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

Lab code	Paper code
WN-5	**7a**
WN-4	**7b**
WN-13	**7c**
WN-6	**7d**
WN-15	**7e**
WN-17	**7f**
WN-16	**7g**

S.1. Chemistry

The chemicals, reagents, and reaction solvents used in this study were gained from Sigma-Aldrich, Alpha Chem, Fluka, and Loba and were not purified further. The melting points (mp), which have not been corrected, were measured using the SMP50 Digital Melting Point App provided by Bibby Scientific in Staffordshire. A Thermo Fisher Nicolet IS10 spectrophotometer was used to detect infrared spectra as solids on the potassium bromide disc (ν_{max} in cm^{-1}) with a resolution of 4.0 cm^{-1}, covering 4000-400 cm^{-1}.^{1}H-NMR and ^{13}C-NMR spectra (400 and 101 MHz) were recorded at the JNM-ECA 500 II Made by JEOL-JAPAN instrument through a solution of deuterated dimethyl sulfoxide. Proton chemical shifts are labeled in part per million (ppm), downfield from tetramethyl silane (TMS, δ=0) as an internal standard, and the following abbreviations (or a combination thereof) are used to describe splitting patterns: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet, and br, broad. Mass spectra were measured with a Thermo Scientific GCMS model (Isq Lt) using the Thermo X-Calibur software (Shimadzu, Kyoto, Japan) at the Regional Center for Mycology and Biotechnology (RCMB), Al-Azhar University, Nasr City, Cairo, Egypt. Elemental studies were conducted at the Regional Center for Microbiology and Biotechnology, Al-Azhar University, Cairo, Egypt, with results accurate to within 0.4%. Thin-layer chromatography (TLC) was carried out on silica gel plates by using DCM:

MeOH (95:5%), as the eluting system. The progress of the reaction and evaluation of product purity was determined using a UV indicator at 254 nm.

S2. Biological testing

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

Materials and methods

Cell line

Mammary gland breast cancer (MCF-7) and (MDA-MB-231) . The cell line was obtained from ATCC via Holding company for biological products and vaccines (VACSERA), Cairo, Egypt.

Sorafenib was used as a standard anticancer drug for comparison.

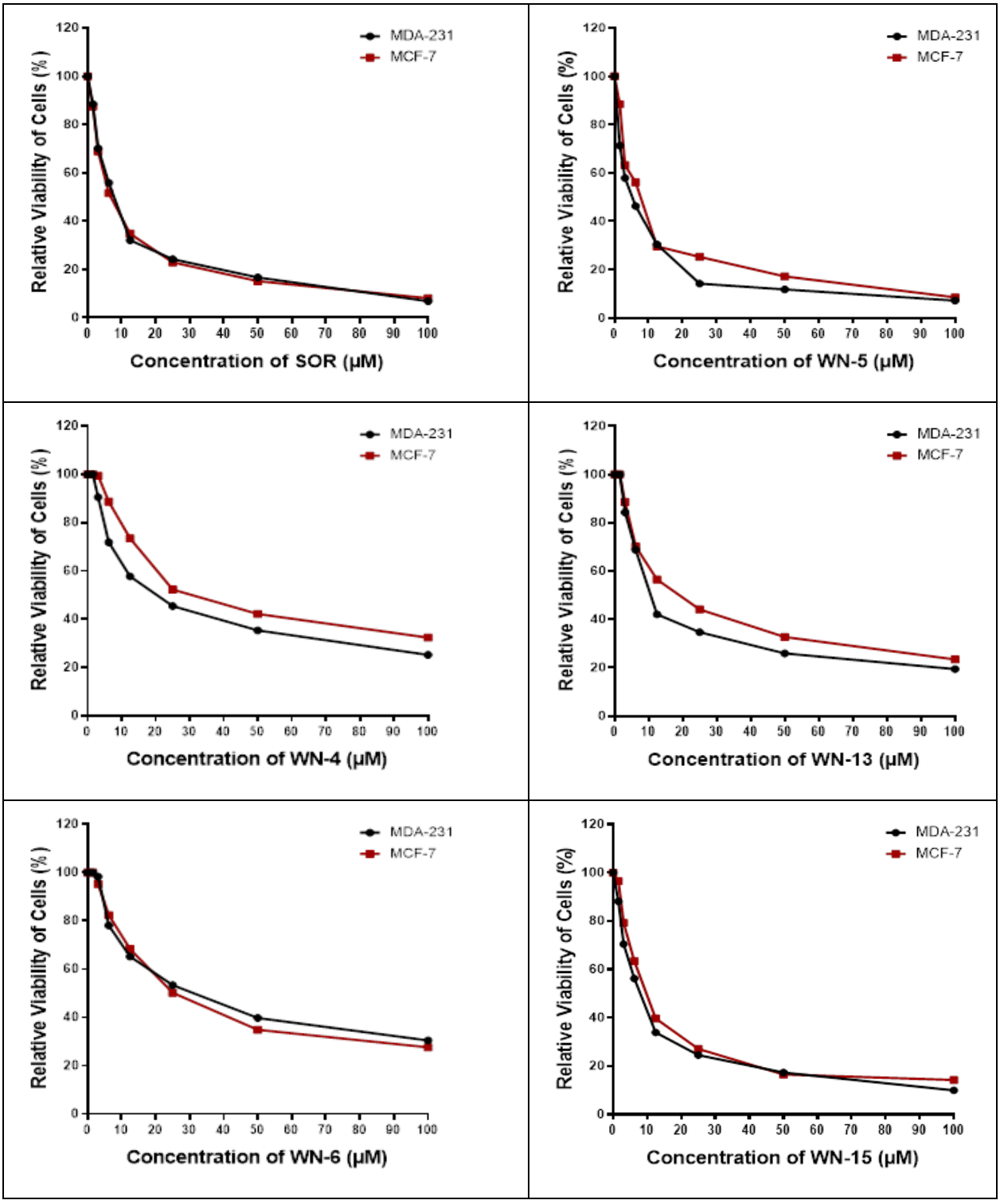
Chemical reagents

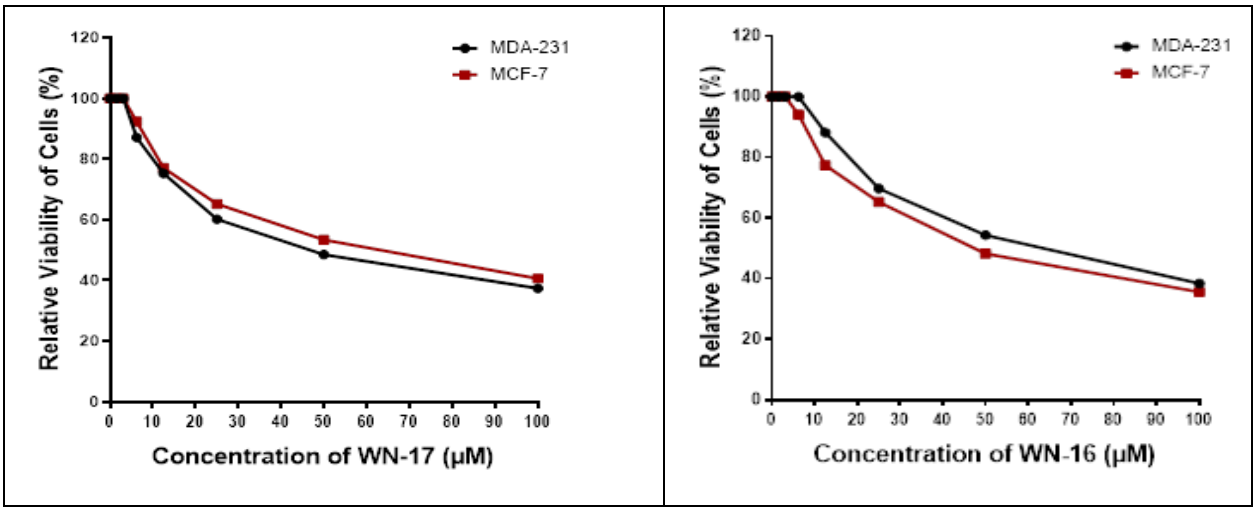
The reagents RPMI-1640 medium , MTT and DMSO (sigma co., St. Louis, USA), Fetal Bovine serum (GIBCO, UK) .

MTT assay

The cell lines mentioned above were used to determine the inhibitory effects of compounds on cell growth using the MTT assay. This colorimetric assay is based on the conversion of the yellow tetrazolium bromide (MTT) to a purple formazan derivative by mitochondrial succinate dehydrogenase in viable cells. Cell lines were cultured in RPMI-1640 medium with 10% fetal bovine serum. Antibiotics added were 100 units/ml penicillin and 100µg/ml streptomycin at 37 C in a 5% Co_2 incubator. The cell lines were seedesin a 96-well plate at a density of $1.0x10^4$ cells/well. at 37 C for 48 h under 5% Co_2. After incubation the cells were treated with different concentration of compounds and incubated for 24 h. After 24 h of drug treatment, 20 µl of MTT solution at 5mg/ml was added and incubated for 4 h. Dimethyl sulfoxide (DMSO) in volume of 100 µl is added into each well to dissolve the purple formazan formed. The colorimetric assay is measured and recorded at absorbance of 570 nm using a plate reader (EXL 800 ,USA). The relative cell viability in percentage was calculated as (A570 of treated samples/A570 of untreated sample) X 100.

***Dose response curves**





S.2.2. Safety assay

Materials and methods

Cell line

Human lung fibroblast cell line (WI-38) and Human amnion (WISH) . The cell line was obtained from ATCC via Holding company for biological products and vaccines (VACSERA) , Cairo, Egypt.

Chemical reagents

The reagents RPMI-1640 medium , MTT and DMSO (sigma co., St. Louis, USA), Fetal Bovine serum (GIBCO, UK) .

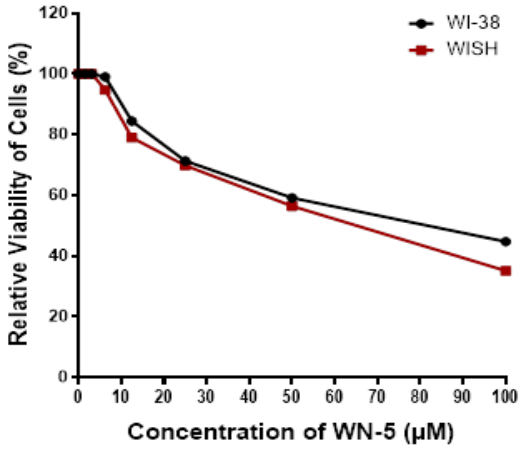
Sorafenib was used as a standard anticancer drug for comparison.

MTT assay

The different cell line mentioned above were used to determine the inhibitory effects of compounds on cell growth using the MTT assay. This colorimetric assay is based on the conversion of the yellow tetrazolium bromide (MTT) to a purple formazan derivative by mitochondrial succinate dehydrogenase in viable cells. The cells were cultured in RPMI-1640 medium with 10% fetal bovine serum. Antibiotics added were 100 units/ml penicillin and 100µg/ml streptomycin at 37 C in a 5% Co_2 incubator. The cellswere seedesin a 96-well plate at a density of $1.0x10^4$ cells/well.at 37 C for 48 h under 5% Co_2. After incubation the cells were treated with different concentration of compounds and incubated for 24 h. After 24 h of drug treatment, 20 µl of MTT solution at 5mg/ml was added and incubated for 4 h. Dimethyl sulfoxide(DMSO) in volume of 100µl is added into each well to dissolve the purple formazan formed. The colorimetric assay is measured and

recorded at absorbance of 570nmusing a plate reader (EXL 800, USA). The relative cell viability in percentage was calculated as (A570 of treated samples/A570 of untreated sample) X 100.

***Dose response curves**



S.2.4. *In vitro* VEGFR-2 inhibition

- The in vitro inhibitory activity of the tested compounds against **VEGFR-2** was accomplished using Kinase Assay Kit (BPS Bioscience, USA) at different concentrations 1, 0.1, and 0.01 µM.
- The activities of Kinases were observed by measuring chemiluminescence using BioTek™ Synergy2 Microplate Reader (BioTek, USA).
- Different standards were used.
- All samples and controls were verified in triplicates to calculate the concentration that caused 50% inhibition of the kinase activity.

S.2.5. Cell cycle analysis

This assay was conducted according to the reported methods. In short, MDA-MB-231 cells were seeded into six-well plates (2 × 105 cells per each well) and incubated for 24 h at 37 °C and 5% CO2. After that, they were treated with 4.64 µM of compound **7a** dissolved in DMSO (1% v/v) for 48 h. After the cells were washed several times with cold phosphate buffered saline (PBS), fixed with ethyl alcohol (70%) and rinsed again with PBS, they were stained with the DNA fluorochrome propidium iodide and kept at 37 °C for 15 min in the

dark. The cell distribution was analyzed using Epics XL-MCL™ Flow Cytometer (Beckman Coulter), and the data were analyzed using Flowing software (version 2.5.1, Turku Centre for Biotechnology, Turku, Finland).

S.2.6. Apoptosis analysis

Apoptosis detection was carried out using Annexin V fluorescein isothiocyanate (V-FITC)/PI kit according to the reported procedures. MDA-MB-231 cells cells were seeded and incubated for 24 h, then treated with (4.64 µM) of compound **7a** for 48 h. The cells were then collected, washed 3 times with PBS, fixed with ice-cold absolute ethanol (70%) and stained with Annexin-V-FITC/propidium iodide (PI) using a double staining kit for 20 min in the dark. Epics XL-MCL™ Flow Cytometer was used to evaluate the apoptosis.

S.2.7. *In vitro* assay for BAX, Bcl-2, and caspase-3

MDA-MB-231 cells were cultured in suspension in complete culture medium (CCM) of RPMI 1640 culture medium (PAA; Pasching, Austria) containing 10% foetal bovine serum (FBS; PAA) and 1% penicillin/streptomycin (GIBCO, Invitrogen, Paisley, UK) into an incubator at 37 °C with 5% CO_2. Cells were cultured at a ratio of 0.2 10^6 cells/well in 96-wells plates and differentiated into macrophage-like cells with 10 ng/mL phorbol myristate acetate (PMA; Sigma Chemical Co.) for a period of 24 h. Cell rested then in CCM without PMA for another 24 h. After differentiation and resting periods, vehicle (DMSO; MERK, Whitehouse Station, NJ, USA) as control and compound 7a (at a concentration of 4.64 µM) were added to each well and plates were pre-incubated at 37 °C with 5% CO_2 for 72 h.

Supernatants from *in vitro* assays were harvested and assayed for BAX, Bcl-2, and caspase- 3 using an enzyme-linked immunosorbent assay (ELISA) kit following the manufacturer's instructions. For caspase-3, DRG® Caspase-3 (human) ELISA (EIA-4860) was used. For BAX and Bcl2, DRG® Human Bax ELISA (EIA-4487) (DRG International, Inc., USA), and Zymed® Bcl-2 ELISA Kit (Invitrogen Corporation,1600 Faraday Avenue, Carlsbad, CA, USA) were used. The absorbance in each well was measured with a microplate reader at 450 nm and corrected at 570 nm. Concentrations of the tested proteins were obtained and the percentage of activation and/or inhibition as compared to control conditions.

S3. *In silico* studies

S.3.1. Docking studies

Protein Preparation: The crystal structure of VEGFR-2 [PDB ID: 4ASD, resolution: 2.03 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 6.0 and saved in MDL-SD file format. The 3D structures of the ligands were protonated, and the structures were optimized by energy minimization using MM2 force-field and 10000 iteration steps of 2 fs. The conformationally optimized ligands were used for docking studies.

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

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

S.3.2. MD simulations

Molecular Dynamic (MD) Simulation:

An unbiased molecular dynamics (MD) simulation was used to evaluate the binding affinity and stability of the VEGFR-2_**7a** complex. A 200 ns trajectory was made for the system using GROMACS 2021. CHARMM-GUI solution builder was used to prepare starting files. The system was solvated using water molecules with a transferable intermolecular potential of three points (TIP3P) in a 10 nm cubic box with a buffer distance of 1 nm. Neutralization was achieved by adding NaCl ions at a concentration of 0.154 M. The CHARMM36m force field was used for the VEGFR-2 protein, TIP3P water, and ions.

All simulation dimensions employed periodic boundary conditions. Potential energy minimization with the steepest descent method was first performed to remove steric conflicts. When the highest force on any atom fell below 100 kJ/(mol·nm) or after 100,000 steps, convergence was reached. Thermodynamic equilibrium was the aim of the two-step equilibration procedure. The average temperature in the first stage (NVT ensemble) was maintained at 310 K by the Velocity Rescale thermostat. The second stage (NPT ensemble) used the Berendsen barostat to regulate pressure at 1 atm and the Velocity Rescale algorithm to maintain a temperature of 310 K. A Parrinello-Rahman barostat was used to regulate pressure at 1 atm, and a Nose-Hoover thermostat was used for temperature control in the production run (NPT ensemble). The hydrogen atom bond lengths were maintained with the LINear Constraint Solver (LINCS) method. Using a 1.2 nm cutoff, the Particle Mesh Ewald (PME) approach was used to calculate electrostatic interactions.

The leap-frog integrator propagated the Newtonian equations of motion using time steps of 1 fs for equilibration and 2 fs for production. Two thousand simulation frames were obtained at intervals of 0.1 ns.

After periodic boundary conditions were eliminated and protein integrity was restored, VMD TK scripts were used to analyse the trajectories. Many calculations were performed, such as the Root Mean Square Deviation (RMSD) of the VEGFR-2 backbone and the ligand system **7a**. Other structural features that were evaluated included the number of hydrogen bonds between ligands and proteins, center of mass distances between ligands and proteins, radius of gyration (RoG), solvent accessible surface area (SASA), and root mean square fluctuation (RMSF). Protein-Ligand Interaction Fingerprints (ProLIF), a Python program, assesses interactions between ligands and amino acids quantified and characterized the frequency and type of interactions in each frame.

Binding free energy calculation using MM-GBSA:

The ligands' binding free energies were ascertained by applying the Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) method through the gmx_MMPBSA program. To determine the contributing amino acids within a 1 nm radius of the ligand, a decomposition analysis was carried out. The ionic strength of 0.154 M and the solvation parameter (igb) of 5 were used in the simulations. The values of 1.0 and 78.5 were selected for the internal and external dielectric constants, respectively. Equation 1 describes in detail the basic free energy calculation approach.

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

where < > is the mean of the enclosed free energy of the ligand, complex, and receptor throughout the computation frames. We employed the entire route (a total of 2000 frames) in our method. Equations 2 through 6 can be used to compute various energy terms in the following ways:

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

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

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

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

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

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.

Principal Component Analysis:

The mobility of the α-carbon of VEGFR-2 was evaluated using Principal Component Analysis (PCA) and the mass-weighted covariance matrix (C). The final equilibrium frame of the VEGFR-2_**7a** system for combined analysis (for the free energy landscape) and the final equilibrium frame of each trajectory for individual trajectories are referred to during the alignment process. Using gmx covar to diagonalize the

covariance matrix in GROMACS, eigenvectors representing the dominating atomic motions were found. Eigenvalues indicated the amount of motion; the first principal component captured the largest variance, while the contributions from the other components decreased. The gmx anaeig command was utilized for analysis with GROMACS. We computed the dimensions of the critical subspace using three parameters. 1) The variance gathered by each additional eigenvector was visualized, and the cumulative sum of the eigenvalues was calculated using extra eigenvectors. Furthermore, 2) a scree plot was generated, which displayed each eigenvalue in relation to its corresponding eigenvector index. The index with the largest slope reduction indicated the critical subspace size. Moreover, 3) the distribution of the eigenvectors was evaluated.

When a distribution other than a Gaussian distribution emerged, it indicated that the corresponding eigenvectors had meaningful dynamics.

We computed the cosine content (ci) of each eigenvector of the C matrix, which may take values ranging from 0 (no cosine) to 1 (perfect cosine). The following is the cosine content equation:

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

Where T is the time of the simulation. Abnormally large ci values, which represent random motion, are related to insufficient sampling. When the cosine content of the first few PCs is near 1, the behavior of proteins on a large scale is analogous to diffusion. Accordingly, the first 10 PCs were used to calculate their cosine content.

Free Energy Landscape (FEL):

A protein's free-energy landscape (FEL) can be used to determine the different conformations it can adopt, which is necessary to understand the protein's dynamics and function. In this investigation, conformational sampling methods are often used, such as molecular dynamics (MD) simulations, which mimic the intrinsic temperature changes in proteins over time. Here, we have revealed the FEL of our protein of interest through molecular dynamics simulations. To improve its visual appeal, we projected this complex onto two chosen response coordinates. The free energy (Gα) of the protein in a state described by these coordinates is related to the probability of detecting the protein in that state via the exponential relationship e-Gα/kT. These computations can be used to determine the entire FEL:

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

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

S.3.4. Density Function Theory (DFT) calculations

DFT calculations

The Gaussian 9 package under B3LYP/6-311+G (d, p) level has been used to calculate density function theory (DFT). Several tools have been used, including Avogadro, GaussView 0.3, and Multiwfn. The following formulas were used to calculate the reactivity descriptors: ionization energy (I), chemical potential (μ), electron affinity (A), electrophilicity (χ), maximal charge acceptance (ΔNmax), softness (σ), and hardness (η):

$IP = -E_{HOMO}$

$EA = -E_{LUMO}$

μ= (IP+EA)/2

η= (IP-EA)

χ= - μ

ω= $\mu^2/(2\ \eta)$

σ= 1/ η

ΔN= -(μ/η)

ΔE = -ω

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

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:**

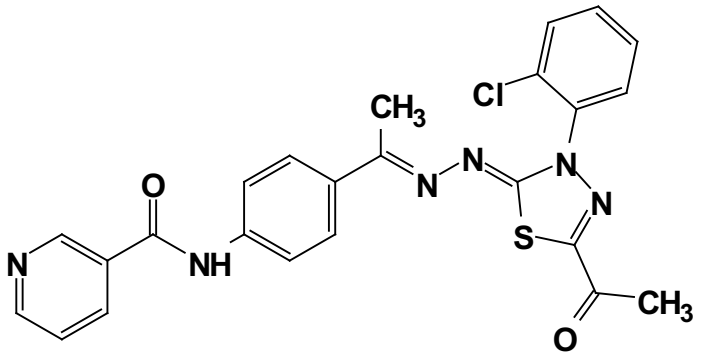
In this protocol, the general-purpose panel was utilized with the activation of the Prepare ligand option. The change ionization was switched on the true option using the Rule based as an ionization method. In Rule based task, we used the carboxylate as an acid ionization. Additionally, the primary,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

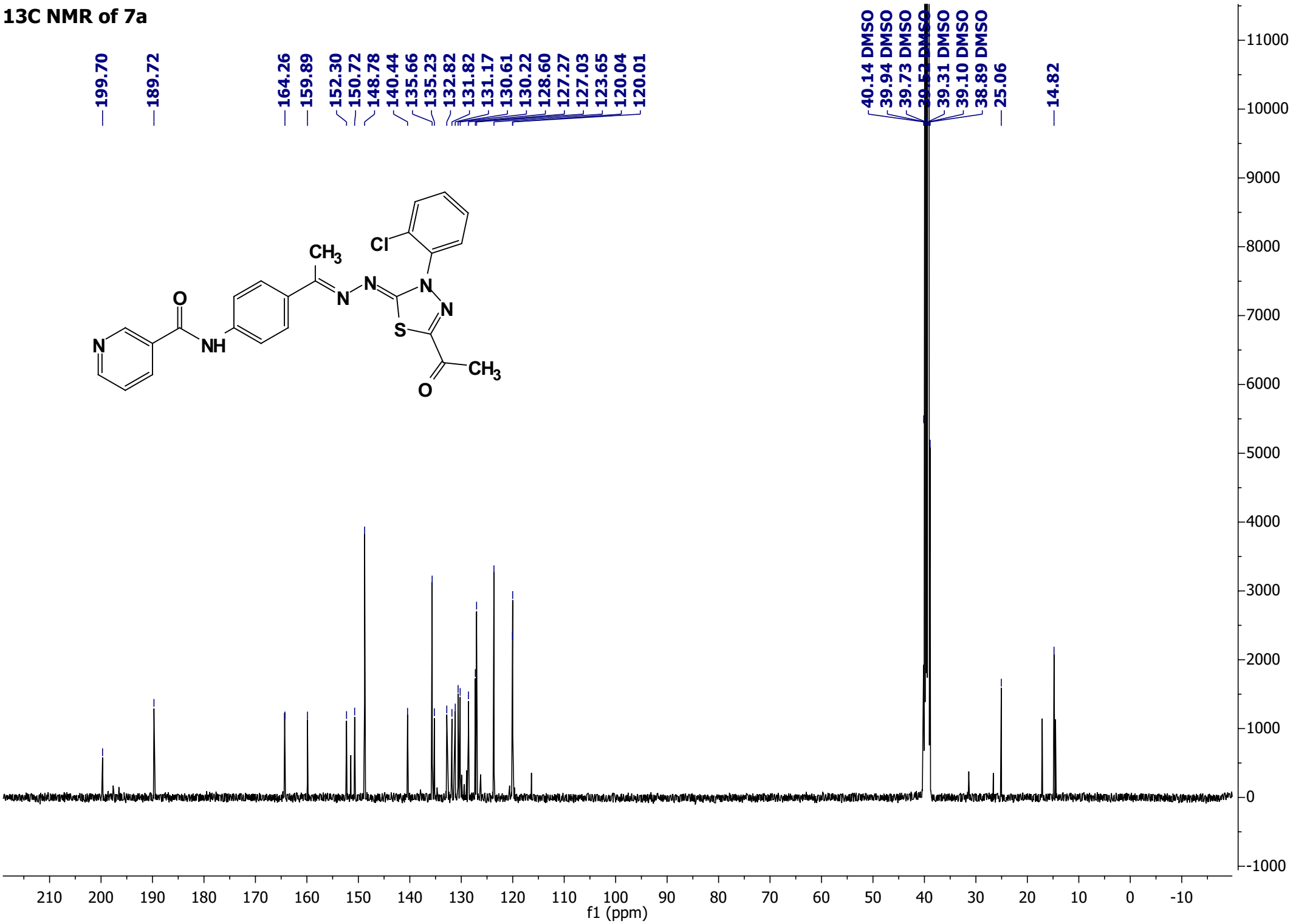

% Transmittance
Wavenumbers (cm-1)
100
90
80
70
60
50
40
30
20
10
0
-10
-20
4000
3800
3600
3400
3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
3322.27
3194.51
3132.30
3105.33
3069.52
3037.14
2998.00
2939.71
2877.07
1914.61
1799.03
1677.26
1657.24
1597.99
1580.79
1558.35
1484.69
1443.41
1409.96
1364.36
1336.23
1280.39
1259.25
1222.49
1149.01
1130.56
1105.69
1091.82
1075.63
1025.80
956.16
900.00
835.80
793.16
772.13
754.02
719.66
689.76
623.01
598.48
562.08
542.25
523.29
499.27
446.10

