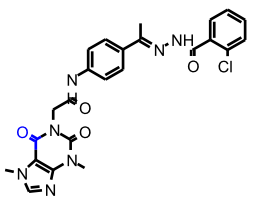
Top features for positive contribution

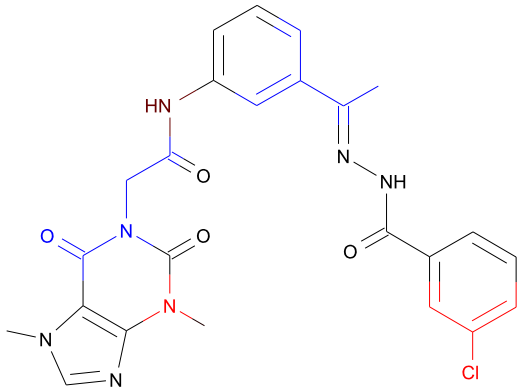
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

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



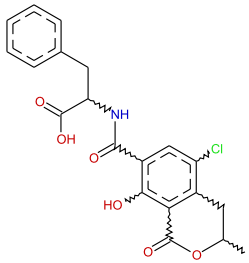
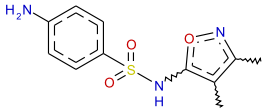
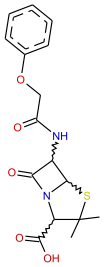
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.0156
Unit: g/kg_body_weight
Mahalanobis Distance: 12.5
Mahalanobis Distance p-value: 1.47e-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	0.777	1.004	1.010
Reference	NCI/NTP TR-358	NCI/NTP TR-138	NCI/NTP TR-336

Model Applicability

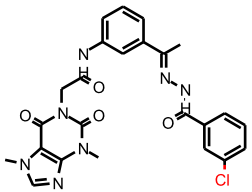
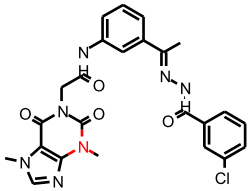
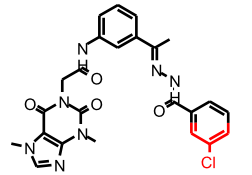
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 507.93. 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: -4.6919. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
4. OPS PC7 out of range. Value: -3.1347. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
5. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown FCFP_2 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown FCFP_2 feature: -1410049896: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown FCFP_2 feature: -1986158408: [*]N([*])C(=O)N([*])[*]
9. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
10. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]
11. Unknown FCFP_2 feature: -885520711: [*]C(=[*])NN=[*]
12. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C

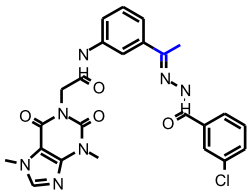
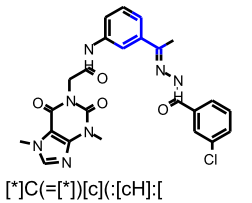
Feature Contribution

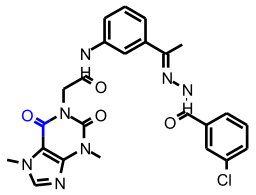
Top features for positive contribution

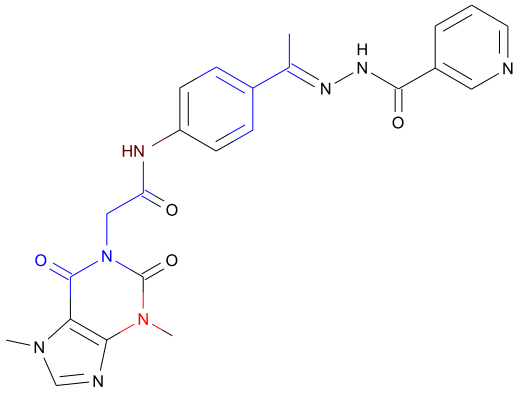
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

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



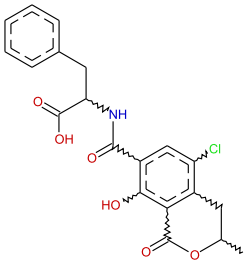
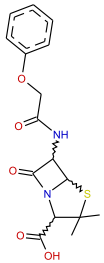
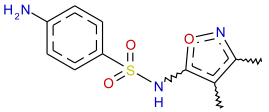
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: 0.113
Unit: g/kg_body_weight
Mahalanobis Distance: 13.2
Mahalanobis Distance p-value: 1.68e-010