1H NMR of 7a


Cl
CH3
N
N
N
N
S
O
NH
N
CH3
O

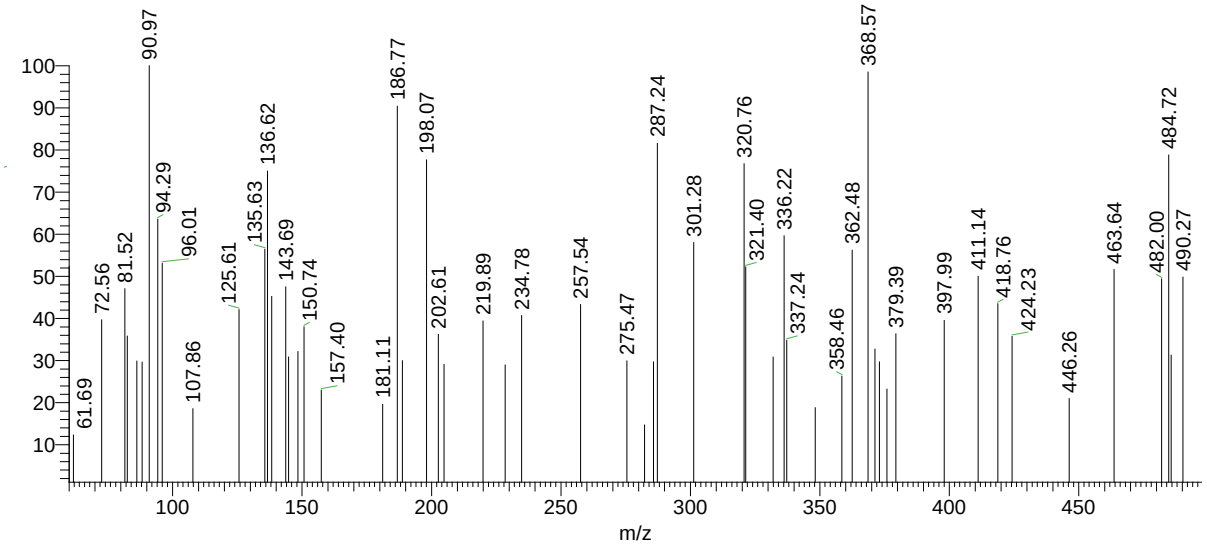
10.61
9.12
9.11
9.11
9.11
8.78
8.77
8.77
8.76
8.75
8.32
8.31
8.31
8.30
8.29
7.88
7.87
7.86
7.86
7.85
7.84
7.84
7.82
7.81
7.80
7.79
7.78
7.77
7.76
7.76
7.63
7.62
7.61
7.61
7.61
7.60
7.59
7.59
7.59
7.58
7.57
7.57
7.56
7.55
3.48
3.44 HDO
2.52
2.51 DMSO
2.50 DMSO
2.50 DMSO
2.50 DMSO
2.49 DMSO
2.20
1.02
1.16
1.35
1.33
4.41
2.09
3.22
3.42
3.06
f1 (ppm)

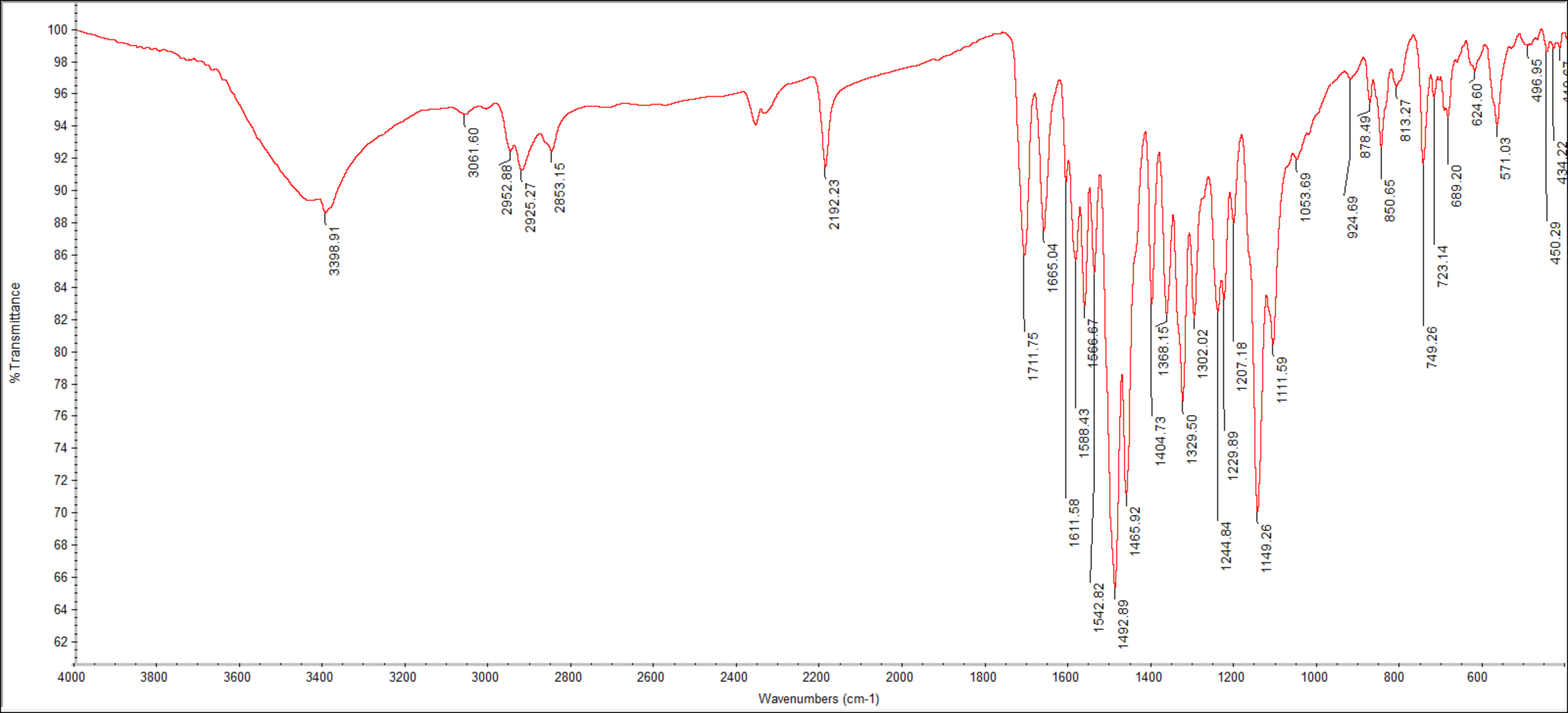
13C NMR of 7a


199.70
189.72
164.26
159.89
152.30
150.72
148.78
140.44
135.66
135.23
132.82
131.82
131.17
130.61
130.22
128.60
127.27
127.03
123.65
120.04
120.01
40.14 DMSO
39.94 DMSO
39.73 DMSO
39.52 DMSO
39.31 DMSO
39.10 DMSO
38.89 DMSO
25.06
14.82
CH3
Cl
O
N
NH
N
N
N
N
S
CH3
O
f1 (ppm)

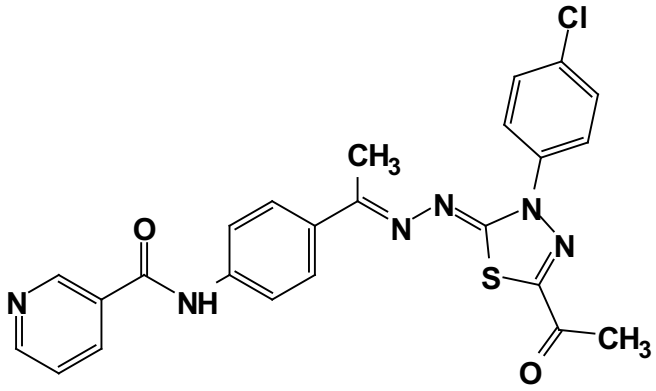
Mass spec. of compound 7a

wn5 #125 RT: 2.11 P: + NL: 2.61E2
T: {0,0} + c EI Full ms [40.00-1000.00]



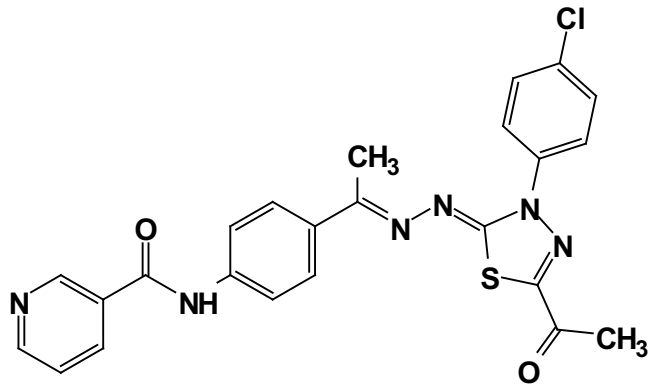

% Transmittance
Wavenumbers (cm-1)
3398.91
3061.60
2952.88
2925.27
2853.15
2192.23
1711.75
1665.04
1611.58
1588.43
1566.67
1542.82
1492.89
1465.92
1404.73
1368.15
1329.50
1302.02
1244.84
1229.89
1207.18
1149.26
1111.59
1053.69
924.69
878.49
850.65
813.27
749.26
723.14
689.20
624.60
571.03
496.95
450.29
434.22

1H NMR of 7b


Cl
CH3
O
N
NH
N
N
N
N
S
CH3
O

10.61
9.13
8.78
8.77
8.33
8.32
8.32
8.31
8.30
8.30
8.14
8.14
8.12
8.11
7.92
7.90
7.88
7.87
7.86
7.69
7.67
7.66
7.65
7.64
7.63
7.63
7.61
7.60
7.58
7.58
7.56
3.32 HDO
2.59
2.51 DMSO
2.50 DMSO
2.50 DMSO
2.49 DMSO
2.49 DMSO
2.44
2.43
1.00
1.01
1.02
1.10
2.20
4.45
2.07
1.24
3.13
3.23
f1 (ppm)

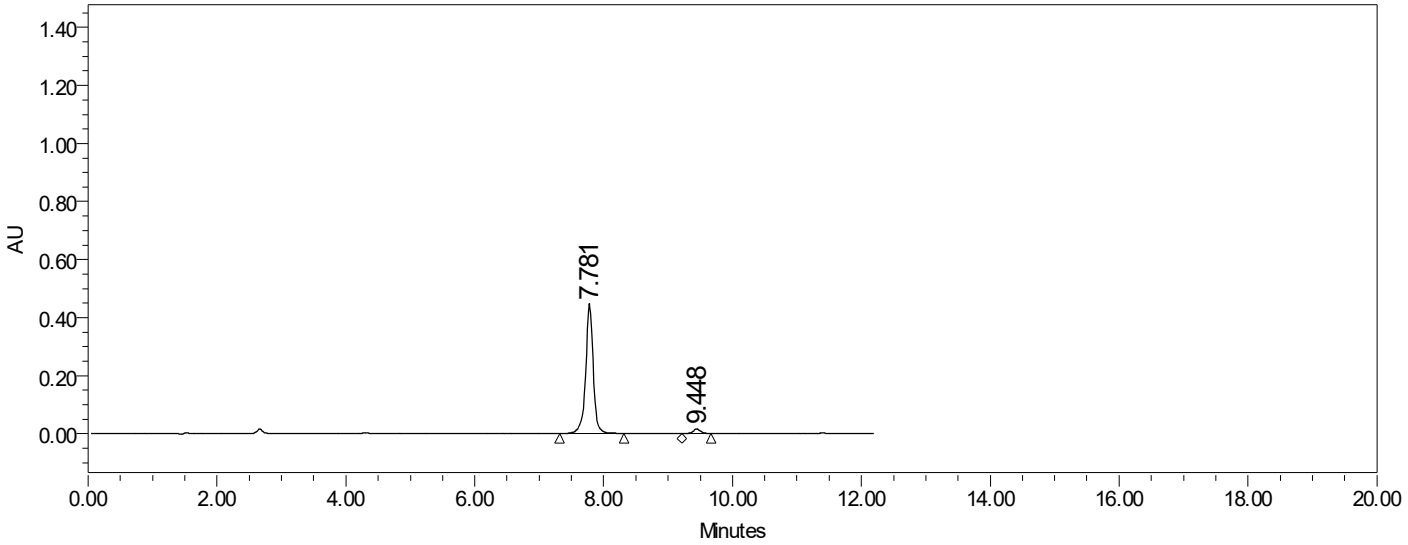
13C NMR of 7b


Cl
CH3
O
NH
N
N
N
N
S
CH3
O

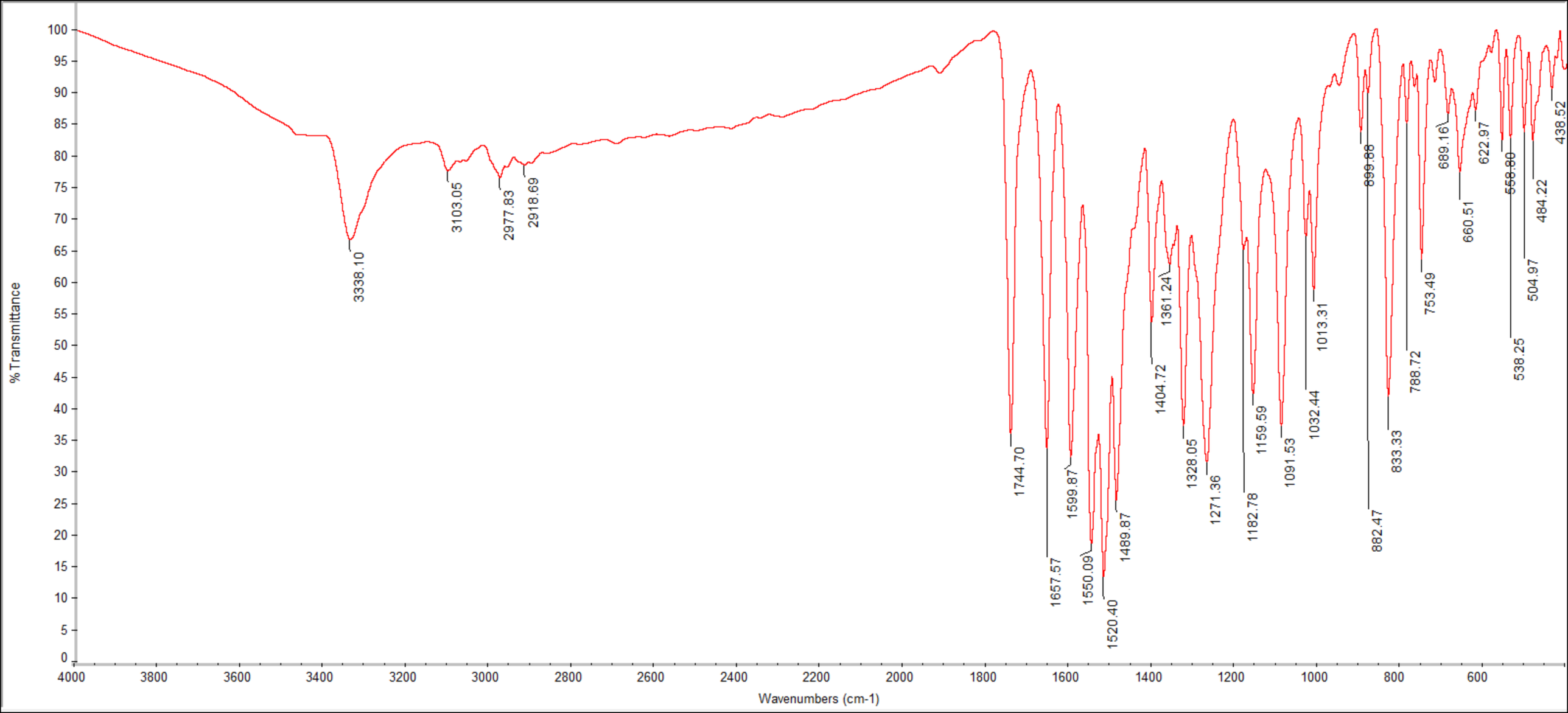
189.77
164.21
163.93
160.26
152.22
151.01
148.74
140.55
137.67
135.56
132.53
131.02
129.12
127.11
123.53
123.36
119.88
40.15 DMSO
40.00 DMSO
39.95 DMSO
39.74 DMSO
39.53 DMSO
39.32 DMSO
39.11 DMSO
38.90 DMSO
25.05
15.33
2500
2000
1500
1000
500
0
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

SAMPLE INFORMATION

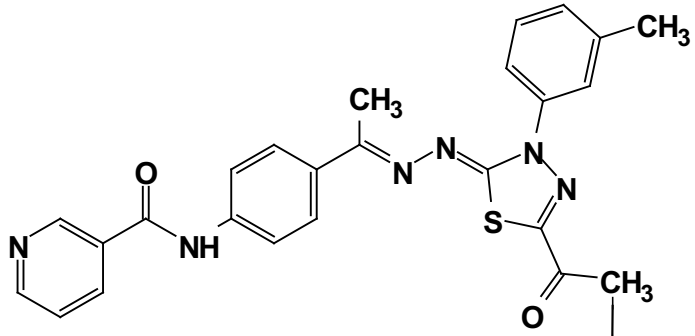
Sample Name:	WN-4	Acquired By:	YARA
Sample Type:	Unknown	Sample Set Name:	
Vial:	35	Acq. Method Set:	Organic
Injection #:	1	Processing Method:	Default
Injection Volume:	4.00 ul	Channel Name:	336.0nm@2
Run Time:	20.0 Minutes	Proc. Chnl. Descr.:	W2996 PDA 336.0 nm (PDA 190.0 to
Date Acquired:	22/05/2024 03:39:26 PM EET		
Date Processed:	23/05/2024 11:54:14 AM EET		



	RT	Area	% Area	Height
1	7.781	3660422	96.71	445448
2	9.448	124459	3.29	15606

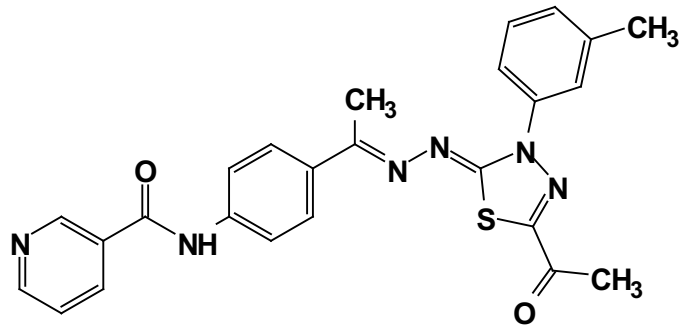

% Transmittance
Wavenumbers (cm-1)
3338.10
3103.05
2977.83
2918.69
1744.70
1657.57
1599.87
1550.09
1520.40
1489.87
1404.72
1361.24
1328.05
1271.36
1182.78
1159.59
1091.53
1032.44
1013.31
899.88
882.47
833.33
788.72
753.49
689.16
660.51
622.97
558.80
538.25
504.97
484.22
438.52

1H NMR of 7c


CH3
CH3
O
N
N
N
N
N
S
NH
CH3
O

10.61
9.13
9.12
8.78
8.78
8.77
8.77
8.33
8.32
8.32
8.31
8.30
8.30
7.91
7.89
7.88
7.86
7.59
7.58
7.57
7.56
7.47
7.45
7.43
7.23
7.21
3.33 HDO
3.18
3.16
2.58
2.50 DMSO
2.50 DMSO
2.49 DMSO
2.49 DMSO
2.41
2.41
2.40
1.00
1.02
1.00
1.07
6.35
1.10
1.14
1.05
3.36
6.15
f1 (ppm)

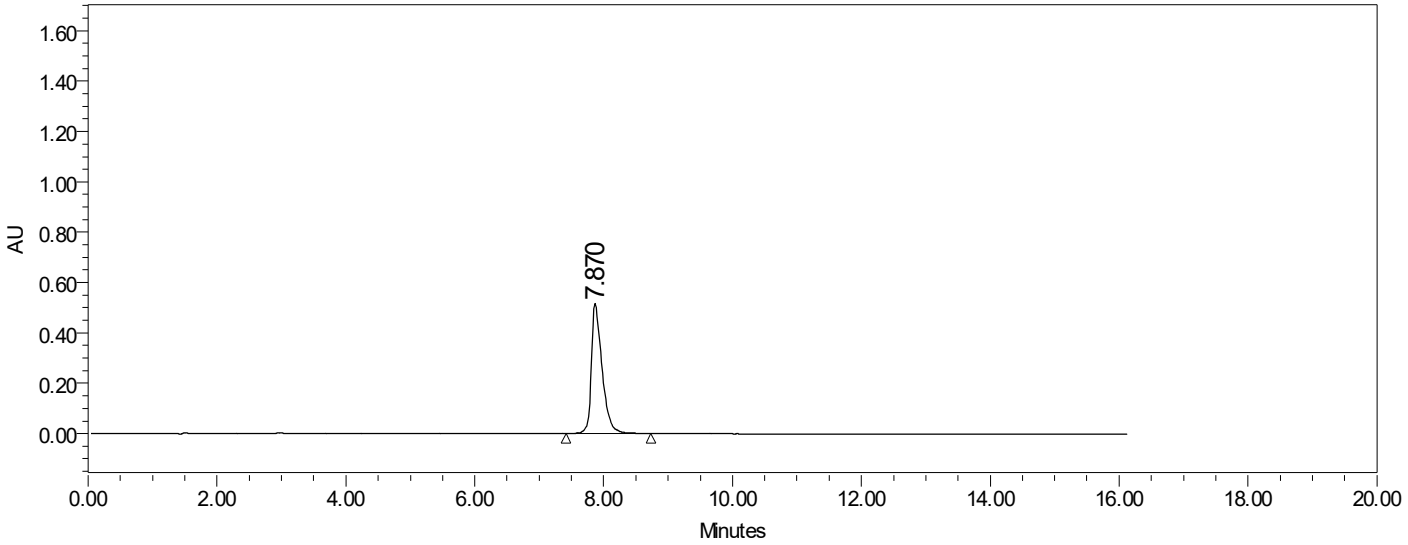
13C NMR of 7c


CH3
CH3
O
N
NH
N
N
N
N
S
CH3
O

189.89
164.32
159.95
152.30
150.59
148.79
140.50
138.85
138.75
135.66
132.75
130.52
129.01
127.89
127.12
123.64
122.50
119.99
119.23
40.20
40.15 DMSO
39.94 DMSO
39.73 DMSO
39.52 DMSO
39.31 DMSO
39.10 DMSO
38.89 DMSO
25.11
21.23
15.26
5000
4500
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 (ppm)

SAMPLE INFORMATION

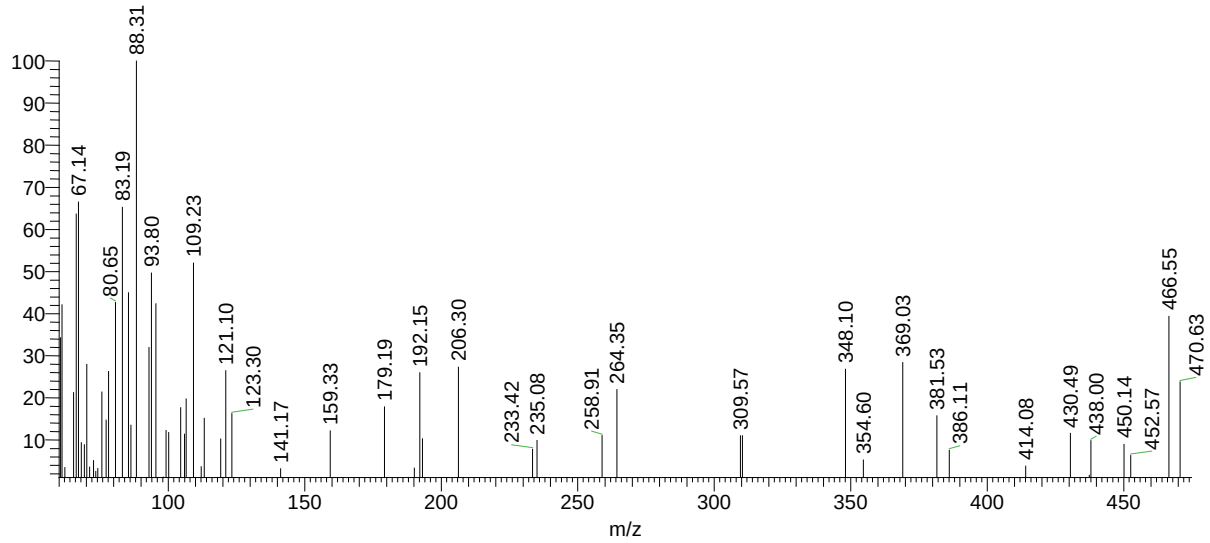
Sample Name:	WN-13	Acquired By:	YARA
Sample Type:	Unknown	Sample Set Name:	
Vial:	41	Acq. Method Set:	Organic
Injection #:	1	Processing Method:	Default
Injection Volume:	4.00 ul	Channel Name:	334.0nm@1
Run Time:	20.0 Minutes	Proc. Chnl. Descr.:	W2996 PDA 334.0 nm (PDA 190.0 to
Date Acquired:	22/05/2024 05:51:59 PM EET		
Date Processed:	23/05/2024 11:52:02 AM EET		