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

Structural Similar Compounds

Name	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	0.916	0.985	1.002
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

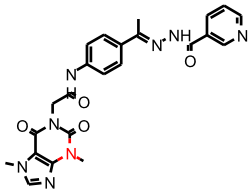
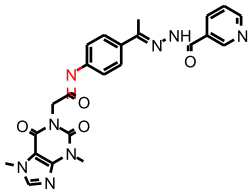
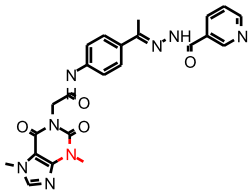
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 474.47. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 7. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. Molecular_PolarSurfaceArea out of range. Value: 141.88. Training min, max, mean, SD: 0, 138.03, 28.978, 32.1.
5. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown FCFP_2 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown FCFP_2 feature: -1410049896: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown FCFP_2 feature: -1986158408: [*]N([*])C(=O)N([*])[*]
9. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
10. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
11. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]
12. Unknown FCFP_2 feature: -885520711: [*]C(=[*])NN=[*]

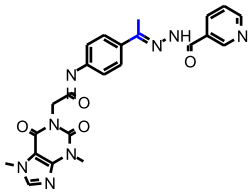
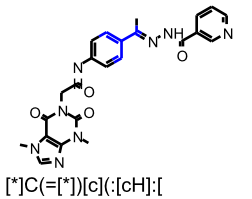
Feature Contribution

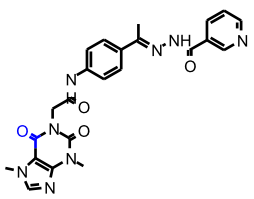
Top features for positive contribution

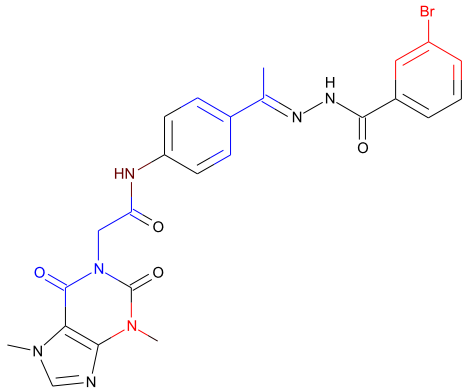
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	1	 [*]N([*])[*]	0.511
FCFP_2	3	 [*]N[*]	0.104
FCFP_2	136627117	 [*]N([*])C	0.0304

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

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



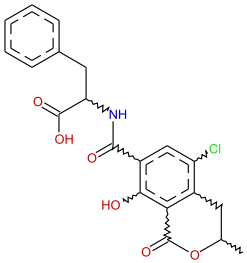
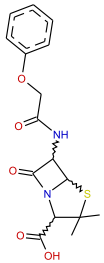
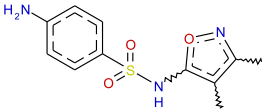
$C_{24}H_{22}BrN_7O_4$
Molecular Weight: 552.37997
ALogP: 2.118
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 0.0151
Unit: g/kg_body_weight
Mahalanobis Distance: 12.9
Mahalanobis Distance p-value: 5.05e-010

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

Structural Similar Compounds

Name	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	0.828	1.065	1.079
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

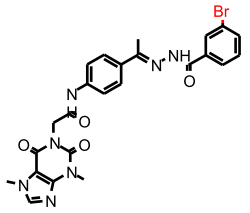
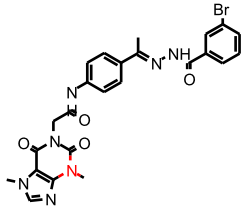
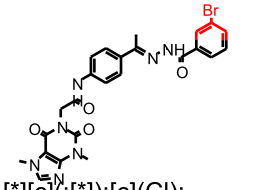
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 552.38. 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: -4.7836. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
4. OPS PC7 out of range. Value: -3.1984. Training min, max, SD, explained variance: -2.8003, 2.9332, 1.16, 0.0416.
5. Unknown FCFP_2 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown FCFP_2 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown FCFP_2 feature: -1410049896: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown FCFP_2 feature: -1986158408: [*]N([*])C(=O)N([*])[*]
9. Unknown FCFP_2 feature: 136150461: [*]:n(:[*])C
10. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
11. Unknown FCFP_2 feature: 581019816: [*]NN=C([*])[*]
12. Unknown FCFP_2 feature: -885520711: [*]C(=[*])NN=[*]

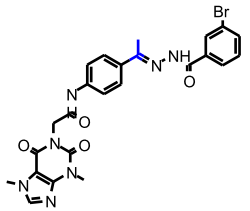
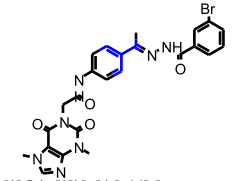
Feature Contribution

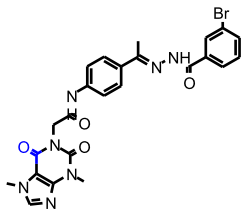
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

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

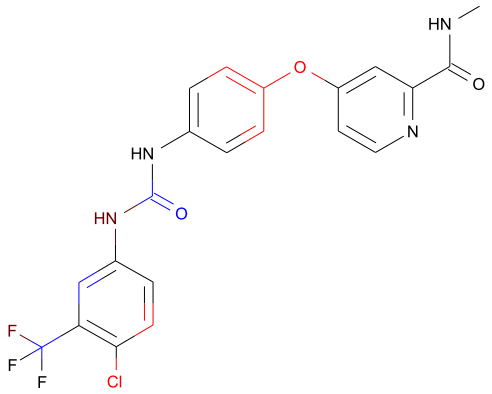
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489
FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406

FCFP_2	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

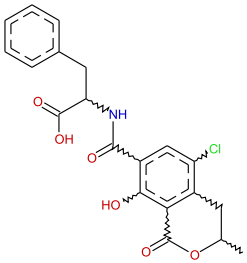
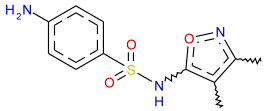
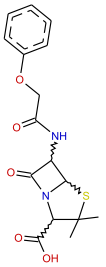
Unit: g/kg_body_weight

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

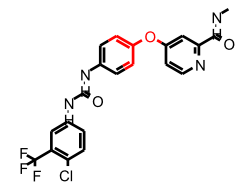
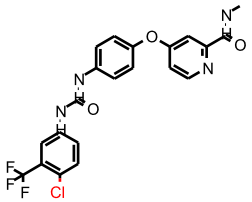
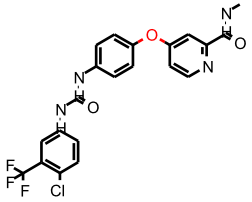
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

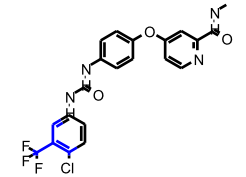
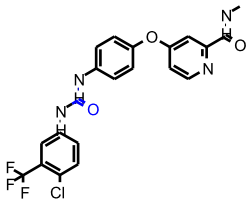
Feature Contribution

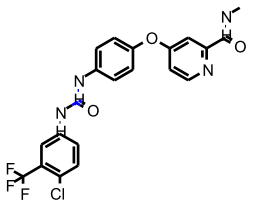
Top features for positive contribution

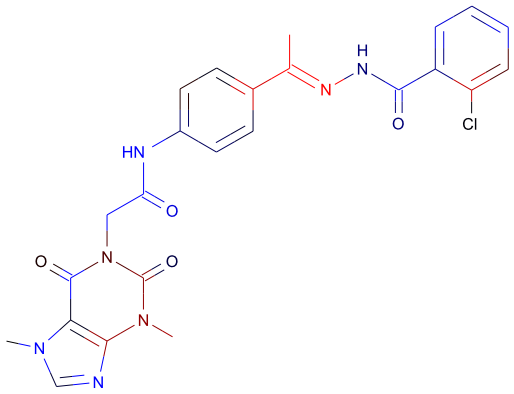
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

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

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



$C_{24}H_{22}ClN_7O_4$

Molecular Weight: 507.92898

ALogP: 2.034

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 1.58

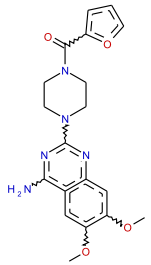
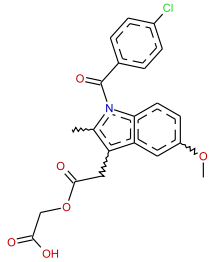
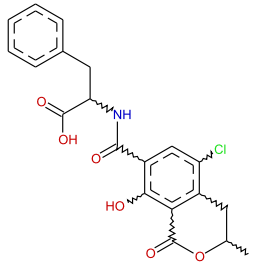
Unit: g/kg_body_weight