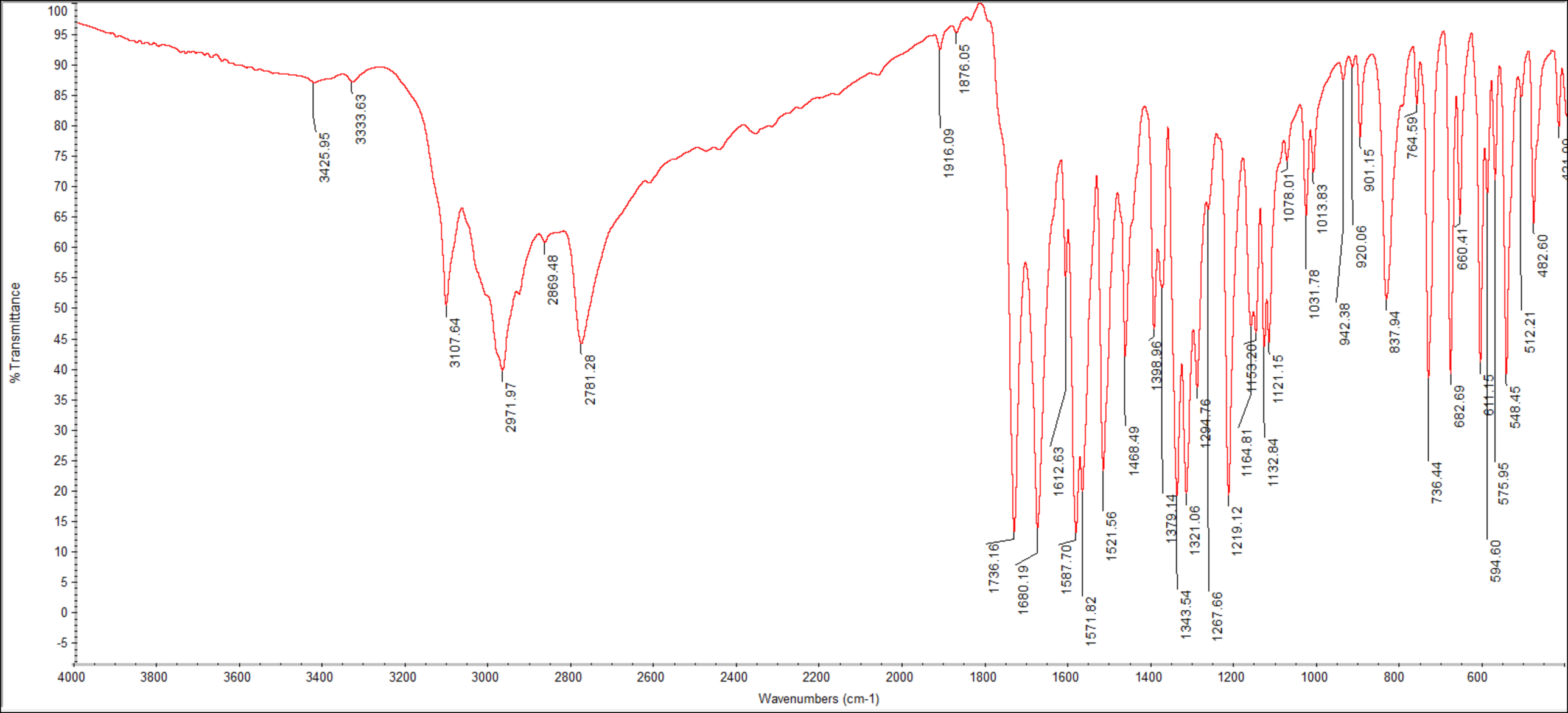


	RT	Area	% Area	Height
1	7.870	5986203	100.00	516604

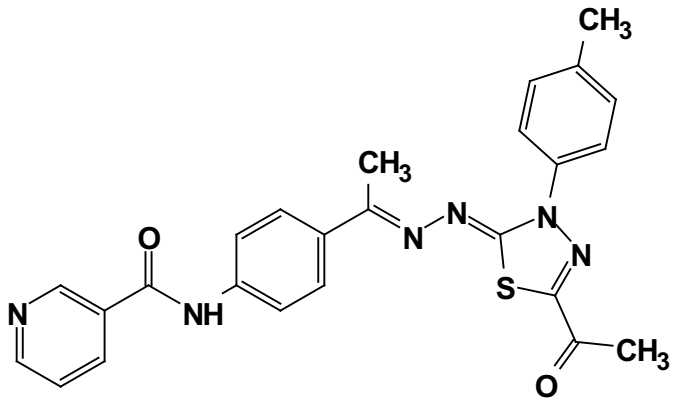
Mass spec. of compound 7c

wn13 #134 RT: 2.26 P: + SB: 6 2.18 , 2.11-2.18 NL: 1.51E3
T: {0,0} + c EI Full ms [40.00-1000.00]



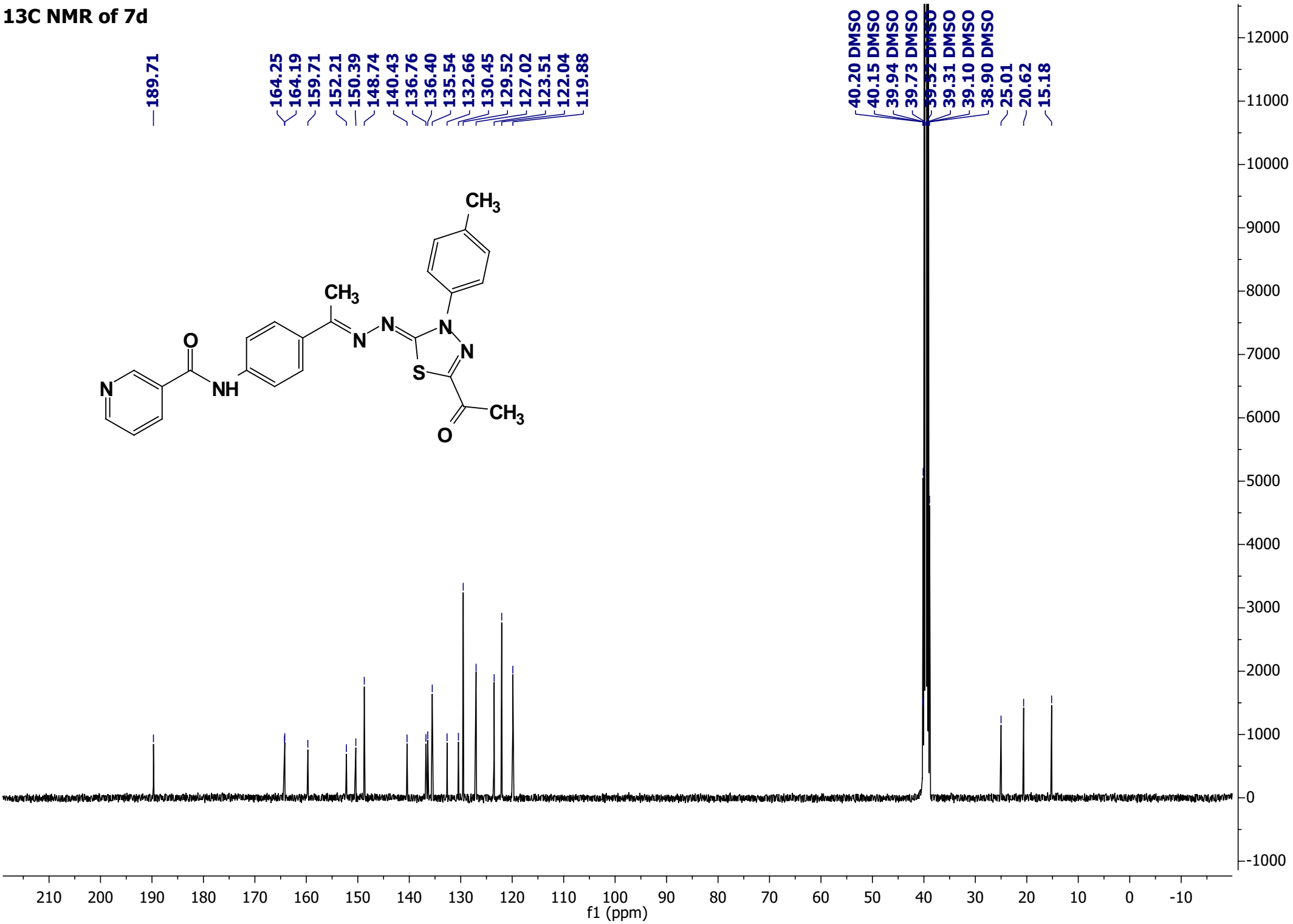

% Transmittance
Wavenumbers (cm-1)
3425.95
3333.63
3107.64
2971.97
2869.48
2781.28
1916.09
1876.05
1736.16
1680.19
1612.63
1587.70
1571.82
1521.56
1468.49
1398.96
1379.14
1343.54
1321.06
1294.76
1267.66
1219.12
1164.81
1153.20
1132.84
1121.15
1078.01
1031.78
1013.83
942.38
920.06
901.15
837.94
764.59
736.44
682.69
660.41
611.15
594.60
575.95
548.45
512.21
482.60

1H NMR of 7d


CH3
CH3
N
N
N
N
S
O
NH
N
CH3
O

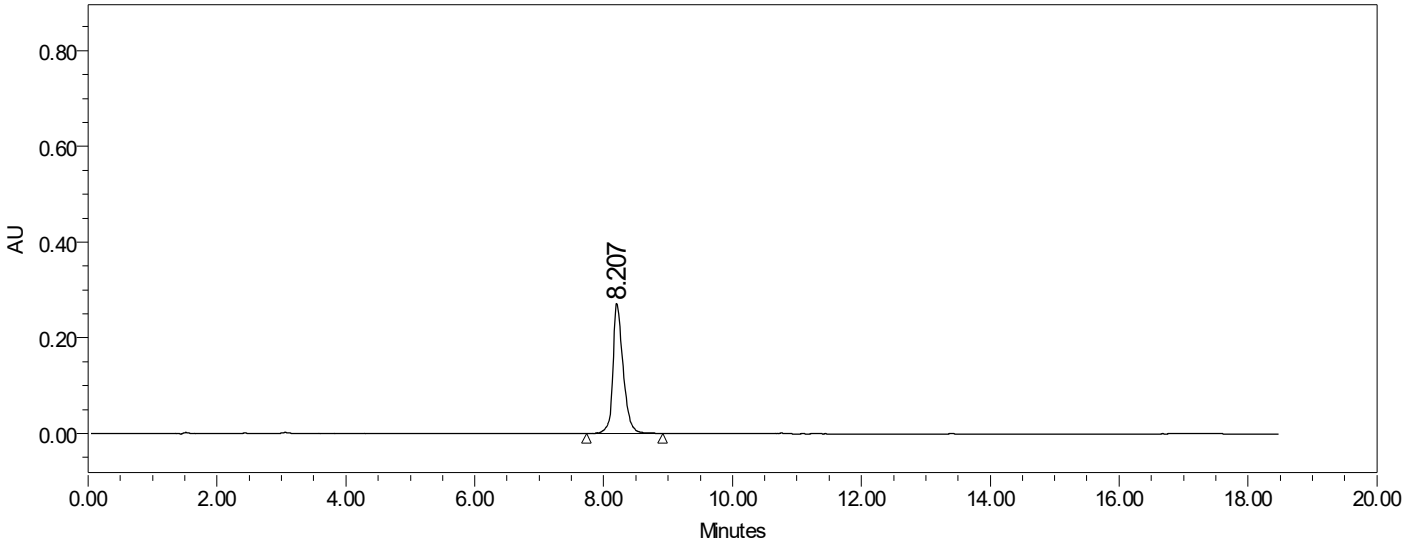
10.61
9.13
9.12
8.78
8.78
8.77
8.77
8.33
8.32
8.32
8.31
8.30
8.30
7.98
7.93
7.92
7.91
7.91
7.89
7.88
7.86
7.86
7.60
7.58
7.58
7.56
7.39
7.37
3.32 HDO
2.57
2.51 DMSO
2.50 DMSO
2.50 DMSO
2.49 DMSO
2.49 DMSO
2.40
2.38
1.00
1.07
1.05
1.09
6.24
1.16
2.20
3.26
3.26
3.20
f1 (ppm)

13C NMR of 7d


189.71
164.25
164.19
159.71
152.21
150.39
148.74
140.43
136.76
136.40
135.54
132.66
130.45
129.52
127.02
123.51
122.04
119.88
40.20 DMSO
40.15 DMSO
39.94 DMSO
39.73 DMSO
39.52 DMSO
39.31 DMSO
39.10 DMSO
38.90 DMSO
25.01
20.62
15.18
CH3
CH3
O
N
NH
N
N
N
N
S
CH3
O
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

SAMPLE INFORMATION

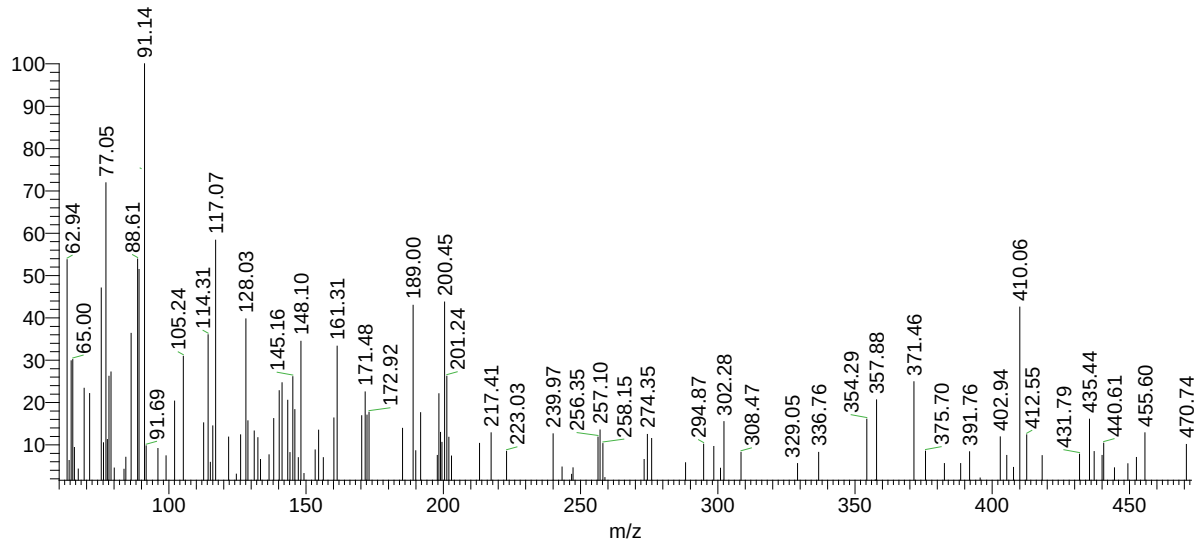
Sample Name:	WN-6	Acquired By:	YARA
Sample Type:	Unknown	Sample Set Name:	
Vial:	37	Acq. Method Set:	Organic
Injection #:	1	Processing Method:	Default
Injection Volume:	4.00 ul	Channel Name:	334.0nm
Run Time:	20.0 Minutes	Proc. Chnl. Descr.:	W2996 PDA 334.0 nm (PDA 190.0 to
Date Acquired:	22/05/2024 04:18:28 PM EET		
Date Processed:	23/05/2024 11:53:49 AM EET		

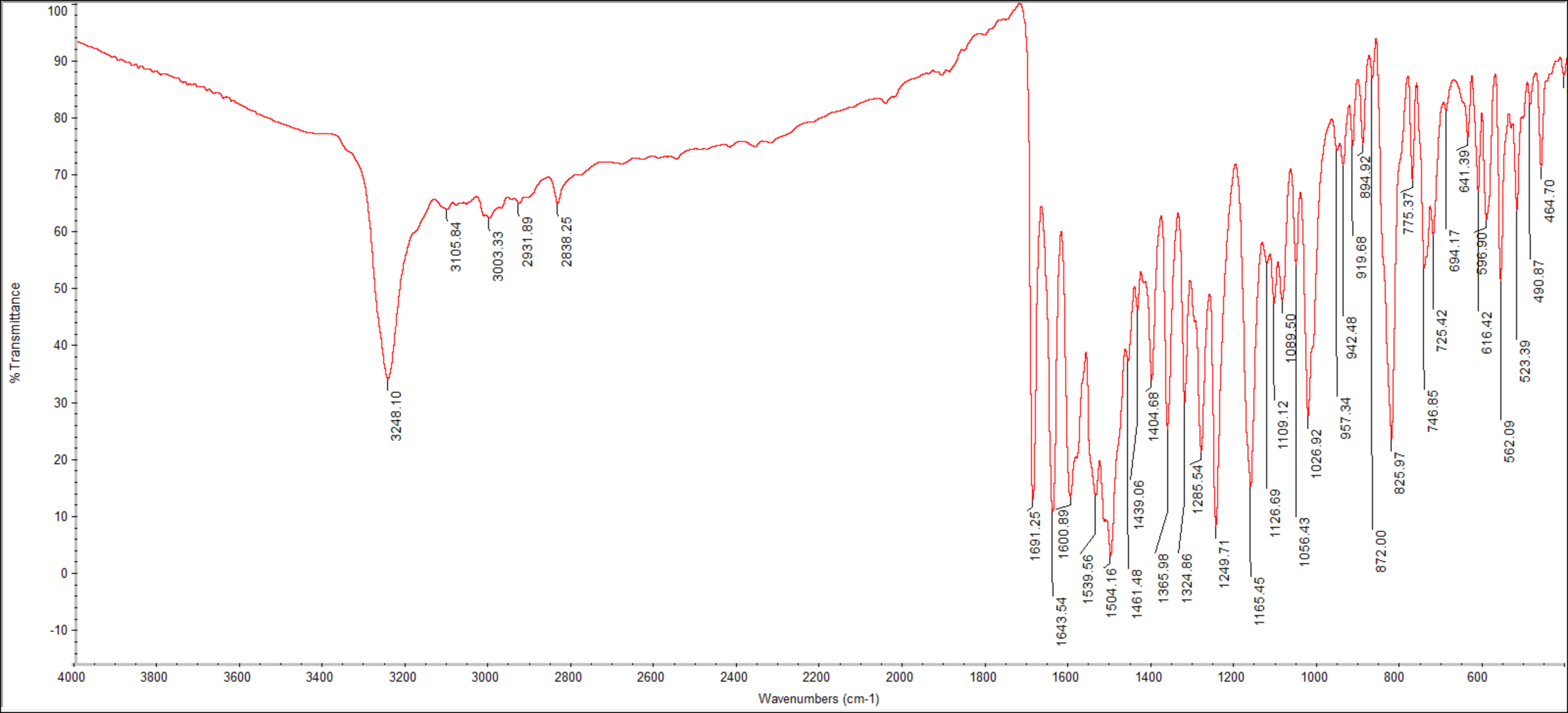


	RT	Area	% Area	Height
1	8.207	3010648	100.00	273113

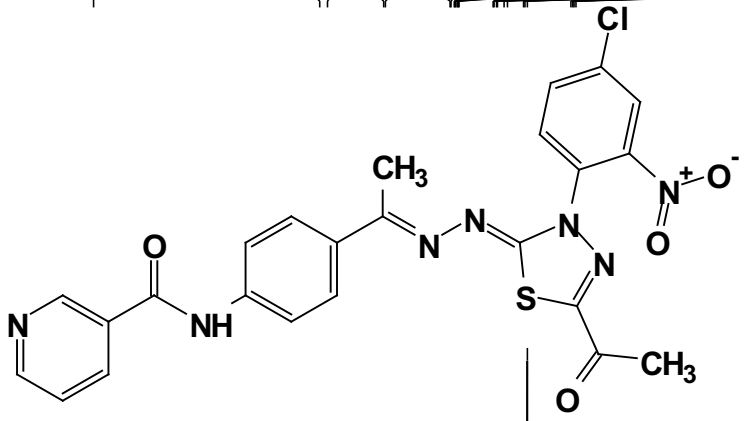
Mass spec. of compound 7d

wn6 #189 RT: 3.18 P: + NL: 1.56E3
T: {0,0} + c EI Full ms [40.00-1000.00]



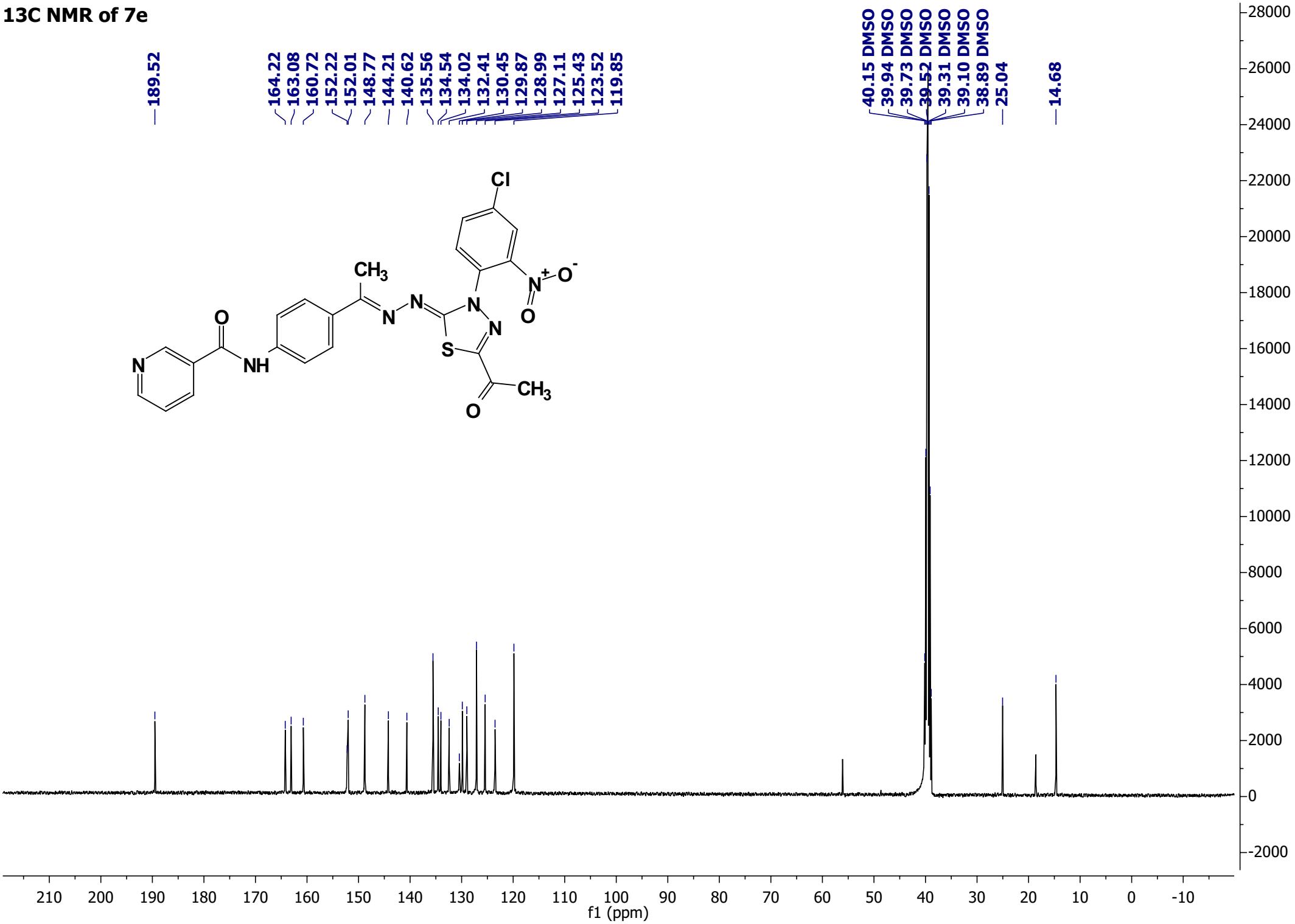

% Transmittance
Wavenumbers (cm-1)
100
90
80
70
60
50
40
30
20
10
0
-10
4000
3800
3600
3400
3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
3248.10
3105.84
3003.33
2931.89
2838.25
1691.25
1643.54
1600.89
1539.56
1504.16
1461.48
1439.06
1404.68
1365.98
1324.86
1285.54
1249.71
1165.45
1126.69
1109.12
1089.50
1056.43
1026.92
957.34
942.48
919.68
894.92
872.00
825.97
775.37
746.85
725.42
694.17
641.39
616.42
596.90
562.09
523.39
490.87
464.70

1H NMR of 7e

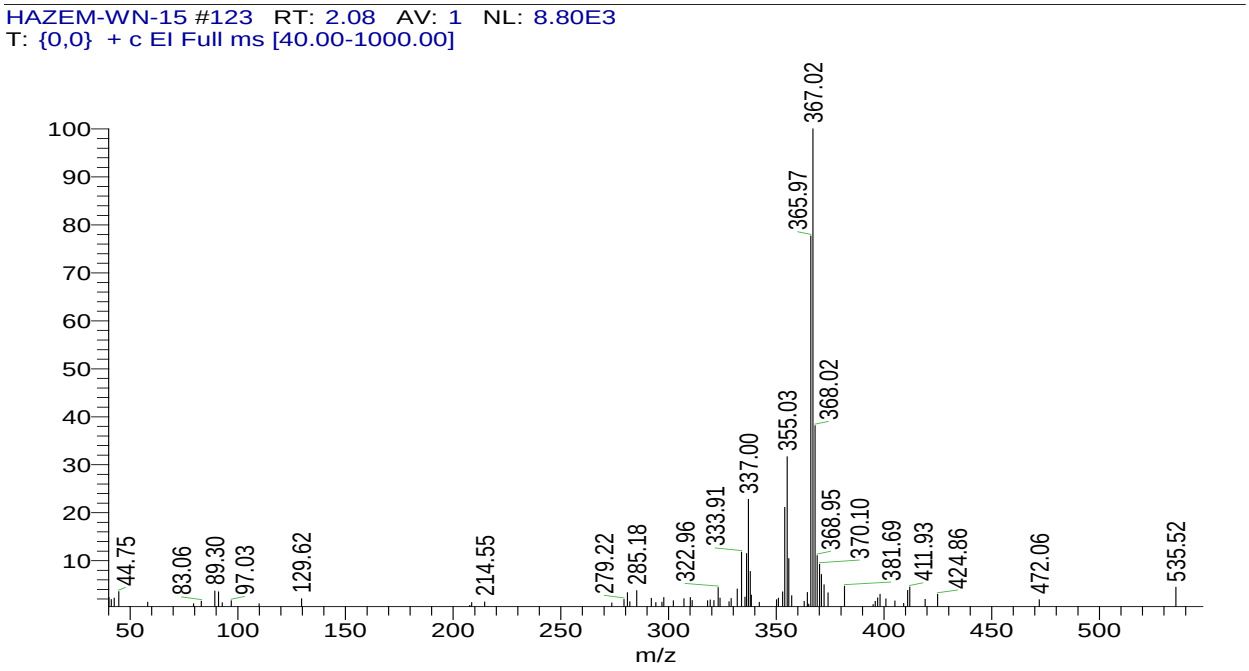

Cl
CH3
N+
O-
O
N
N
N
N
S
O
NH
N
CH3
O

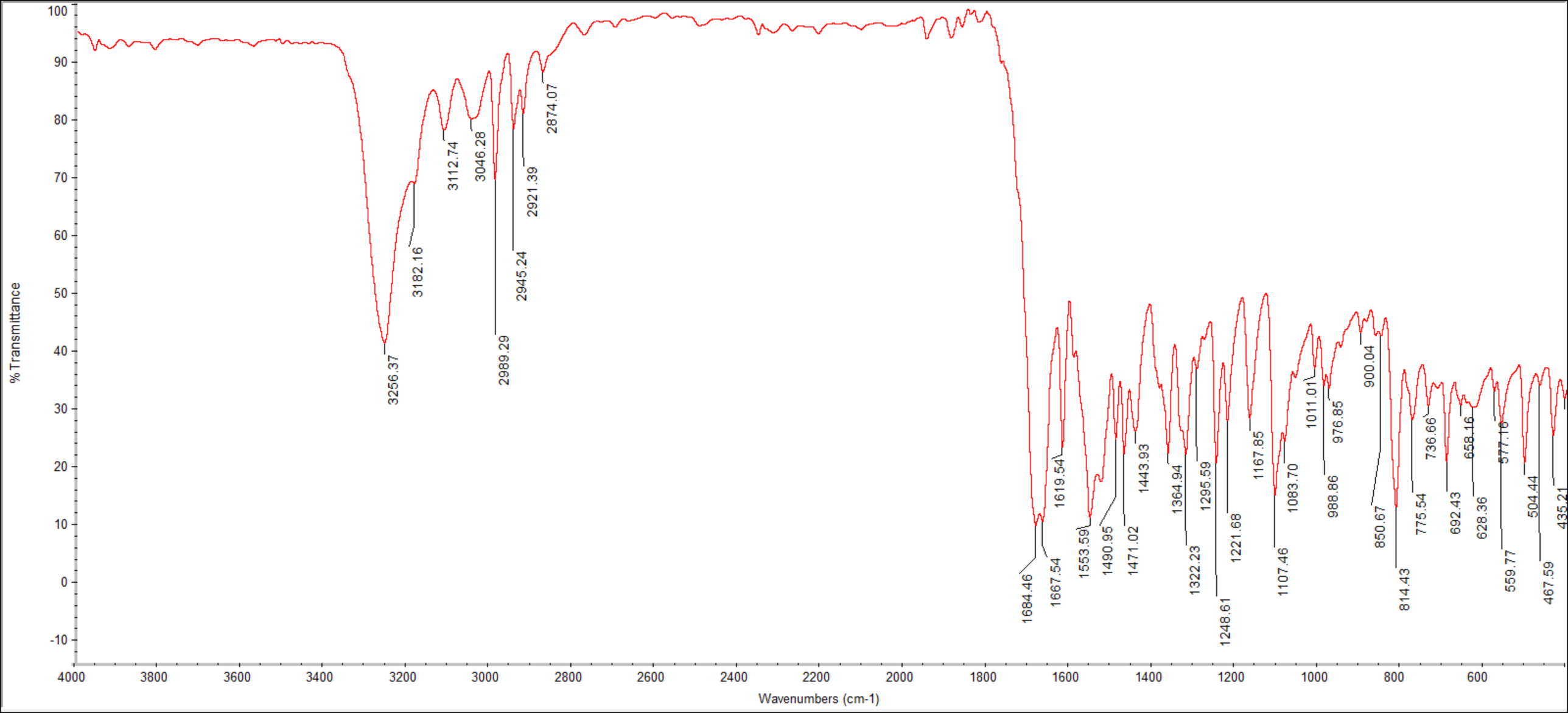
10.60
9.17
9.12
8.77
8.76
8.35
8.34
8.32
8.31
8.31
8.30
8.29
8.29
8.07
8.07
8.05
8.04
8.00
7.98
7.88
7.87
7.86
7.58
7.57
7.56
7.55
3.47 HDO
3.39 HDO
2.55
2.51 DMSO
2.50 DMSO
2.50 DMSO
2.50 DMSO
2.49 DMSO
2.23
2.20
1.00
1.03
1.02
1.03
1.03
1.15
0.96
4.31
1.05
3.18
3.23
f1 (ppm)