Mahalanobis Distance: 24.3

Mahalanobis Distance p-value: 5.05e-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	PRASOZIN .HCl (HCl STRIPPED)	ACEMETACIN	OCHRATOXIN A
Structure			
Actual Endpoint (-log C)	2.294	4.235	4.305
Predicted Endpoint (-log C)	3.00765	3.39415	3.03558
Distance	0.722	0.731	0.748
Reference	NIIRDN 6;688;82	ARZNAD 30;1398;80	FCTXAV 6;479;68

Model Applicability

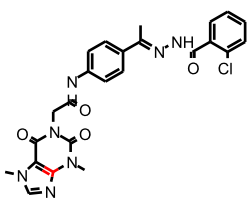
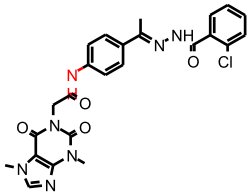
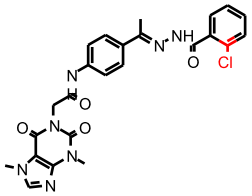
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[cH]:n:1
5. Unknown FCFP_6 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown FCFP_6 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown FCFP_6 feature: -1410049896: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown FCFP_6 feature: 136150461: [*]:n(:[*])C
9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
10. Unknown FCFP_6 feature: 581019816: [*]NN=C([*])[*]
11. Unknown FCFP_6 feature: -885520711: [*]C(=[*])NN=[*]
12. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

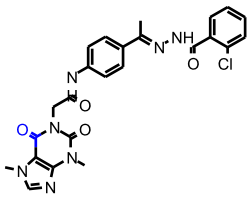
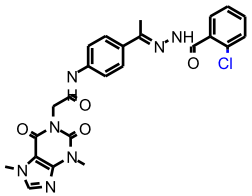
Feature Contribution

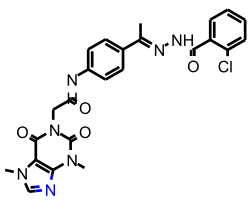
Top features for positive contribution

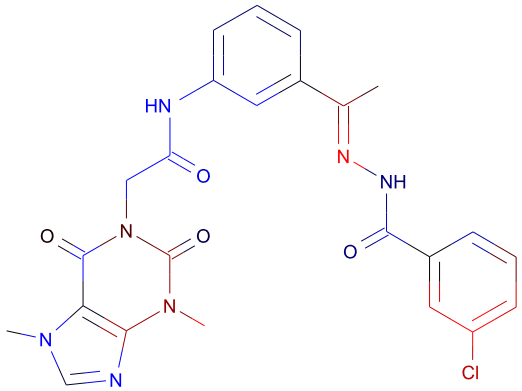
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
ECFP_6	99947387	 [*]:[c](:[*])Cl	0.181

Top Features for negative contribution

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

ECFP_6	655739385	 [*]:n:[*]	-0.239



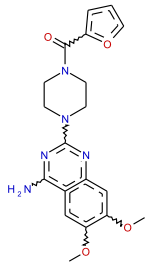
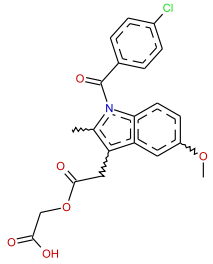
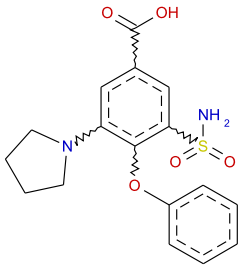
$C_{24}H_{22}ClN_7O_4$
Molecular Weight: 507.92898
ALogP: 2.034
Rotatable Bonds: 6
Acceptors: 6
Donors: 2

Model Prediction

Prediction: 1.31
Unit: g/kg_body_weight
Mahalanobis Distance: 24.3
Mahalanobis Distance p-value: 5.05e-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	PRASOZIN .HCl (HCl STRIPPED)	ACEMETACIN	PIRETANIDE
Structure			
Actual Endpoint (-log C)	2.294	4.235	1.811
Predicted Endpoint (-log C)	3.00765	3.39415	1.83976
Distance	0.721	0.729	0.751
Reference	NIIRDN 6;688;82	ARZNAD 30;1398;80	DRFUD4 2;393;77