13C NMR of 7e

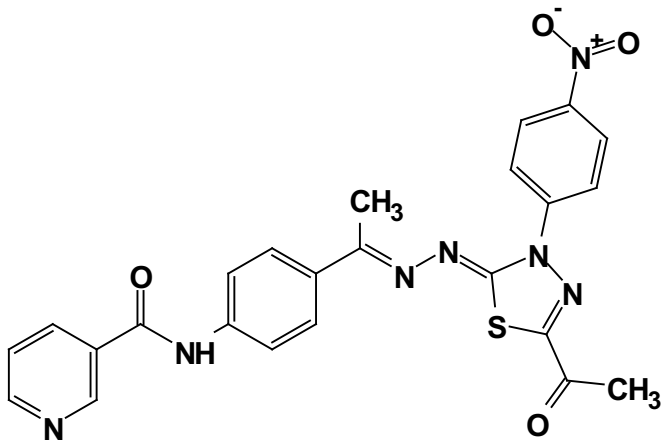

189.52
164.22
163.08
160.72
152.22
152.01
148.77
144.21
140.62
135.56
134.54
134.02
132.41
130.45
129.87
128.99
127.11
125.43
123.52
119.85
40.15 DMSO
39.94 DMSO
39.73 DMSO
39.52 DMSO
39.31 DMSO
39.10 DMSO
38.89 DMSO
25.04
14.68
Cl
CH3
N+
O-
O
N
NH
N
N
N
N
S
CH3
O
f1 (ppm)

Mass spec. of compound 7e




% Transmittance
Wavenumbers (cm-1)
3256.37
3182.16
3112.74
3046.28
2989.29
2945.24
2921.39
2874.07
1684.46
1667.54
1619.54
1553.59
1490.95
1471.02
1443.93
1364.94
1322.23
1295.59
1248.61
1221.68
1167.85
1107.46
1083.70
1011.01
988.86
976.85
900.04
850.67
814.43
775.54
736.66
692.43
658.16
628.36
577.16
559.77
504.44
467.59
435.21

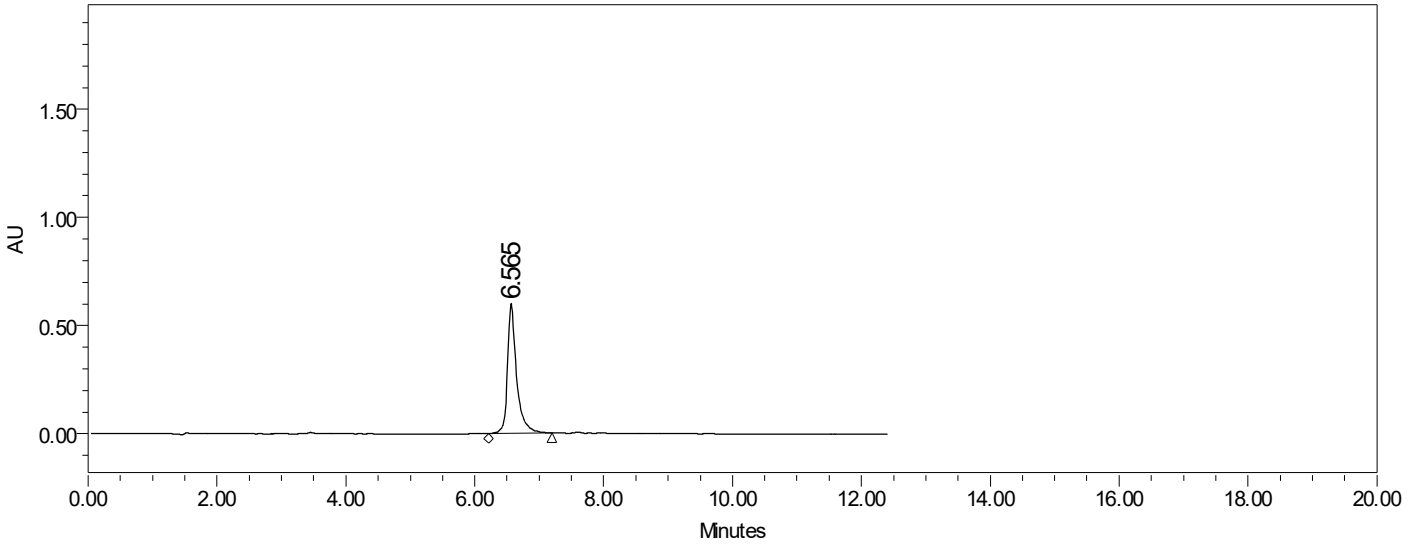
1H NMR of 7f


O
NH
N
CH3
N
N
N
N
S
O
CH3
O
N
O

10.66
9.15
8.81
8.80
8.48
8.33
7.95
7.93
7.61
3.36 HDO
3.34 HDO
2.66
2.60
2.52 DMSO
0.88
0.88
0.76
3.94
0.93
4.22
0.84
4.31
3.17
f1 (ppm)

SAMPLE INFORMATION

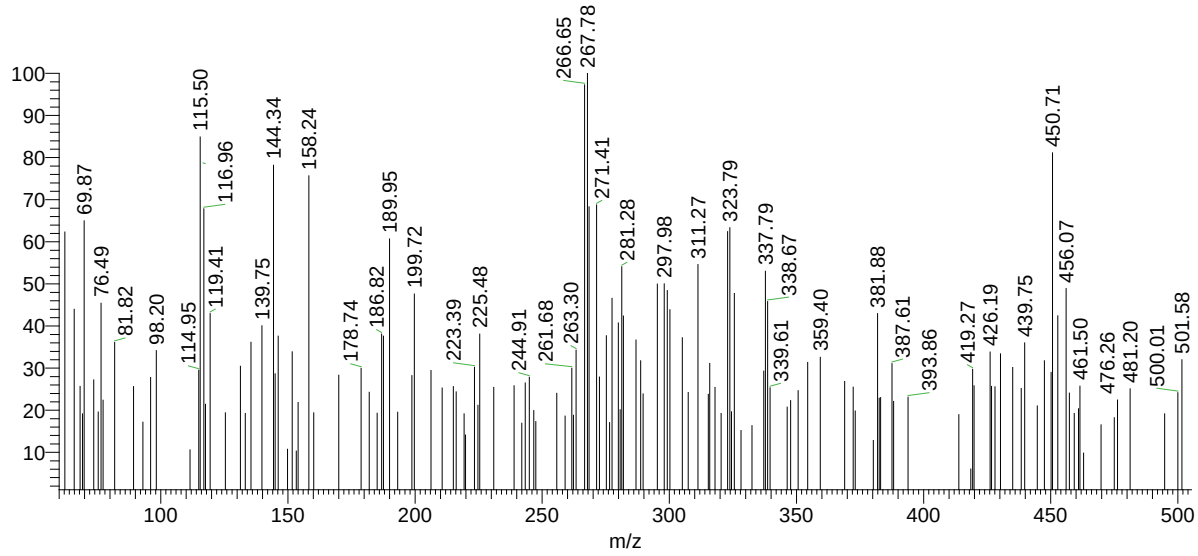
Sample Name:	WN-17	Acquired By:	YARA
Sample Type:	Unknown	Sample Set Name:	
Vial:	43	Acq. Method Set:	Organic
Injection #:	1	Processing Method:	Default
Injection Volume:	4.00 ul	Channel Name:	373.0nm
Run Time:	20.0 Minutes	Proc. Chnl. Descr.:	W2996 PDA 373.0 nm (PDA 190.0 to
Date Acquired:	22/05/2024 06:41:06 PM EET		
Date Processed:	23/05/2024 11:51:39 AM EET		

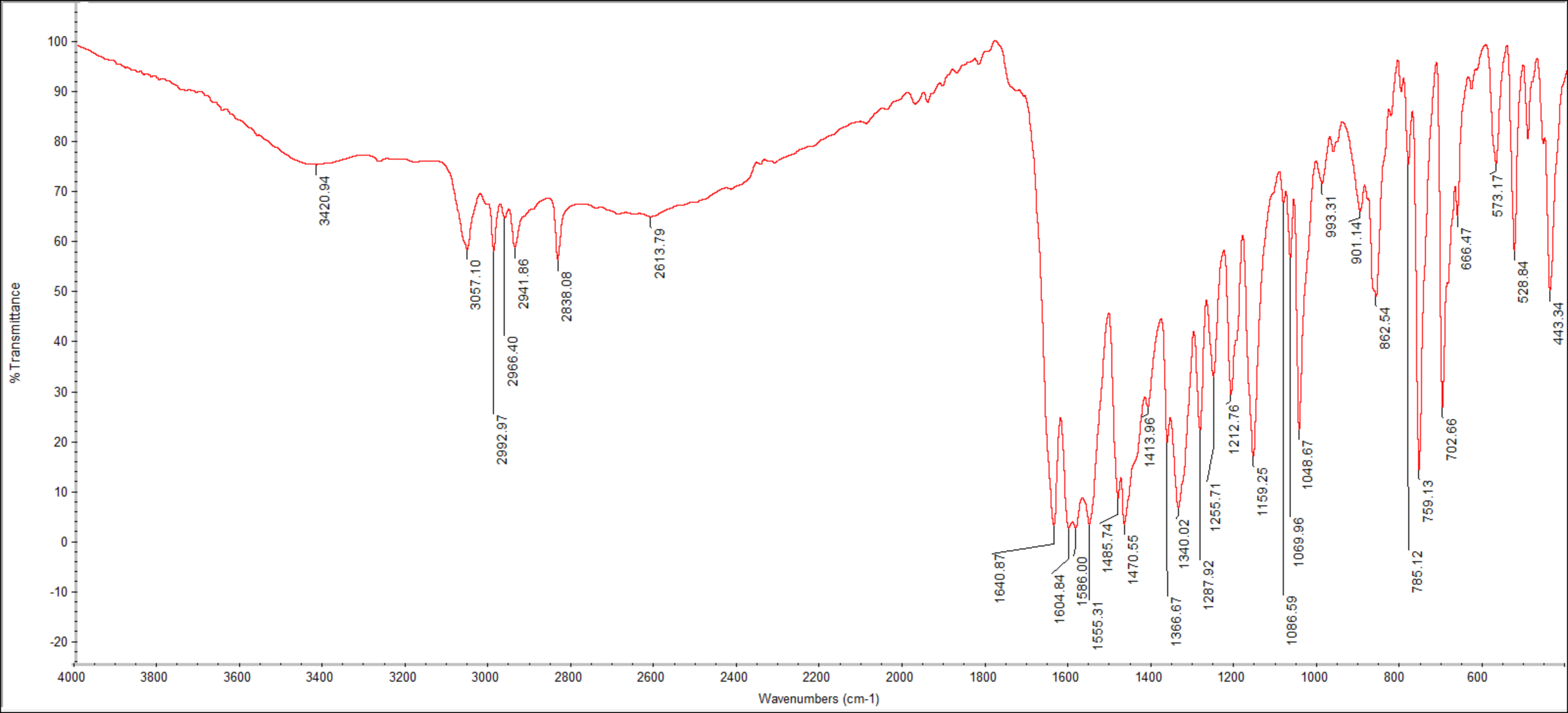


	RT	Area	% Area	Height
1	6.565	5907342	100.00	597572

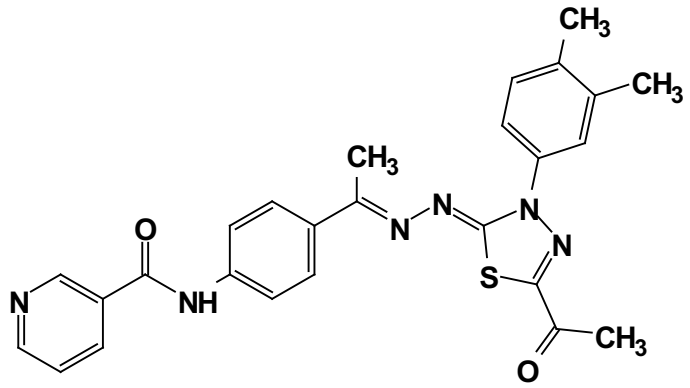
Mass spec. of compound 7f

wn17 #217 RT: 3.65 P: + NL: 3.79E2
T: {0,0} + c EI Full ms [40.00-1000.00]




% Transmittance
Wavenumbers (cm-1)
3420.94
3057.10
2992.97
2966.40
2941.86
2838.08
2613.79
1640.87
1604.84
1586.00
1555.31
1485.74
1470.55
1413.96
1366.67
1340.02
1287.92
1255.71
1212.76
1159.25
1086.59
1069.96
1048.67
993.31
901.14
862.54
785.12
759.13
702.66
666.47
573.17
528.84
443.34

1H NMR of 7g

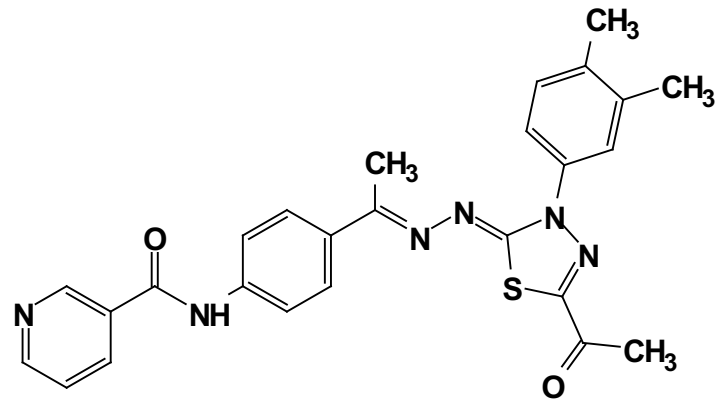


10.60
9.12 9.12 8.78 8.77 8.76 8.76 8.32 8.32 8.31 8.30 8.30 8.29 7.89 7.86 7.82 7.81 7.74 7.74 7.72 7.72 7.59 7.57 7.57 7.55 7.30 7.27
3.41 HDO 3.37 HDO 2.54 2.51 DMSO 2.51 DMSO 2.50 DMSO 2.50 DMSO 2.49 DMSO 2.37 2.29 2.26

1.00 1.01 1.06 1.10 4.37 1.17 1.18 1.14 1.09 3.24 3.35 3.38 3.15

f1 (ppm)

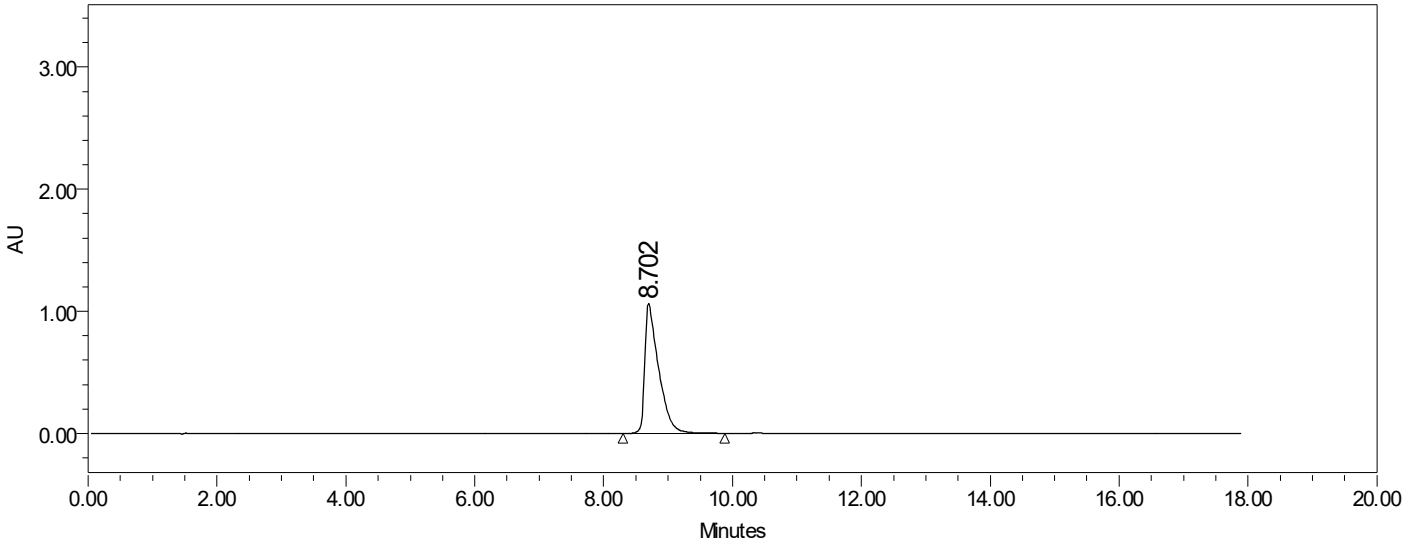
13C NMR of 7g


CH3
CH3
CH3
N
N
N
N
S
O
NH
N
CH3
O

189.68
164.24
164.18
159.59
152.22
150.22
148.75
140.42
137.06
136.63
135.55
135.43
132.67
130.45
129.85
126.99
123.52
122.85
119.85
119.35
40.15 DMSO
39.94 DMSO
39.73 DMSO
39.52 DMSO
39.31 DMSO
39.10 DMSO
38.89 DMSO
24.97
19.68
19.00
18.57
15.10
22000
21000
20000
19000
18000
17000
16000
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
-2000
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

SAMPLE INFORMATION

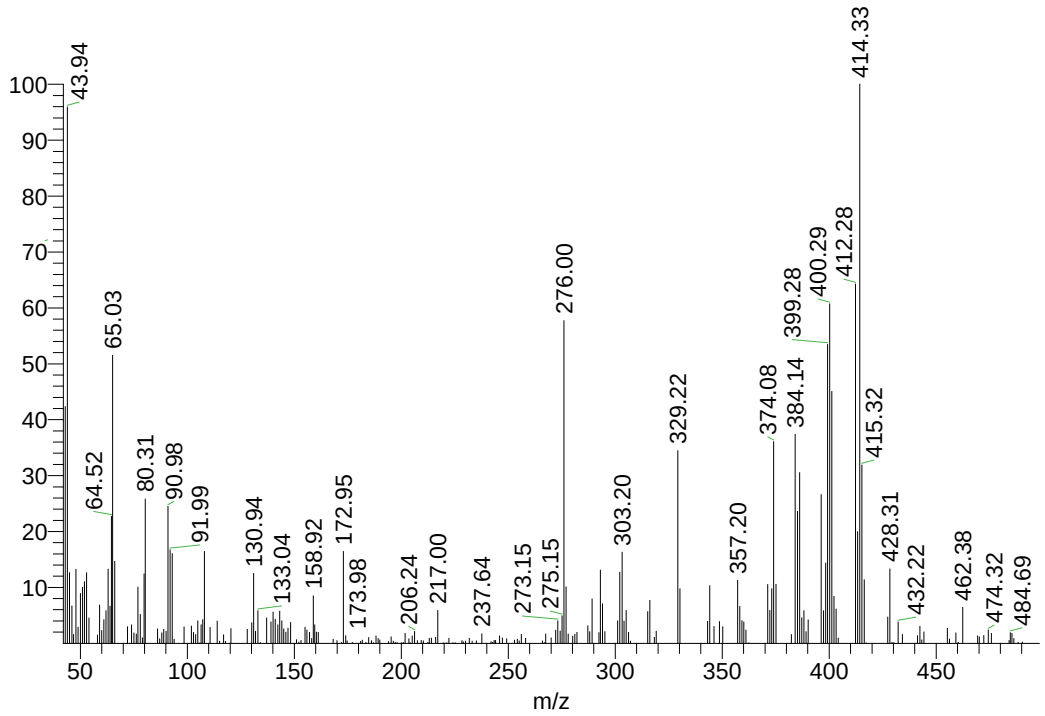
Sample Name:	WN-16	Acquired By:	YARA
Sample Type:	Unknown	Sample Set Name:	
Vial:	42	Acq. Method Set:	Organic
Injection #:	1	Processing Method:	Default
Injection Volume:	4.00 ul	Channel Name:	335.0nm
Run Time:	20.0 Minutes	Proc. Chnl. Descr.:	W2996 PDA 335.0 nm (PDA 190.0 to
Date Acquired:	22/05/2024 06:15:07 PM EET		
Date Processed:	23/05/2024 11:51:48 AM EET		

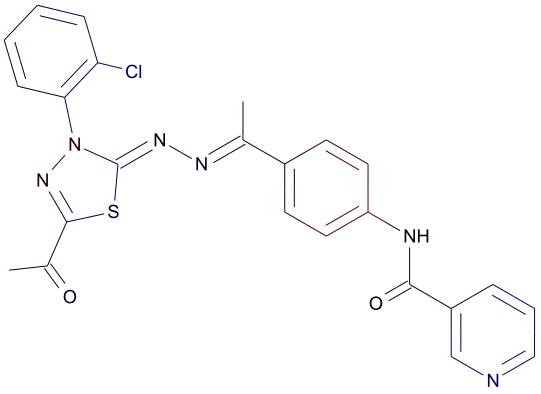


	RT	Area	% Area	Height
1	8.702	15883187	100.00	1063707

Mass spec. of compound 7g

Hazem-Wn-16 #285 RT: 4.79 AV: 1 SB: 2 4.25 , 3.88 NL: 8.11E4
T: {0,0} + c EI Full ms [40.00-1000.00]





$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.607

Enrichment: 1.09

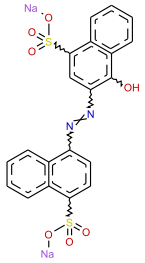
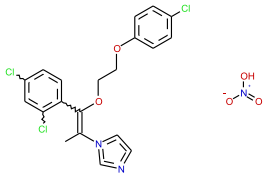
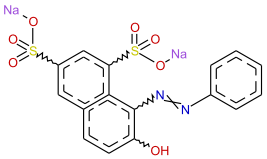
Bayesian Score: -4.84

Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 2.47e-005

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

Structural Similar Compounds

Name	3567-69-9	83621-06-1	1936-15-8
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.574	0.608	0.612
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411

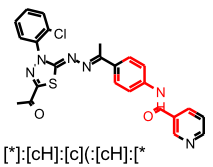
Model Applicability

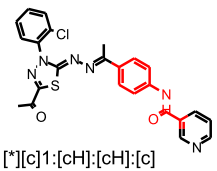
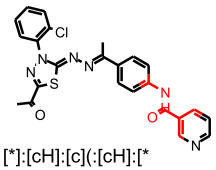
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

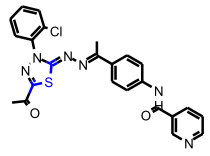
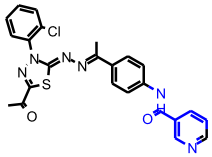
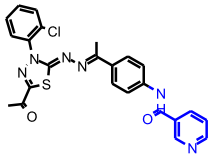
Feature Contribution

Top features for positive contribution

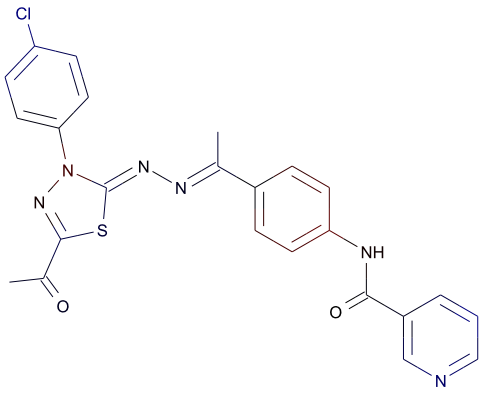
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	555539852	 [*:[cH]:[c](:[cH]:[*])C(=O)N[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.447	22 out of 24

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:n:[cH]:1	-0.762	0 out of 2
SCFP_12	1165971455	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.762	0 out of 2

TOPKAT_Ames_Mutagenicity



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.605

Enrichment: 1.08

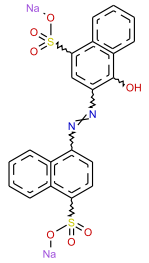
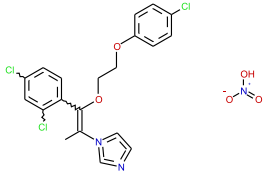
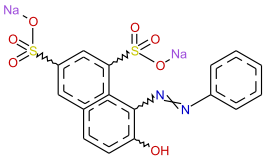
Bayesian Score: -4.88

Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 2.47e-005

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

Structural Similar Compounds

Name	3567-69-9	83621-06-1	1936-15-8
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.577	0.603	0.615
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411

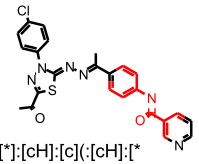
Model Applicability

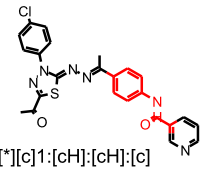
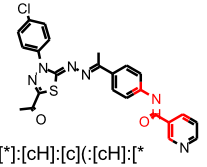
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

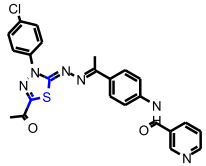
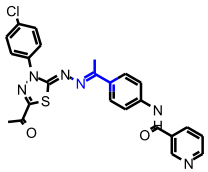
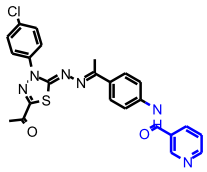
Feature Contribution

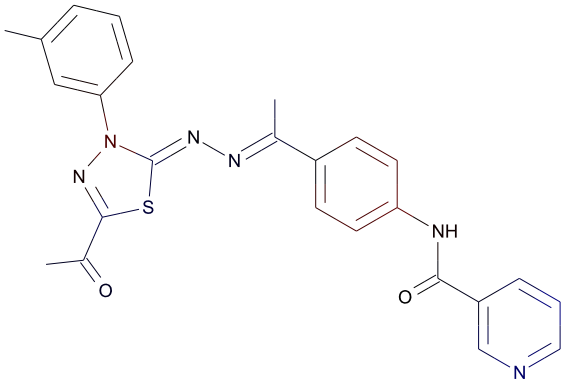
Top features for positive contribution

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

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	1165971455	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.762	0 out of 2



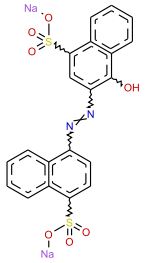
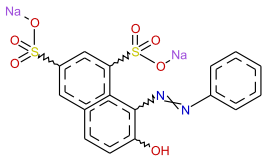
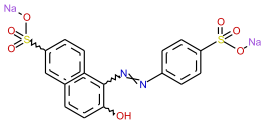
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.674
Enrichment: 1.21
Bayesian Score: -2.75
Mahalanobis Distance: 12.4
Mahalanobis Distance p-value: 9.97e-005

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

Structural Similar Compounds

Name	3567-69-9	1936-15-8	2783-94-0
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.590	0.598	0.600
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411

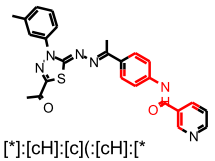
Model Applicability

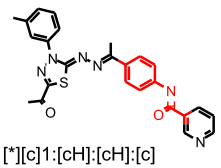
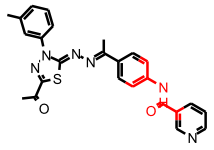
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

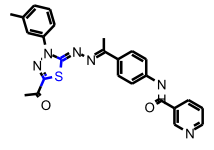
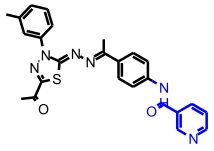
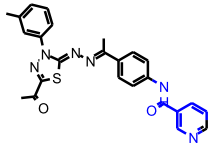
Feature Contribution

Top features for positive contribution

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

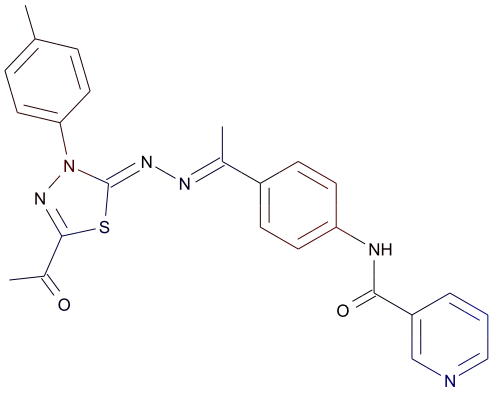
SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	2096901122	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	0.429	33 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	1165971455	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.762	0 out of 2
SCFP_12	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:n:[cH]:1	-0.762	0 out of 2

7d

TOPKAT_Ames_Mutagenicity



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.685

Enrichment: 1.23

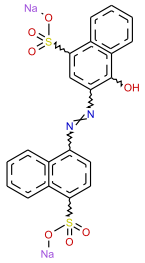
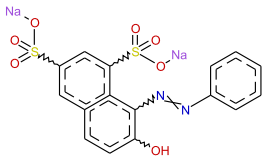
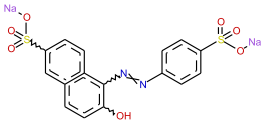
Bayesian Score: -2.35

Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 9.97e-005

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

Structural Similar Compounds

Name	3567-69-9	1936-15-8	2783-94-0
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.593	0.600	0.600
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411

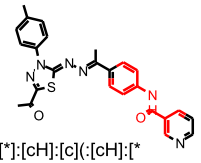
Model Applicability

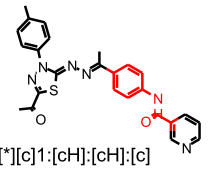
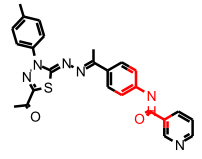
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

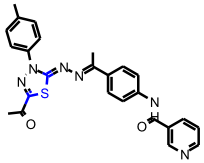
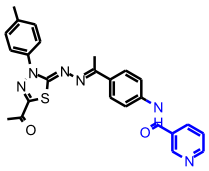
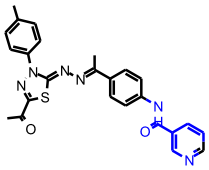
Feature Contribution

Top features for positive contribution

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

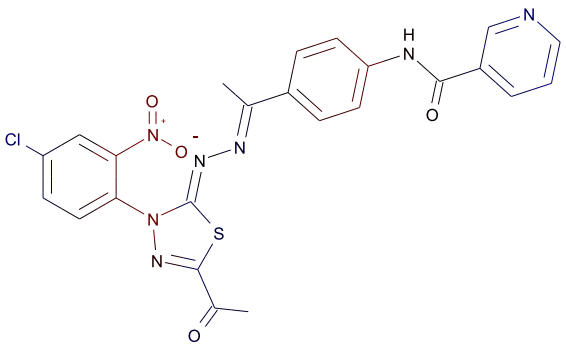
SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13
SCFP_12	2096901122	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	0.429	33 out of 37

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	1165971455	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.762	0 out of 2
SCFP_12	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:n:[cH]:1	-0.762	0 out of 2

7e

TOPKAT_Ames_Mutagenicity



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.688

Enrichment: 1.23

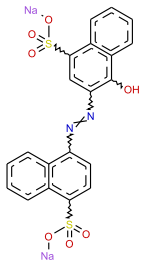
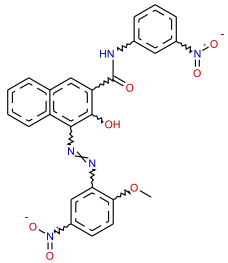
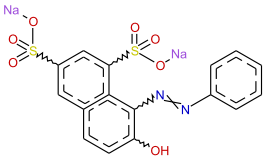
Bayesian Score: -2.21

Mahalanobis Distance: 12.9

Mahalanobis Distance p-value: 8.03e-006

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

Structural Similar Compounds

Name	3567-69-9	6471-49-4	1936-15-8
Structure			
Actual Endpoint	Mutagen	Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Distance	0.599	0.632	0.636
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411

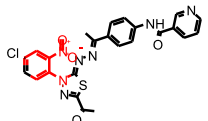
Model Applicability

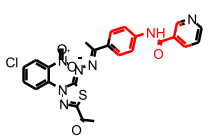
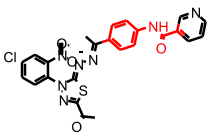
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

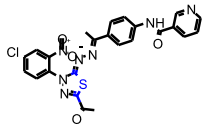
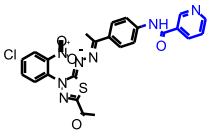
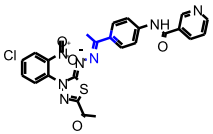
Feature Contribution

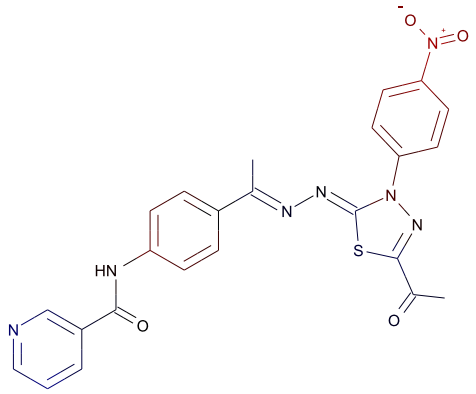
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-990000384	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[c]:1[N+](=O)[O-]	0.507	50 out of 52

SCFP_12	555539852	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.447	22 out of 24
SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	1165971455	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[cH]:n:[cH]:1	-0.762	0 out of 2
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2



$C_{24}H_{19}N_7O_4S$

Molecular Weight: 501.51715

ALogP: 3.488

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Mutagen

Probability: 0.744

Enrichment: 1.33

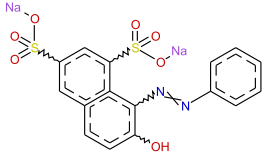
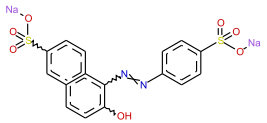
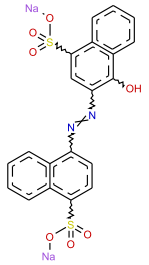
Bayesian Score: 0.275

Mahalanobis Distance: 12.5

Mahalanobis Distance p-value: 5.91e-005

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

Structural Similar Compounds

Name	1936-15-8	2783-94-0	3567-69-9
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.599	0.599	0.632
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411

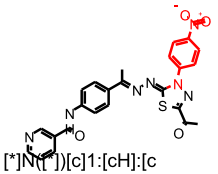
Model Applicability

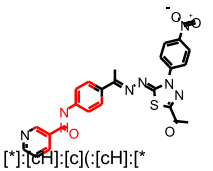
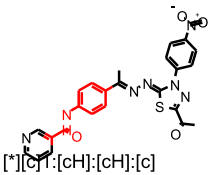
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

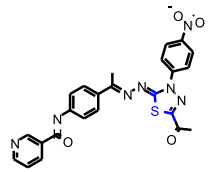
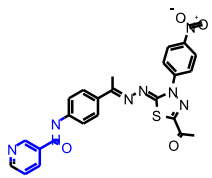
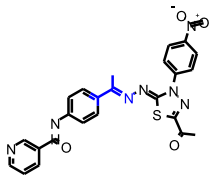
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1729573318	 [*]N([*])[c]1:[cH]:[cH]:[c]([cH]:[cH]:1)[N+](=O)[O-]	0.488	36 out of 38

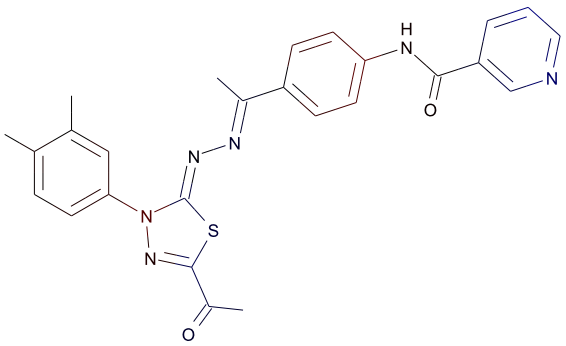
SCFP_12	555539852	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c]1:[cH]:[cH]:[cH]:[*]:[cH]:[cH]:1	0.447	22 out of 24
SCFP_12	818445224	 [*][c]1:[cH]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[cH]:1	0.434	12 out of 13

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:n:[cH]:1	-0.762	0 out of 2
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2

7g

TOPKAT_Ames_Mutagenicity



$C_{26}H_{24}N_6O_2S$

Molecular Weight: 484.57275

ALogP: 4.566

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.683

Enrichment: 1.22

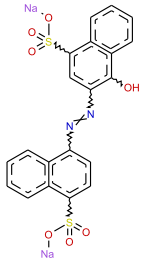
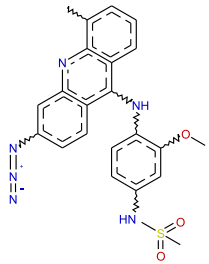
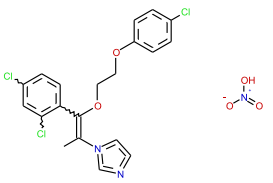
Bayesian Score: -2.42

Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 9.65e-005

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

Structural Similar Compounds

Name	3567-69-9	110004-69-8	83621-06-1
Structure			
Actual Endpoint	Mutagen	Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Distance	0.581	0.621	0.622
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

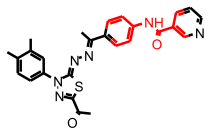
Model Applicability

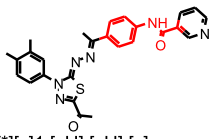
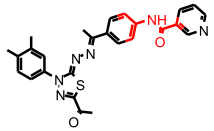
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

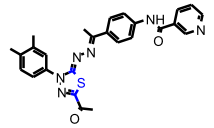
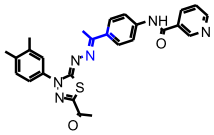
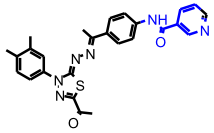
Feature Contribution

Top features for positive contribution

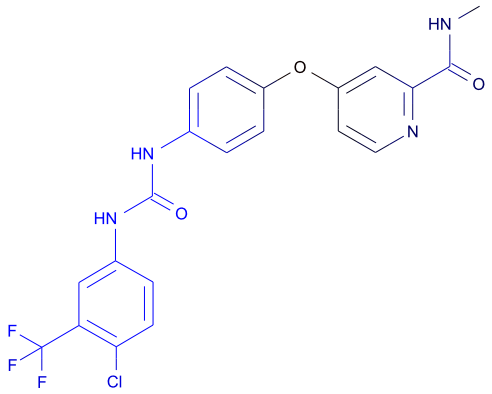
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	555539852	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.447	22 out of 24

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1630519606	 [*]C1=[*][*]C(=[*])S1	-0.998	0 out of 3
SCFP_12	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.762	0 out of 2
SCFP_12	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[c]H]:[*]:n:[cH]:1	-0.762	0 out of 2

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

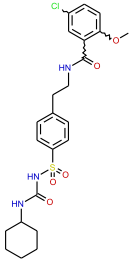
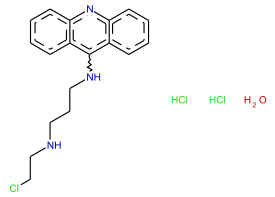
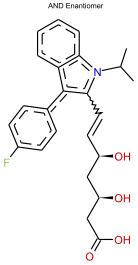
Rotatable Bonds: 6

Acceptors: 4

Donors: 3

TOPKAT_Ames_Mutagenicity

Structural Similar Compounds

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

Model Prediction

Prediction: Non-Mutagen

Probability: 0.0531

Enrichment: 0.0951

Bayesian Score: -19.7

Mahalanobis Distance: 13.1

Mahalanobis Distance p-value: 2.73e-006

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

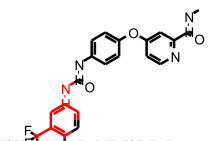
Model Applicability

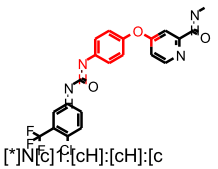
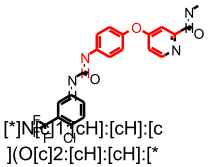
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

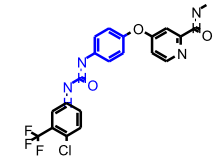
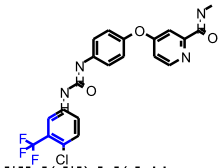
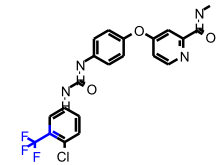
Feature Contribution

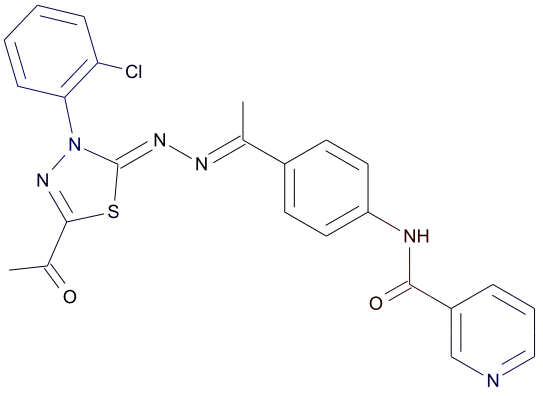
Top features for positive contribution

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

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

Top Features for negative contribution

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



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.42

Enrichment: 0.798

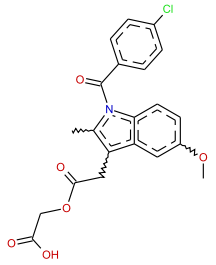
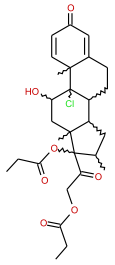
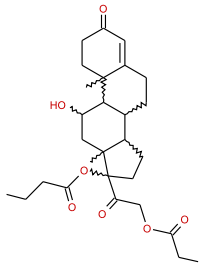
Bayesian Score: -3.74

Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.0037

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

Structural Similar Compounds

Name	Acemetacin	Beclomethasone Dipropionate	Hydrocortisone-17-butyrate-21-propionate
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.632	0.646	0.687
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 21:441-466; 1981

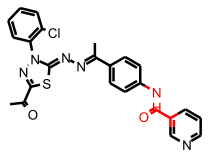
Model Applicability

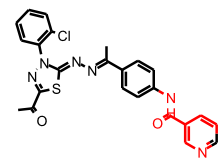
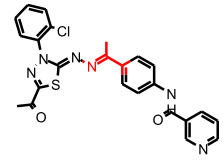
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

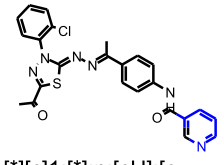
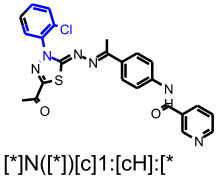
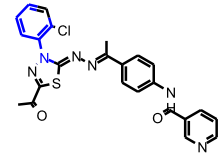
Feature Contribution

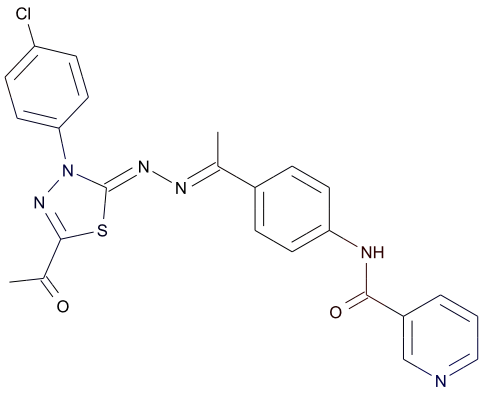
Top features for positive contribution

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

SCFP_6	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:n:[cH]:1	0.271	1 out of 1
SCFP_6	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:[cH]:1	-0.646	2 out of 9
SCFP_6	1557430596	 [*]N([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	-0.594	1 out of 5
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-0.449	6 out of 19



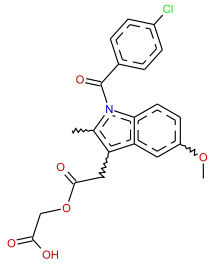
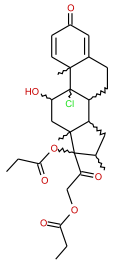
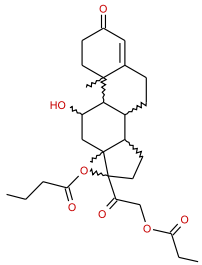
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96465
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.463
Enrichment: 0.88
Bayesian Score: -2.37
Mahalanobis Distance: 10.7
Mahalanobis Distance p-value: 0.00263

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

Structural Similar Compounds

Name	Acemetacin	Beclomethasone Dipropionate	Hydrocortisone-17-butyrate-21-propionate
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.622	0.645	0.686
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 21:441-466; 1981

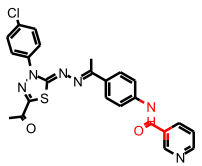
Model Applicability

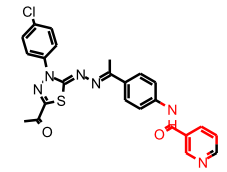
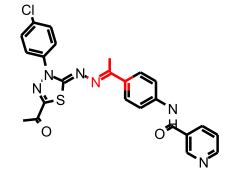
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

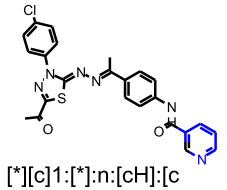
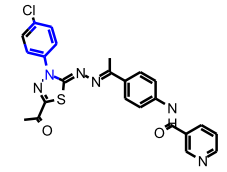
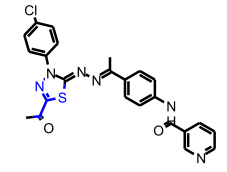
Feature Contribution

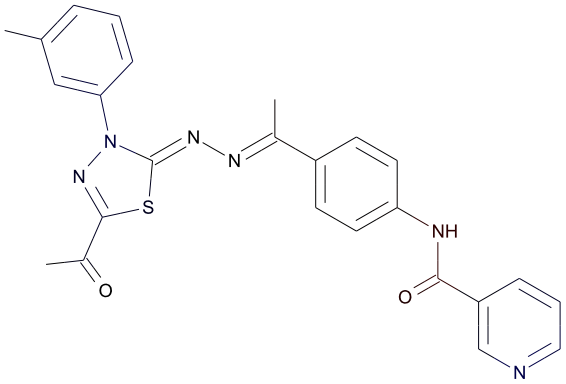
Top features for positive contribution

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

SCFP_6	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:n:[cH]:1	0.271	1 out of 1
SCFP_6	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.646	2 out of 9
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-0.449	6 out of 19
SCFP_6	-704135030	 [*]C(=[*])C1=N[*][*]S1	-0.422	0 out of 1



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.439

Enrichment: 0.834

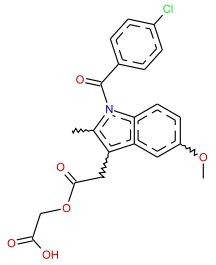
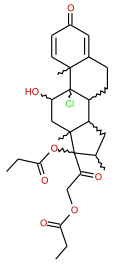
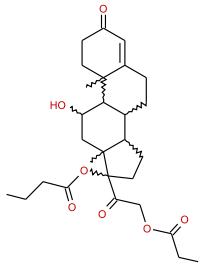
Bayesian Score: -3.12

Mahalanobis Distance: 10.6

Mahalanobis Distance p-value: 0.00398

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

Structural Similar Compounds

Name	Acemetacin	Beclomethasone Dipropionate	Hydrocortisone-17-butyrate-21-propionate
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.625	0.656	0.685
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 21:441-466; 1981

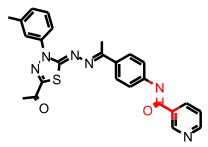
Model Applicability

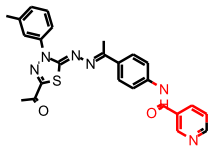
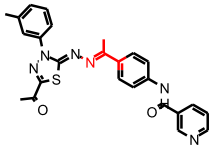
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: 4.5388. Training min, max, SD, explained variance: -3.5766, 3.955, 1.214, 0.0216.

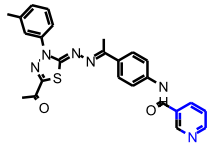
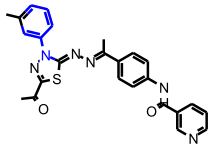
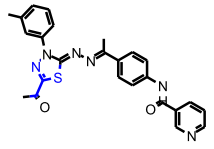
Feature Contribution

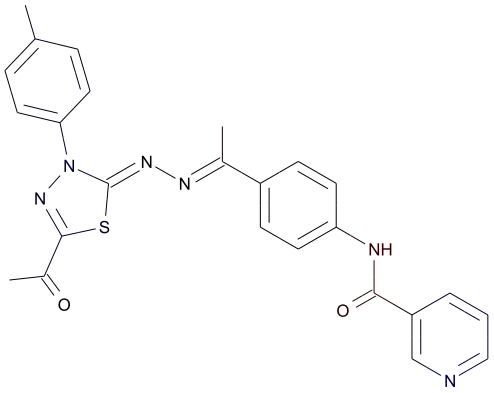
Top features for positive contribution

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

SCFP_6	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:n:[cH]:1	0.271	1 out of 1
SCFP_6	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:[cH]:1	-0.646	2 out of 9
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-0.449	6 out of 19
SCFP_6	-704135030	 [*]C(=[*])C1=N[*][*]S1	-0.422	0 out of 1



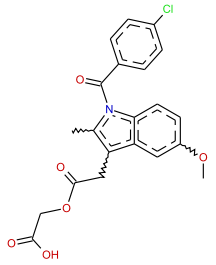
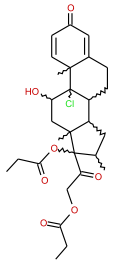
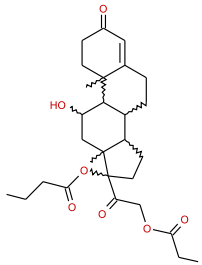
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.456
Enrichment: 0.868
Bayesian Score: -2.57
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.00091

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

Structural Similar Compounds

Name	Acemetacin	Beclomethasone Dipropionate	Hydrocortisone-17-butyrate-21-propionate
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.624	0.656	0.685
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 21:441-466; 1981

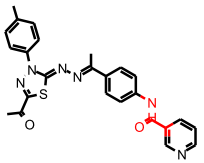
Model Applicability

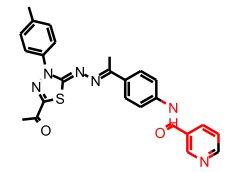
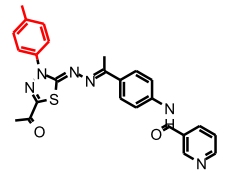
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: 4.4401. Training min, max, SD, explained variance: -3.5766, 3.955, 1.214, 0.0216.