Model Applicability

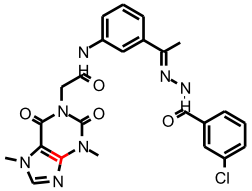
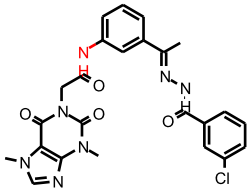
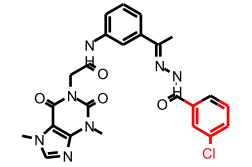
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[cH]:n:1
5. Unknown FCFP_6 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown FCFP_6 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown FCFP_6 feature: -1410049896: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 581019816: [*]NN=C([*])[*]
10. Unknown FCFP_6 feature: -885520711: [*]C(=[*])NN=[*]
11. Unknown FCFP_6 feature: 136150461: [*]:n(:[*])C
12. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

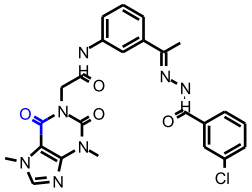
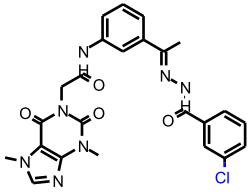
Feature Contribution

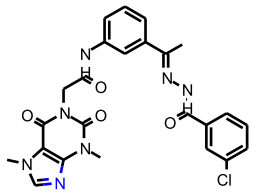
Top features for positive contribution

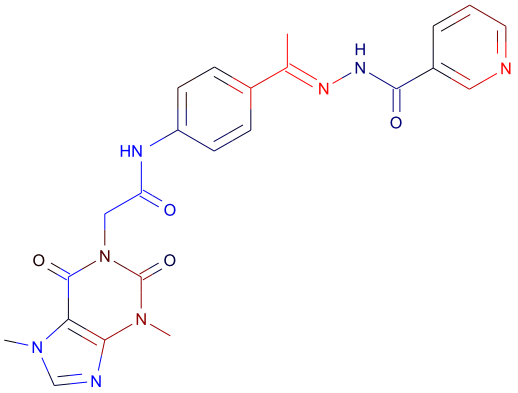
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
ECFP_6	577592657	 [*][c]1:[*]:[cH]:[cH] :[c](Cl):[cH]:1	0.194

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	655739385	 [*]:n:[*]	-0.239



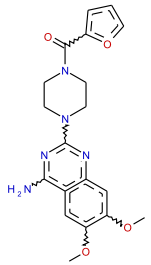
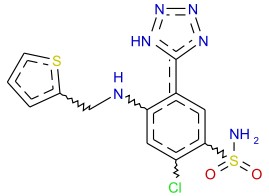
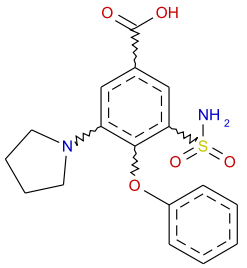
$C_{23}H_{22}N_8O_4$
Molecular Weight: 474.47198
ALogP: 0.219
Rotatable Bonds: 6
Acceptors: 7
Donors: 2

Model Prediction

Prediction: 0.768
Unit: g/kg_body_weight
Mahalanobis Distance: 24.2
Mahalanobis Distance p-value: 1.31e-026

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

Structural Similar Compounds

Name	PRASOZIN .HCl (HCl STRIPPED)	AZOSEMIDE	PIRETANIDE
Structure			
Actual Endpoint (-log C)	2.294	2.163	1.811
Predicted Endpoint (-log C)	3.00765	2.21052	1.83976
Distance	0.756	0.763	0.792
Reference	NIIRDN 6;688;82	IYKEDH 18;666;87	DRFUD4 2;393;77

Model Applicability

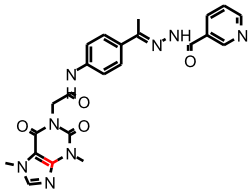
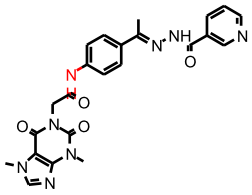
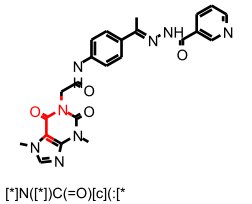
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[cH]:n:1
5. Unknown FCFP_6 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown FCFP_6 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown FCFP_6 feature: -1410049896: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown FCFP_6 feature: 136150461: [*]:n(:[*])C
9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
10. Unknown FCFP_6 feature: 581019816: [*]NN=C([*])[*]
11. Unknown FCFP_6 feature: -885520711: [*]C(=[*])NN=[*]