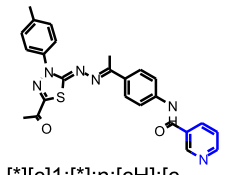
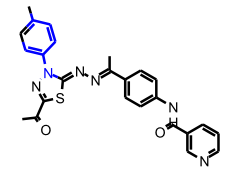
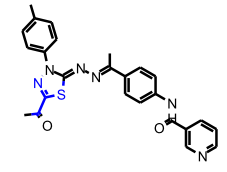
Feature Contribution

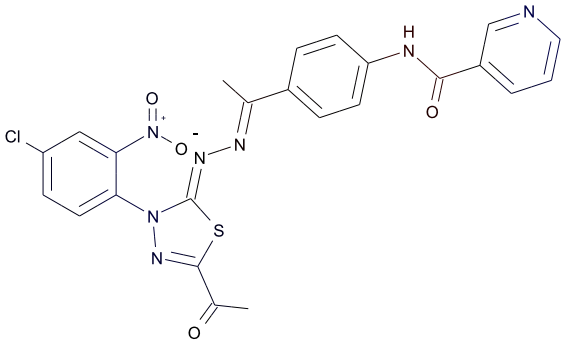
Top features for positive contribution

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

SCFP_6	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:n:[cH]:1	0.271	1 out of 1
SCFP_6	795925860	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.646	2 out of 9
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-0.449	6 out of 19
SCFP_6	-704135030	 [*]C(=[*])C1=N[*][*]S1	-0.422	0 out of 1



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.422

Enrichment: 0.803

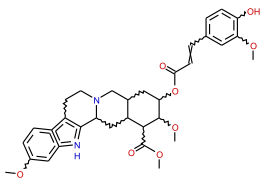
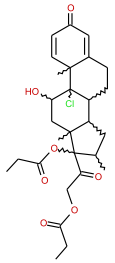
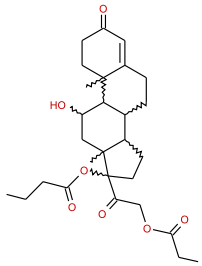
Bayesian Score: -3.65

Mahalanobis Distance: 11.2

Mahalanobis Distance p-value: 0.000497

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

Structural Similar Compounds

Name	Reserpate	Beclomethasone Dipropionate	Hydrocortisone-17-butyrate-21-propionate
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.736	0.756	0.775
Reference	Oyo Yakuri 18:105-124; 1979	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 21:441-466; 1981

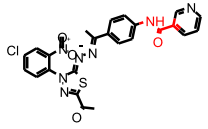
Model Applicability

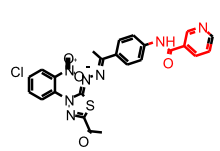
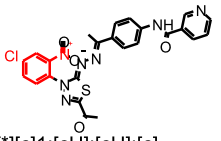
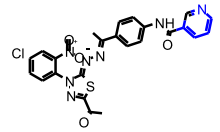
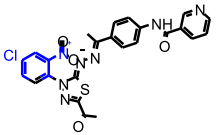
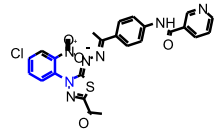
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

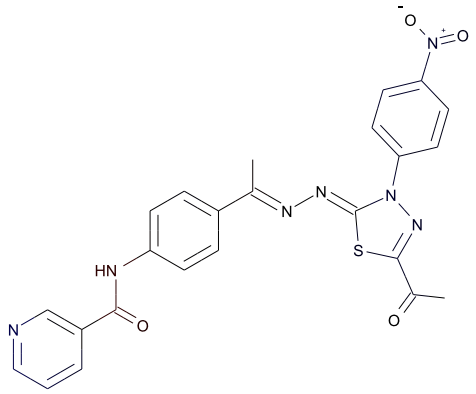
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

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

SCFP_6	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[cH]:[*]:n:[cH]:1	0.271	1 out of 1
SCFP_6	1907765814	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[c]:1[N+](=[*])[*]	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	1807097289	 [*][c]1:[*]:[cH]:[c](Cl):[cH]:[c]:1[N+](=[*])[*]	-0.594	1 out of 5
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-0.449	6 out of 19



$C_{24}H_{19}N_7O_4S$

Molecular Weight: 501.51715

ALogP: 3.488

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.44

Enrichment: 0.836

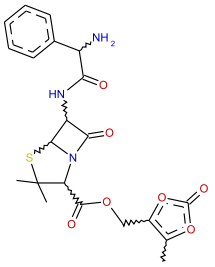
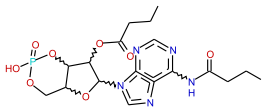
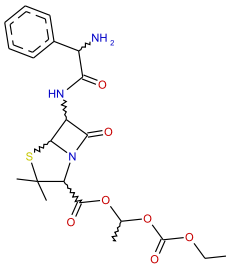
Bayesian Score: -3.09

Mahalanobis Distance: 10.8

Mahalanobis Distance p-value: 0.00224

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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)	Cyclic AMP Bucladesine	Bacampicillin .HCl (Free base form)
Structure			
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Non-Toxic
Distance	0.758	0.759	0.770
Reference	Chemotherapy 32:130-145; 1984	Oyo Yakuri 27(3):585-597; 1984	Chemotherapy 27:30-35; 1979

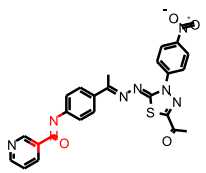
Model Applicability

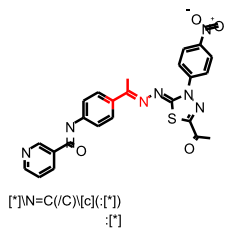
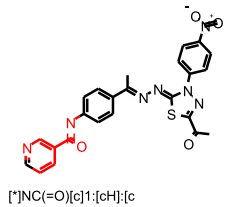
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

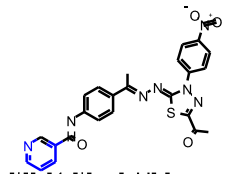
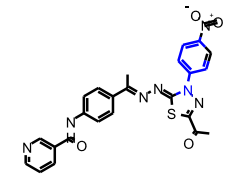
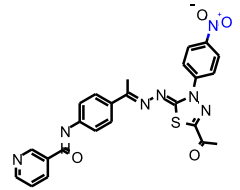
Feature Contribution

Top features for positive contribution

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

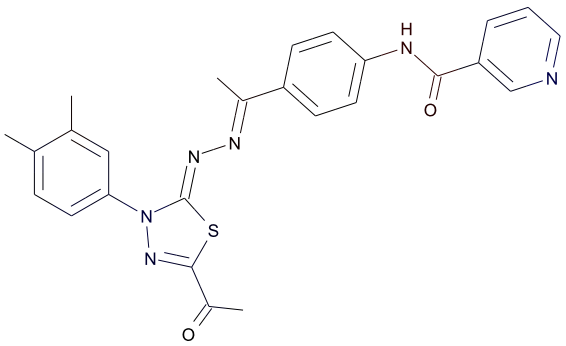
SCFP_6	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	0.271	1 out of 1
SCFP_6	903335088	 [*]NC(=O)[c]1:[cH]:[cH]:[*]:n:[cH]:1	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.646	2 out of 9
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[*]:[c]:1[*]	-0.449	6 out of 19
SCFP_6	1311339974	 [*][N+](=O)[*]	-0.446	3 out of 10

7g

TOPKAT_Developmental_Toxicity_Potential



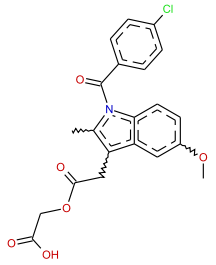
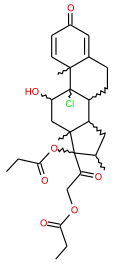
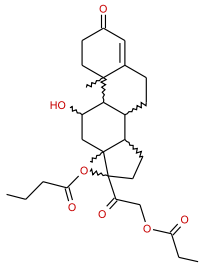
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.456
Enrichment: 0.868
Bayesian Score: -2.57
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.000762

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

Structural Similar Compounds

Name	Acemetacin	Beclomethasone Dipropionate	Hydrocortisone-17-butyrate-21-propionate
Structure			
Actual Endpoint	Non-Toxic	Toxic	Toxic
Predicted Endpoint	Non-Toxic	Toxic	Toxic
Distance	0.639	0.649	0.689
Reference	Oyo Yakuri 22(6):777-786; 1981	Oyo Yakuri 18(6):1021-1038; 1979	Oyo Yakuri 21:441-466; 1981

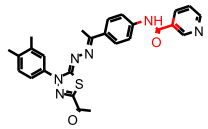
Model Applicability

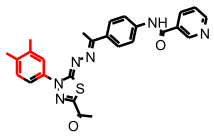
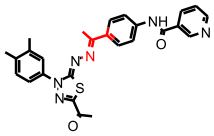
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: 4.4679. Training min, max, SD, explained variance: -3.5766, 3.955, 1.214, 0.0216.

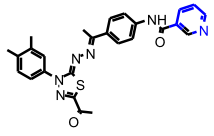
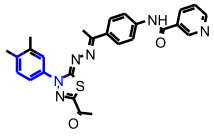
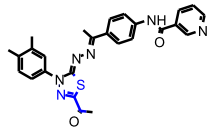
Feature Contribution

Top features for positive contribution

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

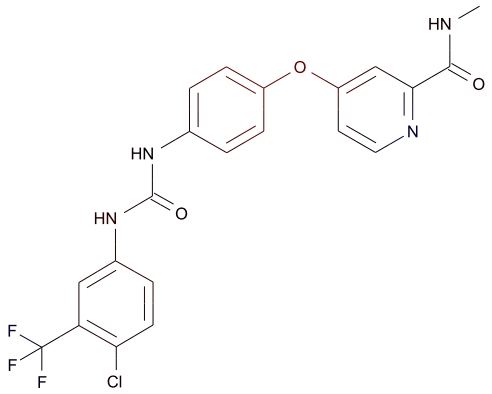
SCFP_6	1789155233	 [*][c]1:[*]:[cH]:[c](C):[c](C):[cH]:1	0.271	1 out of 1
SCFP_6	-331724199	 [*]\N=C(/C)\[c](:[*]) :[*]	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[c H]:[cH]:1	-0.646	2 out of 9
SCFP_6	-1380909229	 [*]N([*])[c]1:[cH]:[c H]:[cH]:[*]:[c]:1[*]	-0.449	6 out of 19
SCFP_6	-704135030	 [*]C(=[*])C1=N[*][*]S 1	-0.422	0 out of 1

Sorafenib

TOPKAT_Developmental_Toxicity_Potential



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Toxic

Probability: 0.592

Enrichment: 1.13

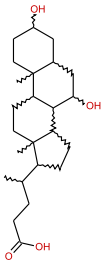
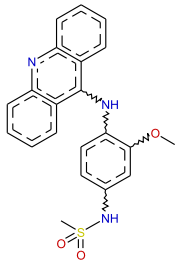
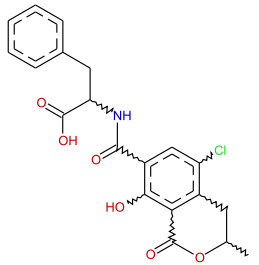
Bayesian Score: 1.15

Mahalanobis Distance: 12.6

Mahalanobis Distance p-value: 2.07e-006

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

Structural Similar Compounds

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

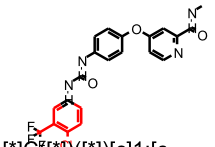
Model Applicability

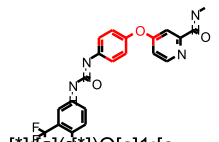
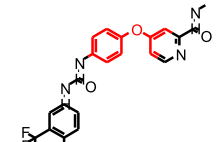
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

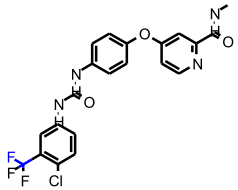
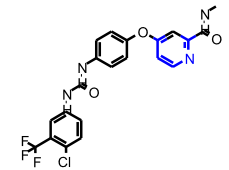
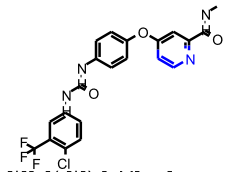
Feature Contribution

Top features for positive contribution

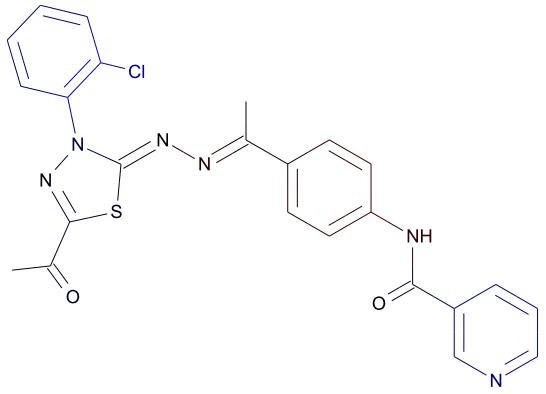
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1559190850	 [*]C([*])([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	0.441	3 out of 3

SCFP_6	-488587948	 [*][c](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

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.208

Enrichment: 0.649

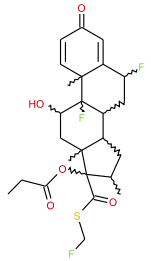
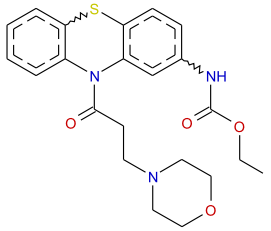
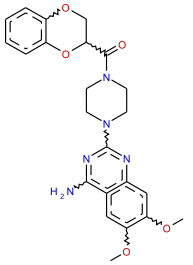
Bayesian Score: -7.85

Mahalanobis Distance: 14.2

Mahalanobis Distance p-value: 7.34e-006

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

Structural Similar Compounds

Name	Fluticasone	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.661	0.666	0.695
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

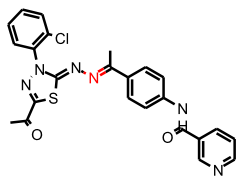
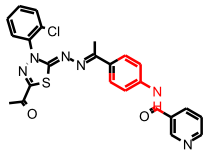
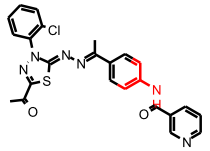
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC28 out of range. Value: 3.6547. Training min, max, SD, explained variance: -2.8298, 3.1935, 1.043, 0.0111.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
7. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1

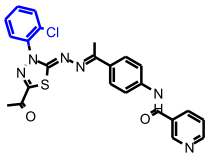
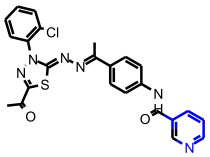
Feature Contribution

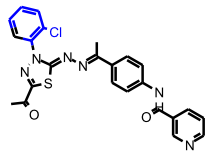
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

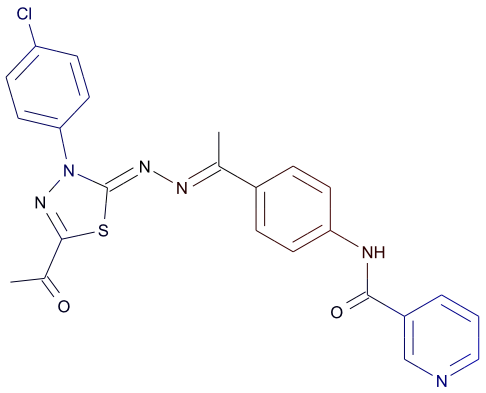
ECFP_6	-1087070950	 [*]N=[*]	0.724	10 out of 14
ECFP_6	738938915	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.617	2 out of 2
ECFP_6	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1625071872	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[c]:1Cl	-0.935	0 out of 5
ECFP_6	2013347047	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.805	0 out of 4

ECFP_6	1641317964	 [*][c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.789	1 out of 11

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.205

Enrichment: 0.64

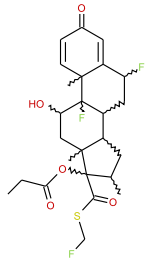
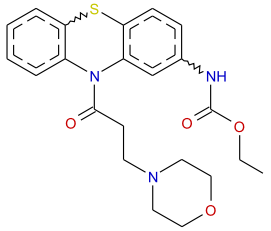
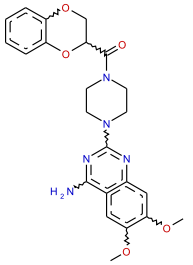
Bayesian Score: -6.85

Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 2.29e-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	Fluticasone	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.661	0.681	0.699
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

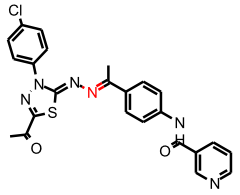
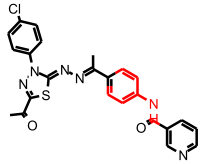
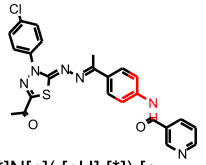
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

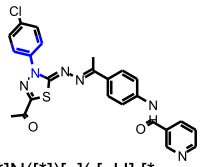
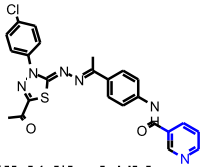
Feature Contribution

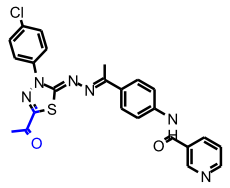
Top features for positive contribution

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	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9

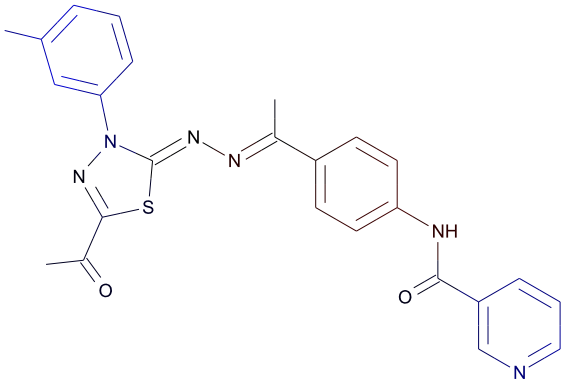
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.805	0 out of 4
ECFP_6	2013347047	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.805	0 out of 4

ECFP_6	129482634	 [*]C(=[*])C(=O)C	-0.657	0 out of 3

7c

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



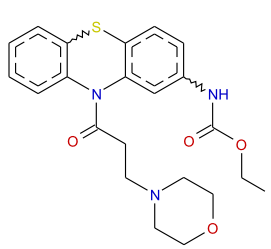
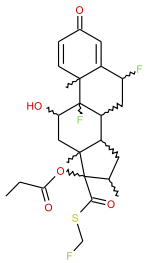
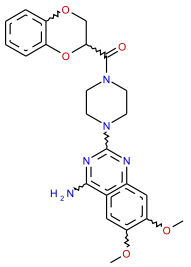
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.214
Enrichment: 0.668
Bayesian Score: -8.94
Mahalanobis Distance: 14
Mahalanobis Distance p-value: 1.35e-005

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

Structural Similar Compounds

Name	Moricizine	Fluticasone	Doxazosin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.656	0.667	0.676
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

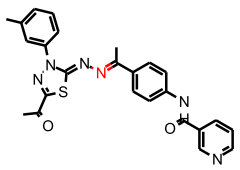
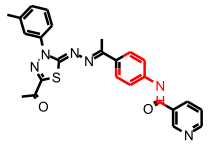
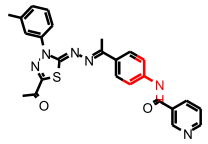
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

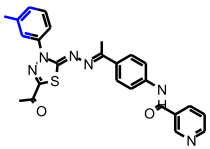
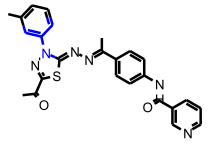
Feature Contribution

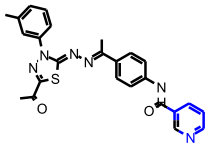
Top features for positive contribution

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	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9

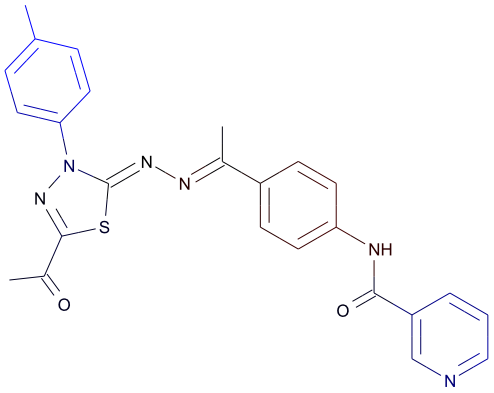
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-179515162	 [*]:[cH]:[c](C):[cH]:[*]	-1.41	0 out of 10
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.805	0 out of 4

ECFP_6	2013347047	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.805	0 out of 4

7d

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



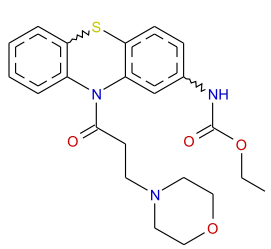
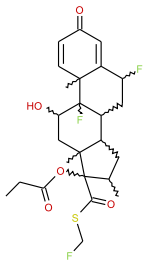
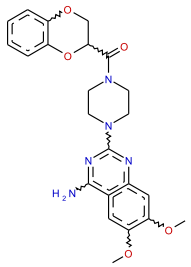
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.226
Enrichment: 0.706
Bayesian Score: -10.3
Mahalanobis Distance: 13.6
Mahalanobis Distance p-value: 6.08e-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	Moricizine	Fluticasone	Doxazosin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.658	0.666	0.679
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

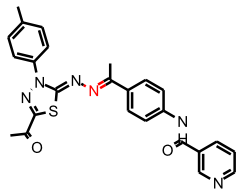
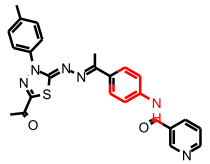
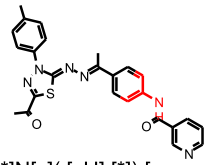
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

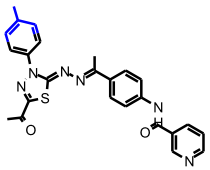
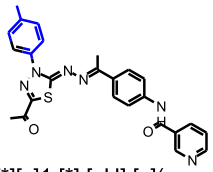
Feature Contribution

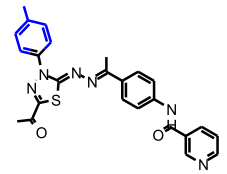
Top features for positive contribution

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	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9

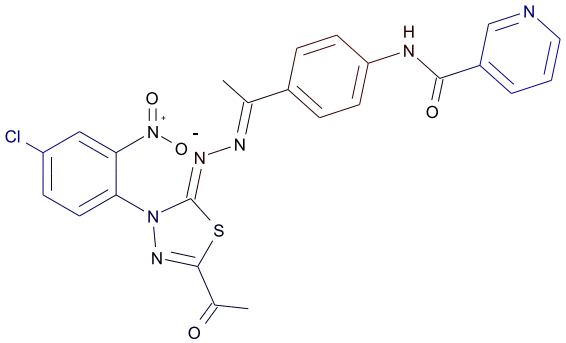
Top Features for negative contribution

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

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

7e

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.209

Enrichment: 0.652

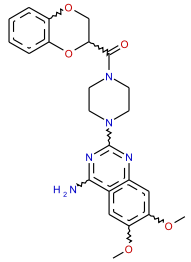
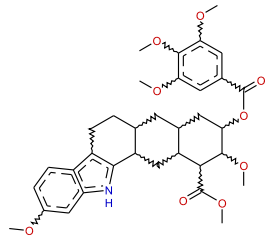
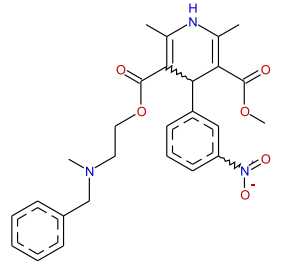
Bayesian Score: -8.12

Mahalanobis Distance: 16.1

Mahalanobis Distance p-value: 1.72e-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	Doxazosin	Reserpine	Nicardipine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.778	0.802	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

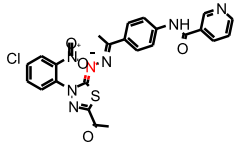
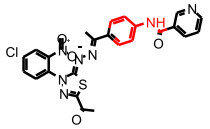
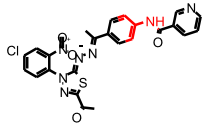
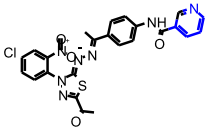
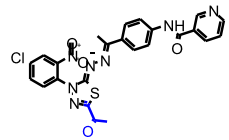
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

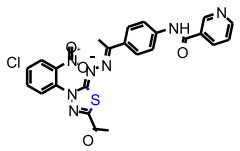
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1043790491: [*][N+](=[*])[*]
3. Unknown ECFP_2 feature: 781519895: [*][O-]
4. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
5. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
6. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
7. Unknown ECFP_2 feature: -1956535100: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
8. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
9. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
10. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
11. Unknown ECFP_2 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
12. Unknown ECFP_2 feature: 2104376220: [*][N+](=O)[*]
13. Unknown ECFP_2 feature: -659271057: [*][N+](=[*])[O-]

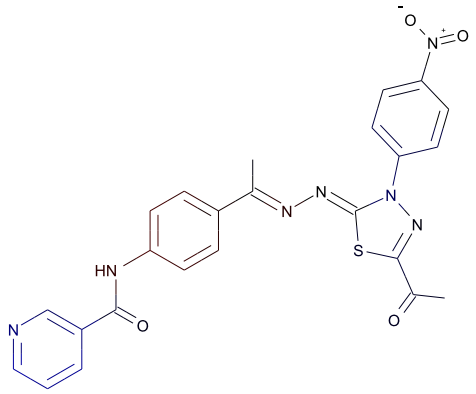
Feature Contribution

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	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
ECFP_6	2013347047	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.805	0 out of 4
ECFP_6	129482634	 [*]C(=[*])C(=O)C	-0.657	0 out of 3

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



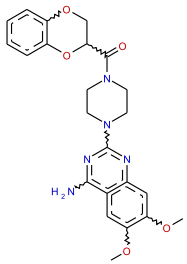
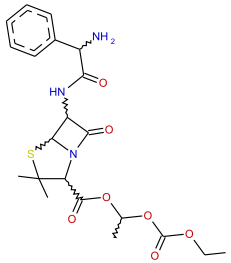
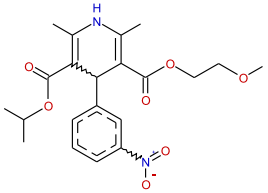
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.205
Enrichment: 0.639
Bayesian Score: -6.25
Mahalanobis Distance: 14.3
Mahalanobis Distance p-value: 5.32e-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	Doxazosin	Bacampicillin	Nimodipine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.737	0.782	0.794
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