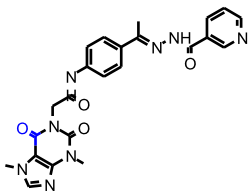
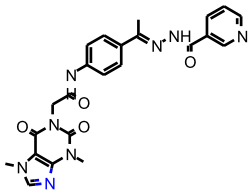
Feature Contribution

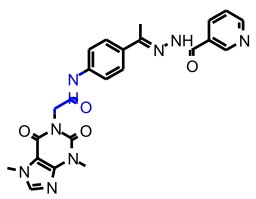
Top features for positive contribution

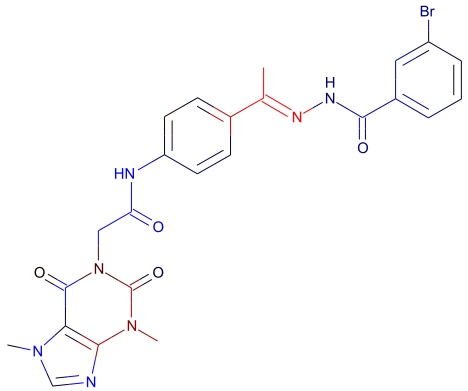
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	-1549163031	 [*]N([*])C(=O)[c](:[*]):[*]	0.171

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	655739385	 [*]:n:[*]	-0.239

FCFP_6	566058135	 [*]CC(=O)N[*]	-0.216



$C_{24}H_{22}BrN_7O_4$

Molecular Weight: 552.37997

ALogP: 2.118

Rotatable Bonds: 6

Acceptors: 6

Donors: 2

Model Prediction

Prediction: 1.78

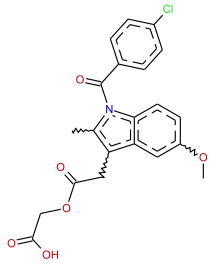
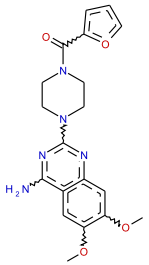
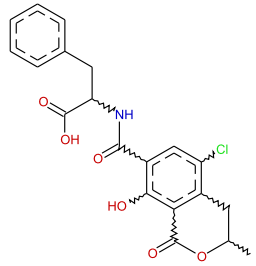
Unit: g/kg_body_weight

Mahalanobis Distance: 25

Mahalanobis Distance p-value: 5.19e-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	ACEMETACIN	PRASOZIN .HCl (HCl STRIPPED)	OCHRATOXIN A
Structure			
Actual Endpoint (-log C)	4.235	2.294	4.305
Predicted Endpoint (-log C)	3.39415	3.00765	3.03558
Distance	0.774	0.777	0.802
Reference	ARZNAD 30;1398;80	NIIRDN 6;688;82	FCTXAV 6;479;68

Model Applicability

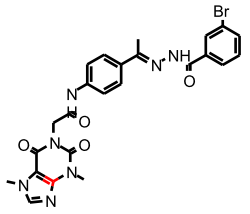
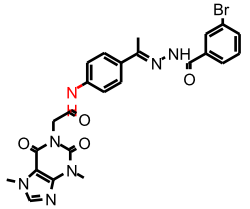
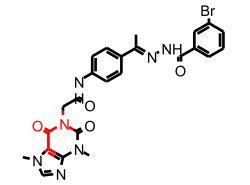
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC61 out of range. Value: 4.9641. Training min, max, SD, explained variance: -4.2722, 4.4724, 0.9936, 0.0036.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown FCFP_6 feature: 16: [*]:[cH]:[*]
4. Unknown FCFP_6 feature: 1747237384: [*][c]1:[*]:[*]:[cH]:n:1
5. Unknown FCFP_6 feature: -124685461: [*]n1:[*]:[*]:n:[cH]:1
6. Unknown FCFP_6 feature: -306856457: [*][c]1:[*]:[*]:[cH]:n:1C
7. Unknown FCFP_6 feature: -1410049896: [*]N([*])[c]1:n:[*]:[*]:[c]:1[*]
8. Unknown FCFP_6 feature: 136150461: [*]:n(:[*])C
9. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
10. Unknown FCFP_6 feature: 581019816: [*]NN=C([*])[*]
11. Unknown FCFP_6 feature: -885520711: [*]C(=[*])NN=[*]
12. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

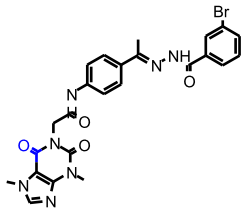
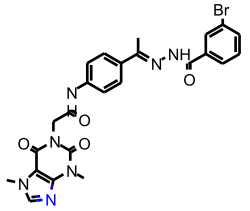
Feature Contribution

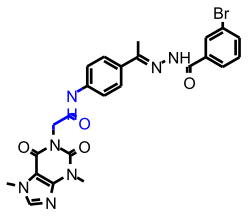
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	-1549163031	 [*]N([*])C(=O)[c](:[*]):[*]	0.171

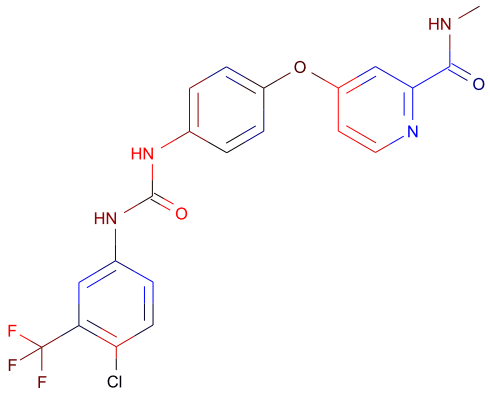
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	2106656448	 [*]C(=O)[*]	-0.352
ECFP_6	655739385	 [*]:n:[*]	-0.239

FCFP_6	566058135	 [*]CC(=O)N[*]	-0.216

Sorafenib

TOPKAT_Rat_Oral_LD50



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

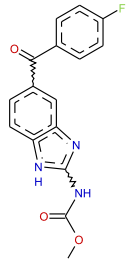
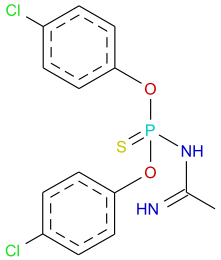
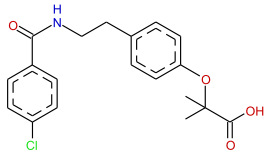
ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Structural Similar Compounds

Name	FLUBENDAZOLE	PHOSPHORAMIDOTHIOIC ACID; ACETIMIDOYL-; O;O-bis-(p-CHLOROPHENYL)ESTER	BEZAFIBRATE
Structure			
Actual Endpoint (-log C)	2.088	5.006	1.946
Predicted Endpoint (-log C)	2.69288	3.23989	2.54395
Distance	0.697	0.703	0.721
Reference	YRTMA6 9;11;78	FMCHA2 -;C149;89	ARZNAD 30;2023;80

Model Prediction

Prediction: 0.823

Unit: g/kg_body_weight

Mahalanobis Distance: 21

Mahalanobis Distance p-value: 1.93e-012

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

Model Applicability

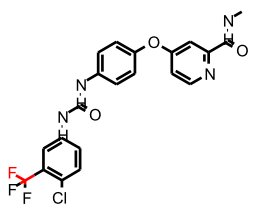
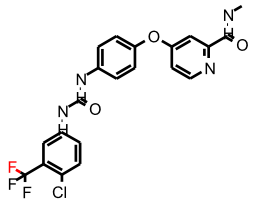
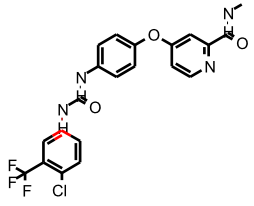
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

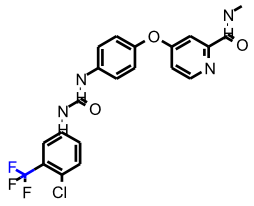
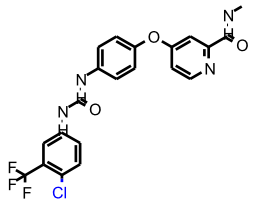
Feature Contribution

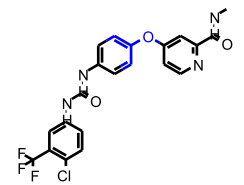
Top features for positive contribution

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