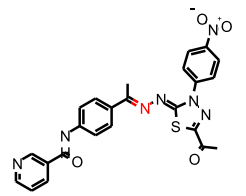
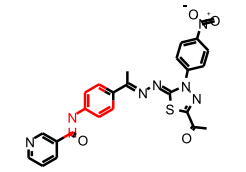
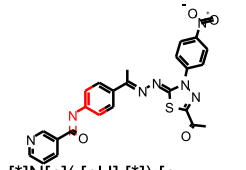
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1043790491: [*][N+](=[*])[*]
3. Unknown ECFP_2 feature: 781519895: [*][O-]
4. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
5. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
6. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
7. Unknown ECFP_2 feature: -179073144: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
8. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
9. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
10. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
11. Unknown ECFP_2 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
12. Unknown ECFP_2 feature: 2104376220: [*][N+](=O)[*]
13. Unknown ECFP_2 feature: -659271057: [*][N+](=[*])[O-]

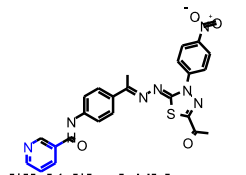
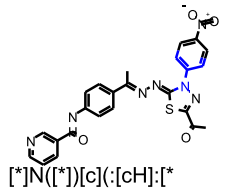
Feature Contribution

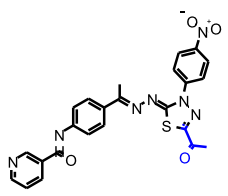
Top features for positive contribution

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	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9

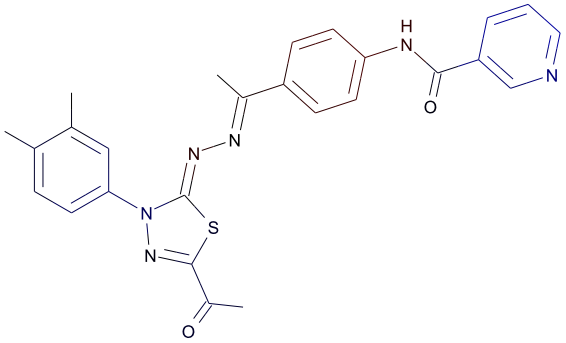
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	2013347047	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.805	0 out of 4
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.805	0 out of 4

ECFP_6	129482634	 [*]C(=[*])C(=O)C	-0.657	0 out of 3

7g

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



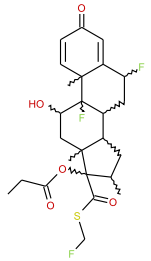
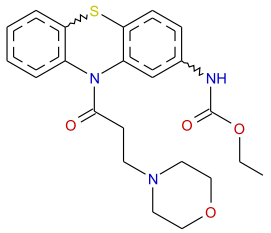
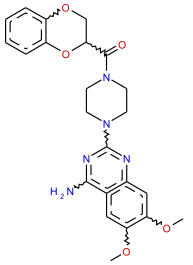
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.205
Enrichment: 0.641
Bayesian Score: -5.58
Mahalanobis Distance: 14.1
Mahalanobis Distance p-value: 1.02e-005

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

Structural Similar Compounds

Name	Fluticasone	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.663	0.686	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

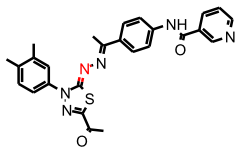
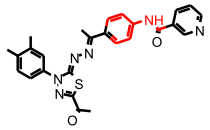
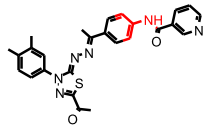
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

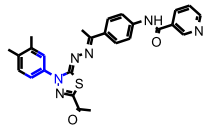
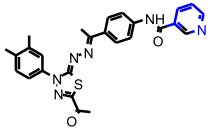
Feature Contribution

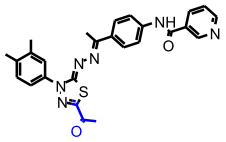
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_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	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9

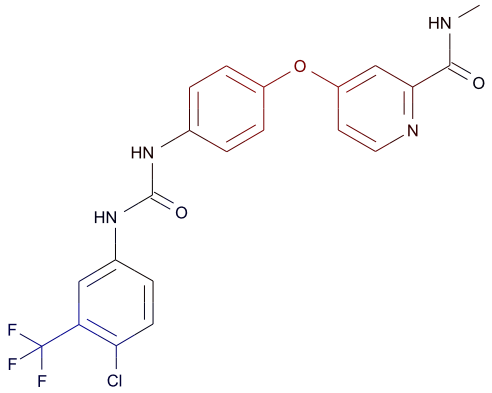
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-175021654	 [*]N([*])[c](:[cH]:[*]):[cH]:[*]	-0.805	0 out of 4
ECFP_6	2013347047	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.805	0 out of 4

ECFP_6	129482634	 [*]C(=[*])C(=O)C	-0.657	0 out of 3

Sorafenib

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Carcinogen

Probability: 0.257

Enrichment: 0.801

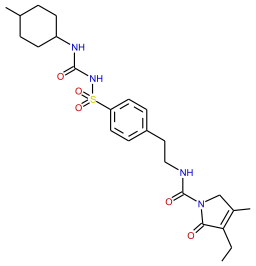
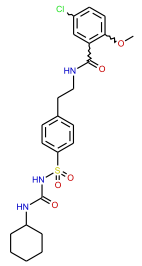
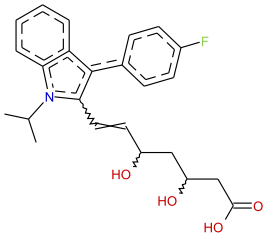
Bayesian Score: -0.321

Mahalanobis Distance: 14.9

Mahalanobis Distance p-value: 4.21e-007

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

Structural Similar Compounds

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

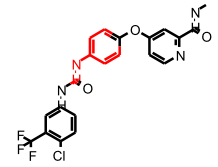
Model Applicability

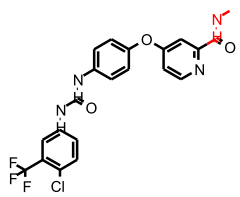
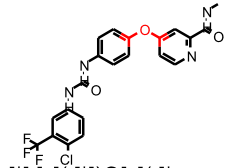
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

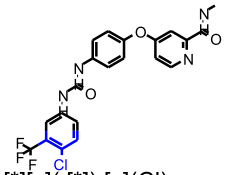
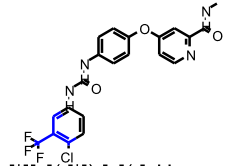
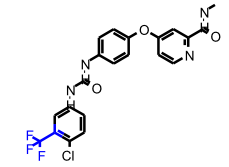
Feature Contribution

Top features for positive contribution

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

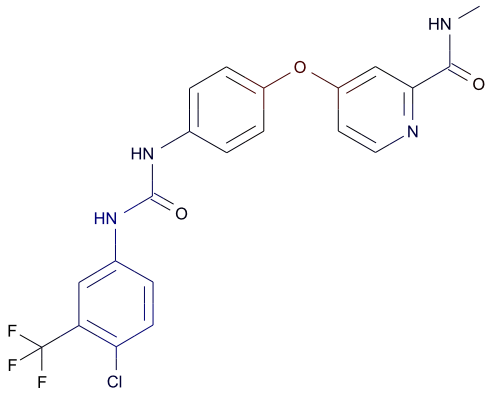
ECFP_6	1338334141	 [*]C(=[*])NC	0.442	2 out of 3
ECFP_6	1305253718	 [*]:[c](:[*])O[c](:[*]):[*]	0.424	1 out of 1

Top Features for negative contribution

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

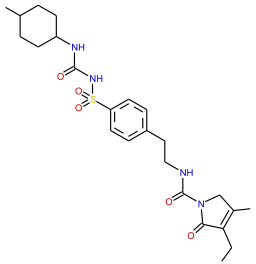
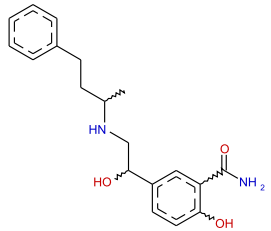
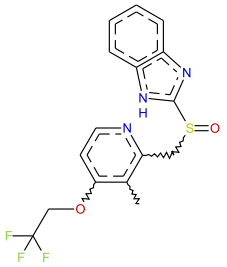
Bayesian Score: -3.89

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

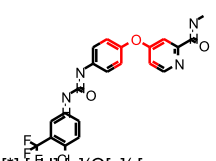
Model Applicability

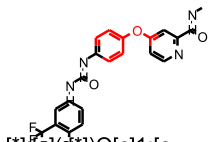
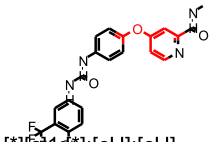
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

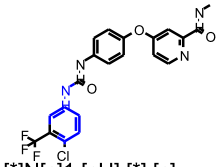
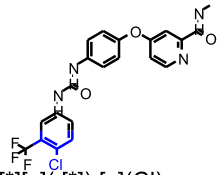
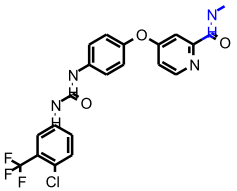
Feature Contribution

Top features for positive contribution

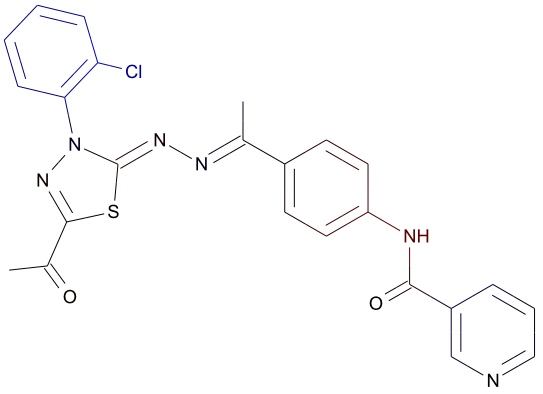
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
ECFP_4	-834094296	 [*]:[cH]:[c](O[c](:[cH]:[*]):[cH]:[*]):[cH]:[*]	0.351	1 out of 1

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

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.229

Enrichment: 0.779

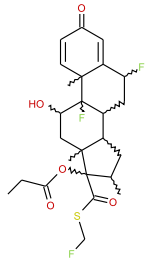
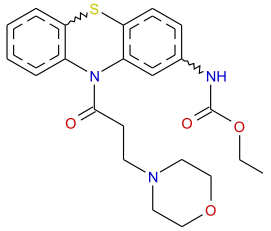
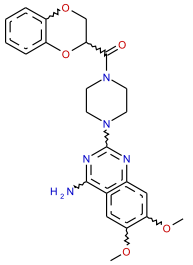
Bayesian Score: -3.3

Mahalanobis Distance: 18.4

Mahalanobis Distance p-value: 5.07e-015

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

Structural Similar Compounds

Name	Fluticasone	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.645	0.661	0.680
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

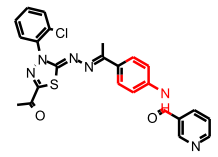
Model Applicability

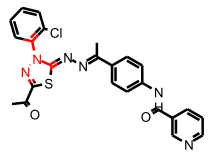
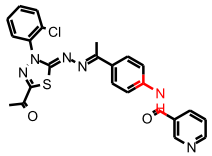
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC16 out of range. Value: -3.3134. Training min, max, SD, explained variance: -3.1219, 5.3717, 1.291, 0.0175.

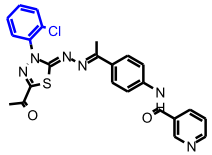
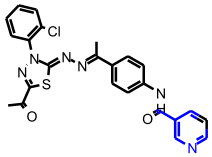
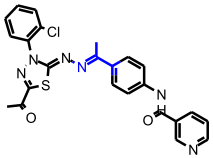
Feature Contribution

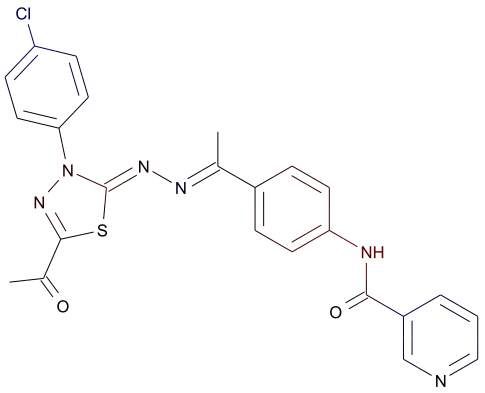
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7

FCFP_6	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.46	1 out of 1
FCFP_6	1294255210	 [*]C(=[*])N[c](:[*]):[*]	0.441	12 out of 28

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-2095752315	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[c]:1Cl	-1.13	0 out of 8
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.281

Enrichment: 0.956

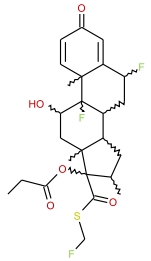
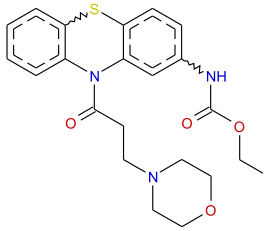
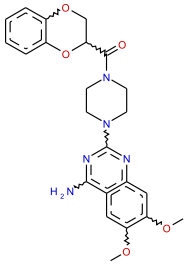
Bayesian Score: -0.944

Mahalanobis Distance: 18.1

Mahalanobis Distance p-value: 1.78e-014

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

Structural Similar Compounds

Name	Fluticasone	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.644	0.670	0.688
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

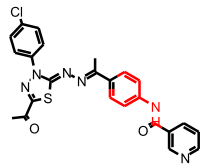
Model Applicability

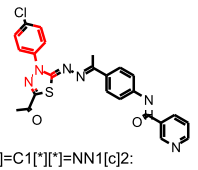
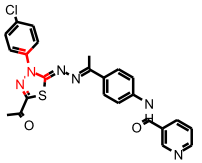
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

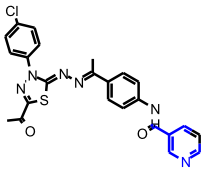
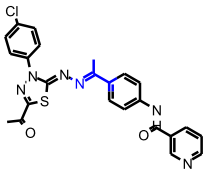
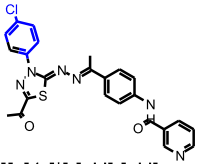
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7

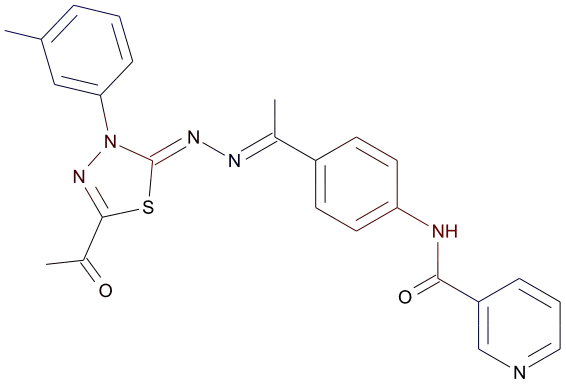
FCFP_6	-319371573	 [*]=C1[*][*]=NN1[c]2:[cH]:[cH]:[*]:[cH]:[cH]:2	0.46	1 out of 1
FCFP_6	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21
FCFP_6	551850122	 [*][c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.433	8 out of 49

7c

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



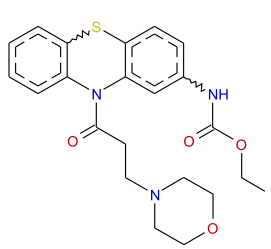
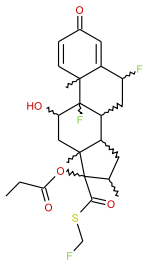
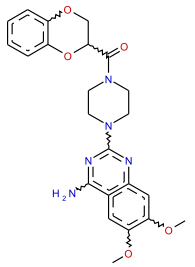
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.285
Enrichment: 0.968
Bayesian Score: -0.796
Mahalanobis Distance: 20.4
Mahalanobis Distance p-value: 8.37e-020

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

Structural Similar Compounds

Name	Moricizine	Fluticasone	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.642	0.654	0.664
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

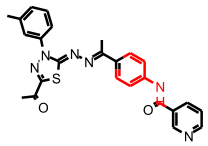
Model Applicability

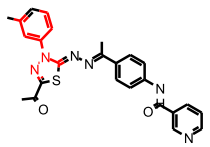
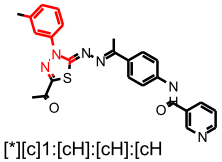
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

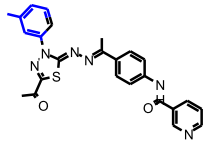
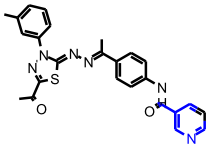
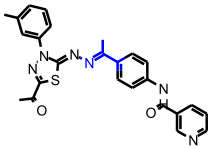
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7

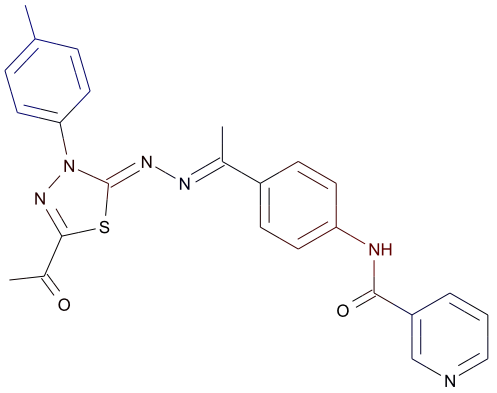
FCFP_6	-319371573	 [*]=C1[*][*]=NN1[c]2:[cH]:[cH]:[*]:[cH]:[cH]:2	0.46	1 out of 1
FCFP_6	-198775421	 [*][c]1:[cH]:[cH]:[cH]:[c](:[cH]:1)N2N=[*][*]C2=[*]	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	 [*][c]1:[*]:[cH]:[cH]:[c](C):[cH]:1	-1.29	0 out of 10
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21

7d

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.264

Enrichment: 0.896

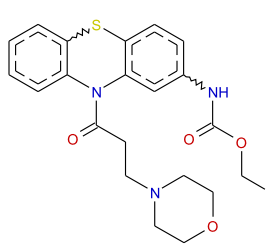
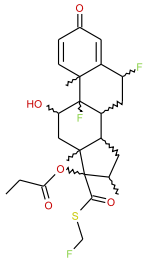
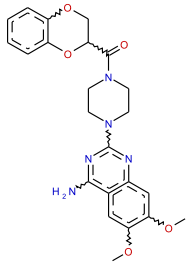
Bayesian Score: -1.68

Mahalanobis Distance: 18.7

Mahalanobis Distance p-value: 7.74e-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	Moricizine	Fluticasone	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.646	0.653	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

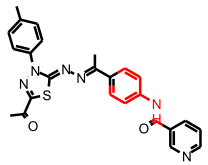
Model Applicability

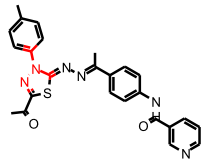
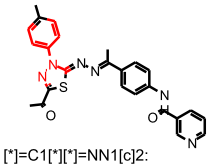
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

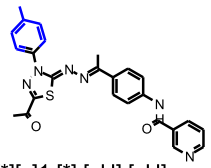
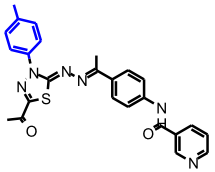
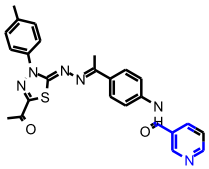
Feature Contribution

Top features for positive contribution

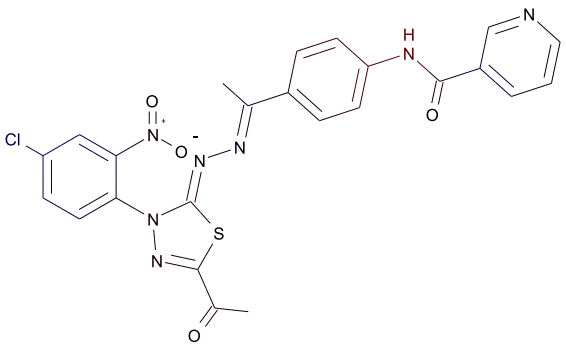
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7

FCFP_6	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.46	1 out of 1
FCFP_6	-319371573	 [*]=C1[*][*]=NN1[c]2:[cH]:[cH]:[*]:[cH]:[cH]:2	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	 [*][c]1:[*]:[cH]:[cH]:[c](C):[cH]:1	-1.29	0 out of 10
FCFP_6	2109043264	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.947	0 out of 6
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.282

Enrichment: 0.958

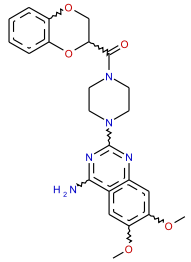
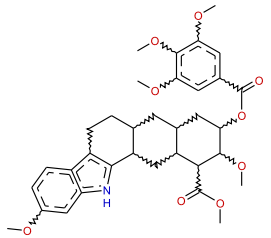
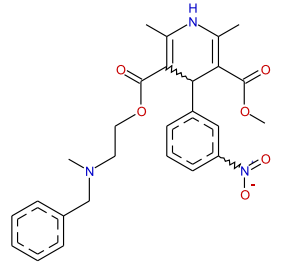
Bayesian Score: -0.917

Mahalanobis Distance: 18.5

Mahalanobis Distance p-value: 2.22e-015

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

Structural Similar Compounds

Name	Doxazosin	Reserpine	Nicardipine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.772	0.795	0.807
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

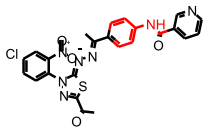
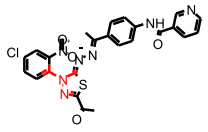
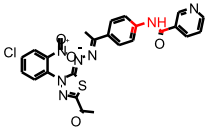
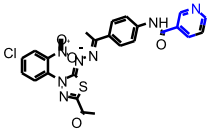
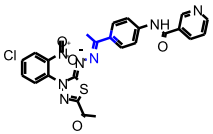
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

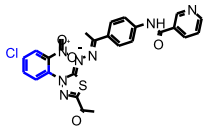
1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 5: [*][O-]
3. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
4. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
5. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
6. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

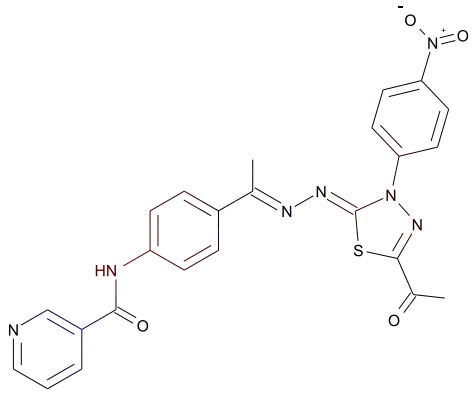
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.46	1 out of 1
FCFP_6	1294255210	 [*]C(=[*])N[c](:[*]):[*]	0.441	12 out of 28
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21

FCFP_6	551850122	 [*][c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.433	8 out of 49



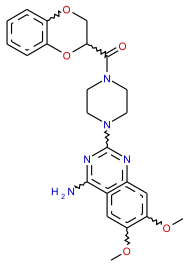
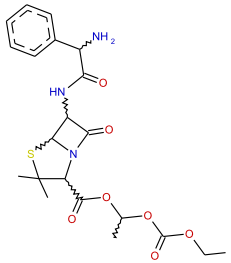
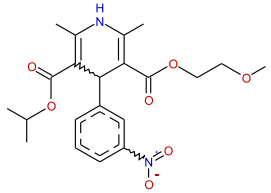
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.336
Enrichment: 1.14
Bayesian Score: 1.04
Mahalanobis Distance: 18
Mahalanobis Distance p-value: 2.98e-014

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

Structural Similar Compounds

Name	Doxazosin	Bacampicillin	Nimodipine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.727	0.771	0.774
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

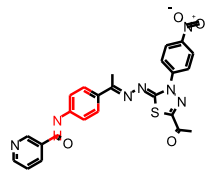
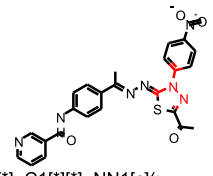
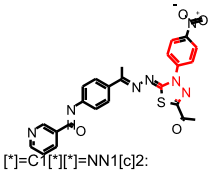
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 5: [*][O-]
3. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
4. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
5. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
6. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

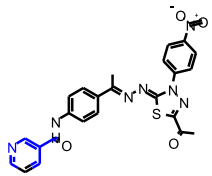
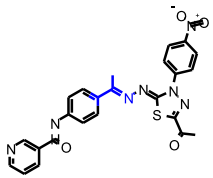
Feature Contribution

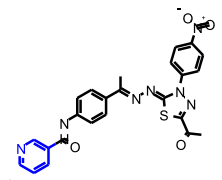
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	675799546	 [*]=C1[*][*]=NN1[c](:[*]):[*]	0.46	1 out of 1
FCFP_6	-319371573	 [*]=C1[*][*]=NN1[c]2:[cH]:[cH]:[*]:[cH]:[cH]:2	0.46	1 out of 1

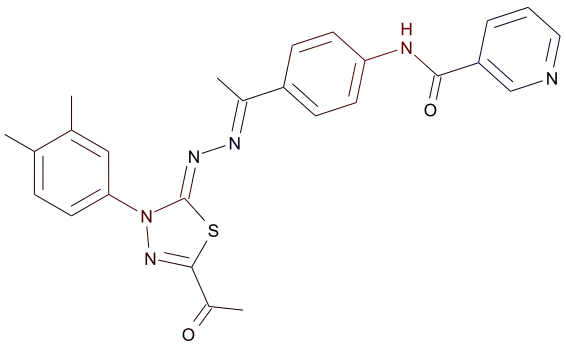
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21

FCFP_6	1888947587	 [*][c]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.423	0 out of 2

7g

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



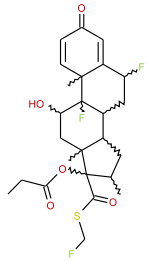
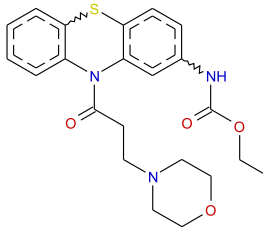
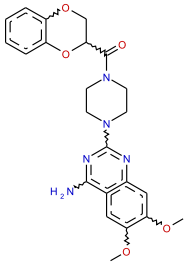
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.35
Enrichment: 1.19
Bayesian Score: 1.5
Mahalanobis Distance: 20.3
Mahalanobis Distance p-value: 1.69e-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	Fluticasone	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.650	0.677	0.705
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

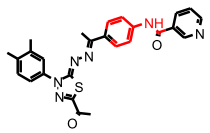
Model Applicability

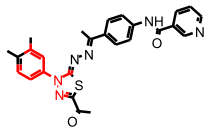
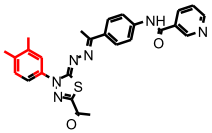
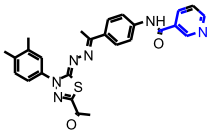
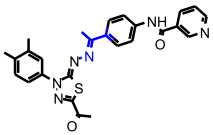
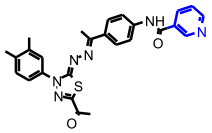
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

Feature Contribution

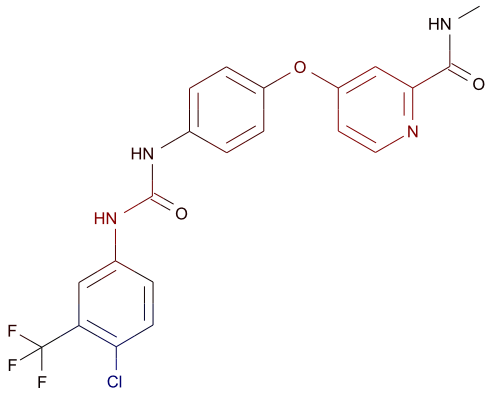
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7

FCFP_6	-319371573	 [*]=C1[*][*]=NN1[c]2:[cH]:[cH]:[*]:[cH]:[cH]:2	0.46	1 out of 1
FCFP_6	1075826633	 [*][c]1:[*]:[cH]:[c](C):[c](C):[cH]:1	0.46	1 out of 1
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
FCFP_6	1153798395	 [*]C(=[*])[c]1:[cH]:[*]:[cH]:n:[cH]:1	-0.582	0 out of 3
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]):[*]	-0.489	3 out of 21
FCFP_6	1888947587	 [*][c]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.423	0 out of 2

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

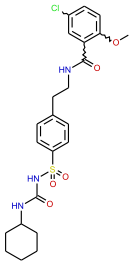
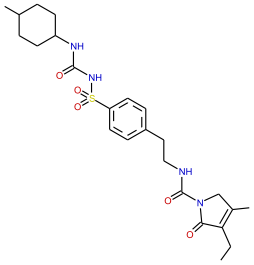
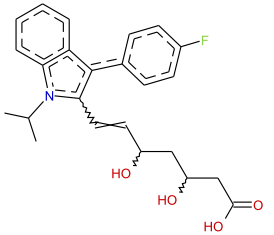
Bayesian Score: 4.21

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

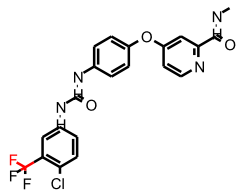
Model Applicability

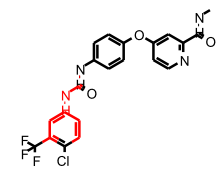
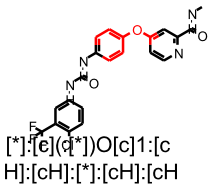
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

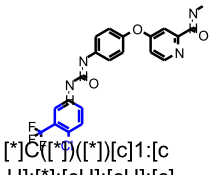
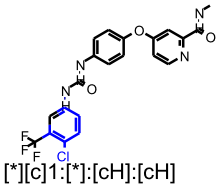
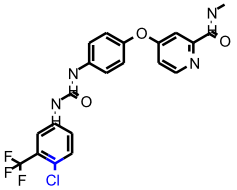
Feature Contribution

Top features for positive contribution

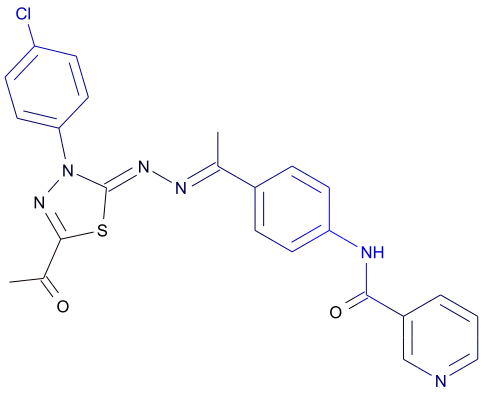
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	71953198	 [*]C([*])([*])F	0.612	12 out of 23

FCFP_6	-1838187238	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.565	4 out of 7
FCFP_6	-1270820019	 [*]:[c](:[*])O[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.46	1 out of 1

Top Features for negative contribution

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

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.142

Enrichment: 0.471

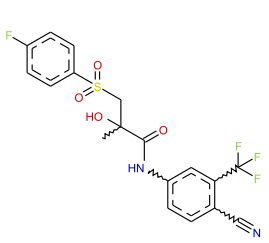
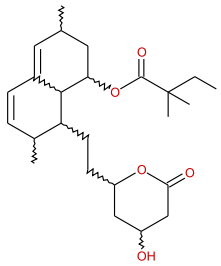
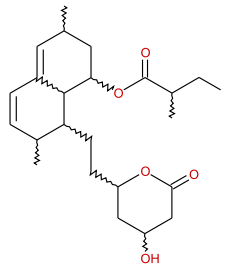
Bayesian Score: -14.4

Mahalanobis Distance: 16.5

Mahalanobis Distance p-value: 1.72e-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	Simvastatin	Lovastatin
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.833	0.843	0.847
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

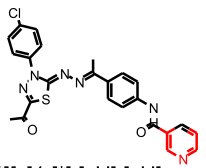
Model Applicability

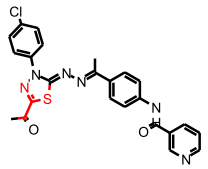
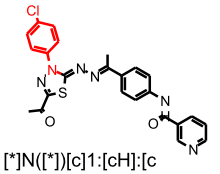
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC2 out of range. Value: 4.2927. Training min, max, SD, explained variance: -5.2888, 4.2744, 2.566, 0.1229.
2. OPS PC5 out of range. Value: 3.9023. Training min, max, SD, explained variance: -3.5268, 3.8048, 1.733, 0.0560.

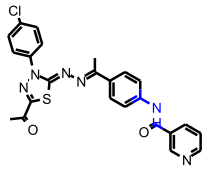
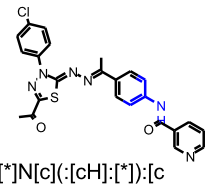
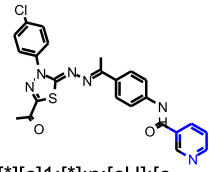
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	547884906	 [*][c]1:[*]:[cH]:[cH] :n:[cH]:1	0.4	1 out of 1

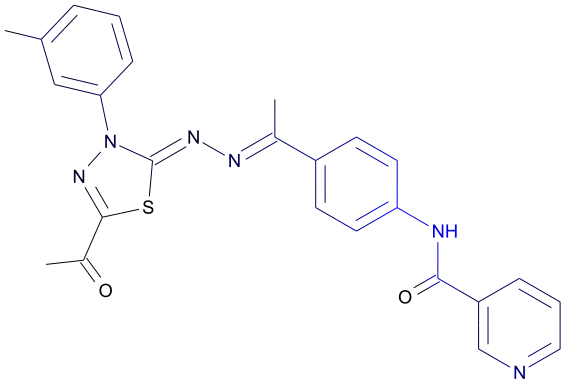
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14
FCFP_12	567484887	 [*]N([*])[c]1:[cH]:[c H]:[c](Cl):[cH]:[cH] :1	0.174	1 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]C(=[*])N[c](:[*]): [*]	-1.63	0 out of 12
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	730557100	 [*][c]1:[*]:n:[cH]:[c H]:[cH]:1	-0.859	0 out of 4

7c

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.15

Enrichment: 0.498

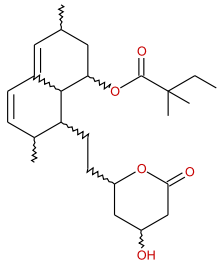
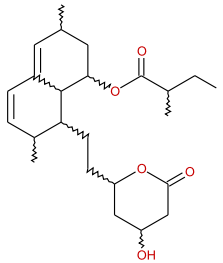
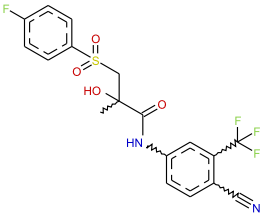
Bayesian Score: -11.5

Mahalanobis Distance: 17.9

Mahalanobis Distance p-value: 1.42e-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	Simvastatin	Lovastatin	Bicalutamide
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.832	0.833	0.838
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

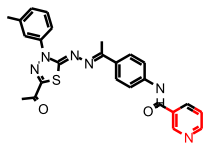
Model Applicability

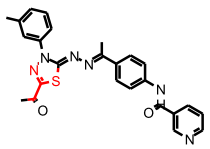
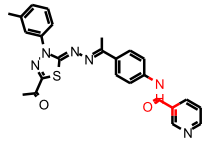
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: 2.6425. Training min, max, SD, explained variance: -3.4599, 2.3291, 1.246, 0.0290.

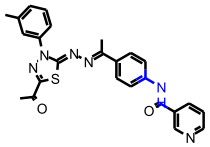
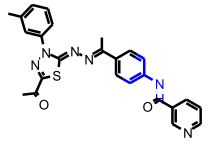
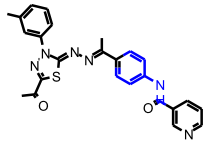
Feature Contribution

Top features for positive contribution

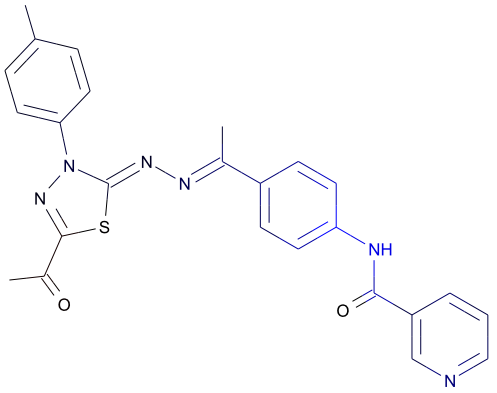
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	547884906	 [*][c]1:[*]:[cH]:[cH] :n:[cH]:1	0.4	1 out of 1

FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.168	3 out of 7

Top Features for negative contribution

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

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



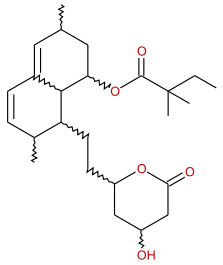
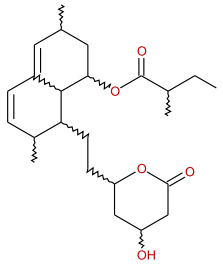
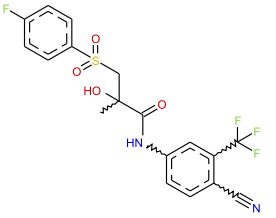
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.147
Enrichment: 0.49
Bayesian Score: -10.8
Mahalanobis Distance: 18.6
Mahalanobis Distance p-value: 4.71e-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	Simvastatin	Lovastatin	Bicalutamide
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.833	0.833	0.838
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

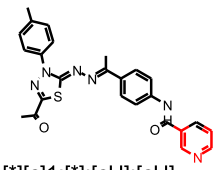
Model Applicability

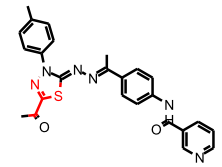
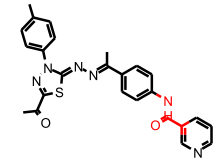
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: 2.3406. Training min, max, SD, explained variance: -3.4599, 2.3291, 1.246, 0.0290.

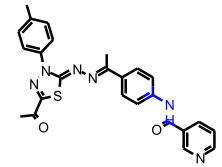
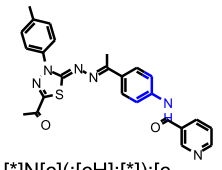
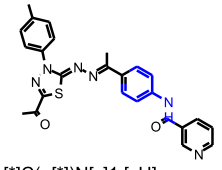
Feature Contribution

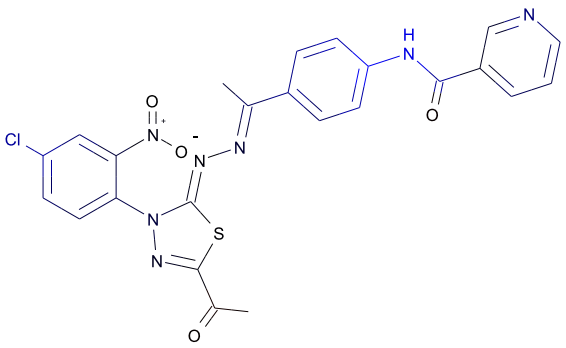
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	547884906	 [*][c]1:[*]:[cH]:[cH] :n:[cH]:1	0.4	1 out of 1

FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14
FCFP_12	-1549103449	 [*]NC(=O)[c](:[*]):[*]	0.168	3 out of 7

Top Features for negative contribution

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



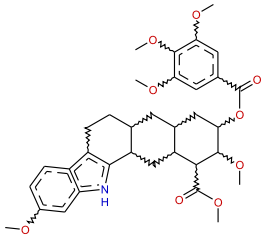
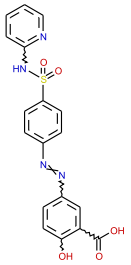
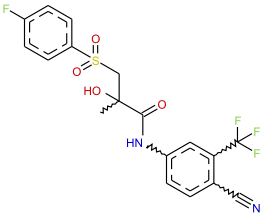
$C_{24}H_{18}ClN_7O_4S$
Molecular Weight: 535.96222
ALogP: 4.153
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.149
Enrichment: 0.494
Bayesian Score: -13.6
Mahalanobis Distance: 17.9
Mahalanobis Distance p-value: 1.53e-008

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

Structural Similar Compounds

Name	Reserpine	Sulfasalazine	Bicalutamide
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.814	0.966	1.038
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

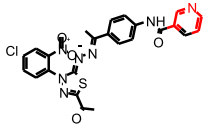
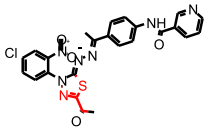
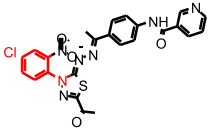
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC2 out of range. Value: 4.412. Training min, max, SD, explained variance: -5.2888, 4.2744, 2.566, 0.1229.
2. OPS PC5 out of range. Value: 3.8935. Training min, max, SD, explained variance: -3.5268, 3.8048, 1.733, 0.0560.
3. Unknown FCFP_2 feature: 5: [*][O-]
4. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
5. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
6. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
7. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

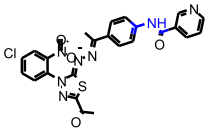
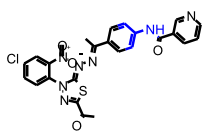
Feature Contribution

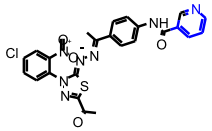
Top features for positive contribution

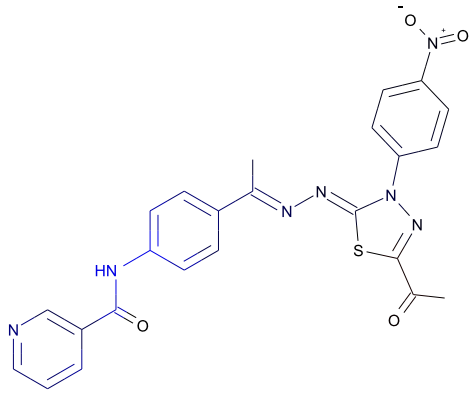
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set

FCFP_12	547884906	 [*][c]1:[*]:[cH]:[cH] :n:[cH]:1	0.4	1 out of 1
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14
FCFP_12	567484887	 [*]N([*])[c]1:[cH]:[c H]:[c](Cl):[cH]:[cH] :1	0.174	1 out of 2

Top Features for negative contribution

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

FCFP_12	730557100	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.859	0 out of 4



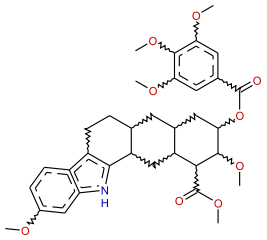
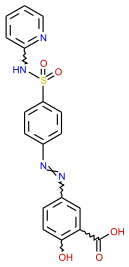
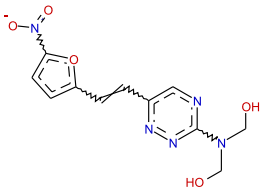
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.151
Enrichment: 0.501
Bayesian Score: -11.7
Mahalanobis Distance: 17.6
Mahalanobis Distance p-value: 2.38e-008

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

Structural Similar Compounds

Name	Reserpine	Sulfasalazine	Dihydroxymethylfuratrizine
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.872	0.932	1.010
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

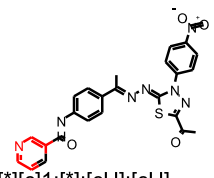
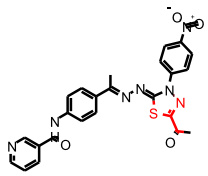
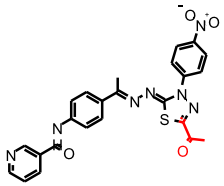
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 5: [*][O-]
3. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
4. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
5. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
6. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

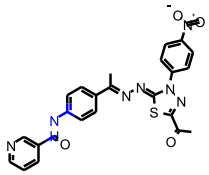
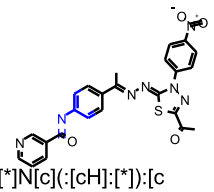
Feature Contribution

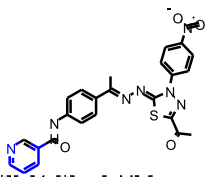
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set

FCFP_12	547884906	 [*][c]1:[*]:[cH]:[cH] :n:[cH]:1	0.4	1 out of 1
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.194	6 out of 14
FCFP_12	565968762	 [*]C(=[*])C(=O)C	0.168	3 out of 7

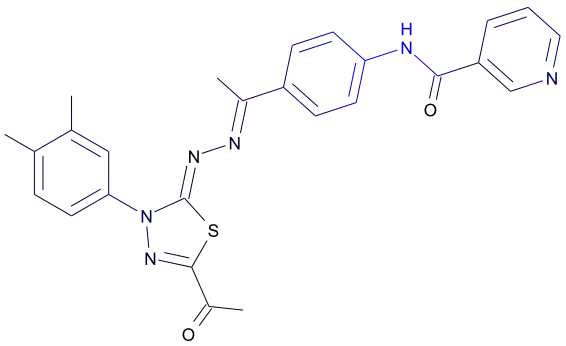
Top Features for negative contribution

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

FCFP_12	730557100	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.859	0 out of 4

7g

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



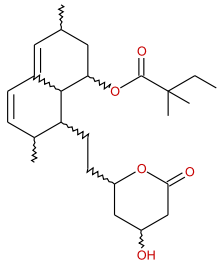
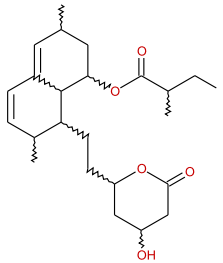
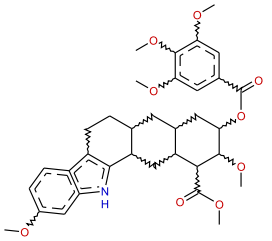
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.148
Enrichment: 0.491
Bayesian Score: -10.9
Mahalanobis Distance: 18.7
Mahalanobis Distance p-value: 3.9e-009

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

Structural Similar Compounds

Name	Simvastatin	Lovastatin	Reserpine
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.833	0.839	0.854
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

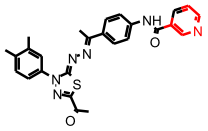
Model Applicability

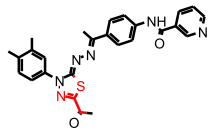
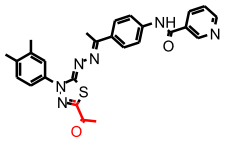
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: 2.3509. Training min, max, SD, explained variance: -3.4599, 2.3291, 1.246, 0.0290.

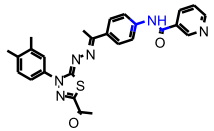
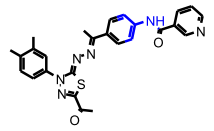
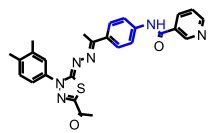
Feature Contribution

Top features for positive contribution

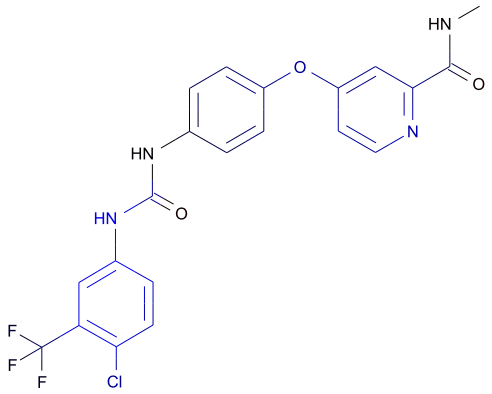
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	547884906	 [*][c]1:[*]:[cH]:[cH] :n:[cH]:1	0.4	1 out of 1

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

Top Features for negative contribution

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

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

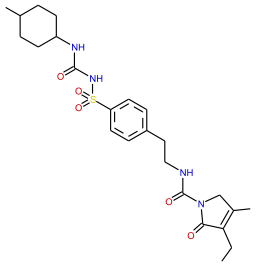
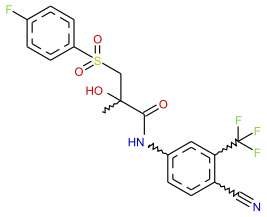
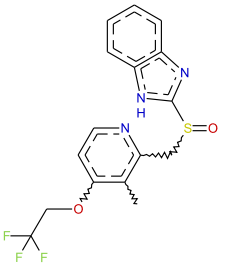
Mahalanobis Distance: 21.3

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

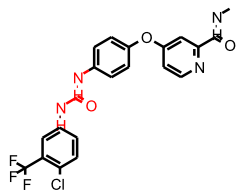
Model Applicability

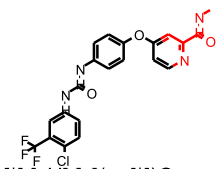
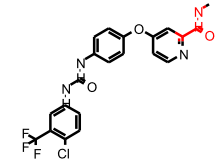
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

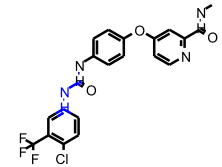
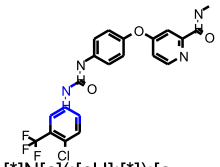
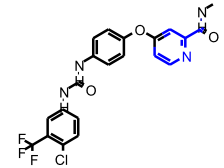
Feature Contribution

Top features for positive contribution

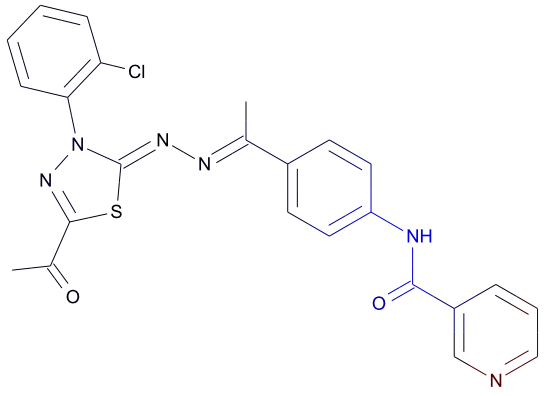
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1499521844	 [*]NC(=O)N[*]	0.39	5 out of 9

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

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.629

Enrichment: 0.914

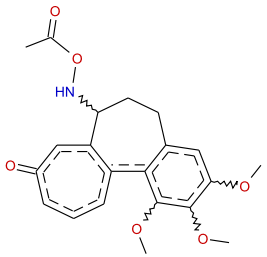
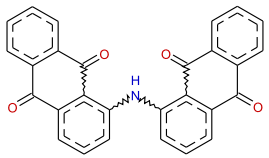
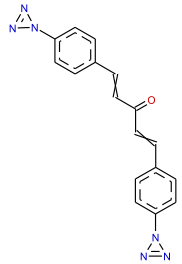
Bayesian Score: -4.62

Mahalanobis Distance: 11.2

Mahalanobis Distance p-value: 0.00185

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.763	0.804	0.853
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

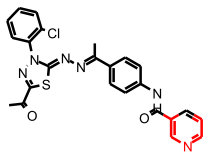
Model Applicability

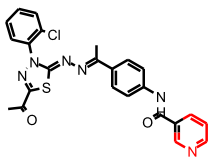
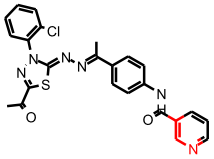
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

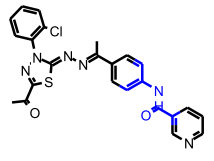
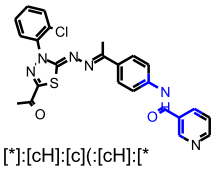
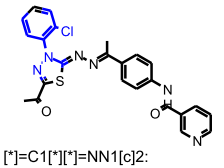
Feature Contribution

Top features for positive contribution

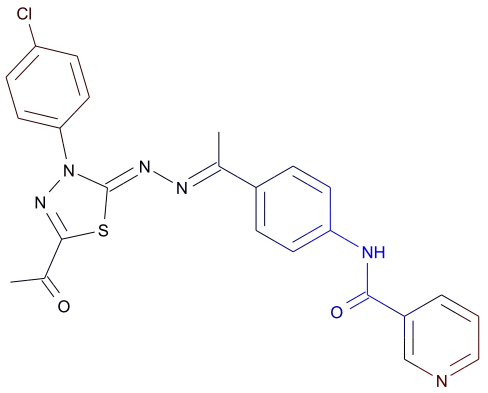
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	547884906	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	0.317	4 out of 4

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	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-1.29	0 out of 4
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4
FCFP_10	-759256695	 [*]=C1[*][*]=NN1[c]2: [cH]:[cH]:[*]:[cH]:[c]:2Cl	-0.507	0 out of 1

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.722

Enrichment: 1.05

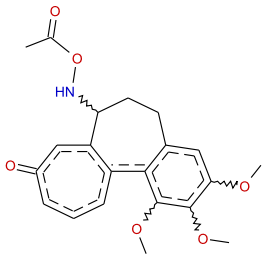
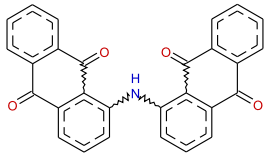
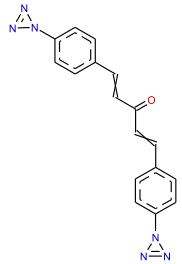
Bayesian Score: -3.11

Mahalanobis Distance: 11.2

Mahalanobis Distance p-value: 0.00185

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.763	0.808	0.853
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

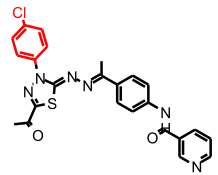
Model Applicability

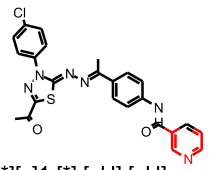
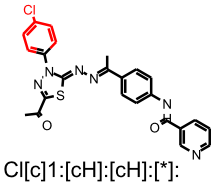
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

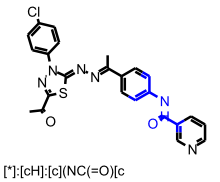
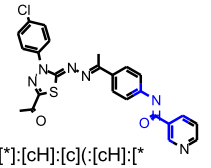
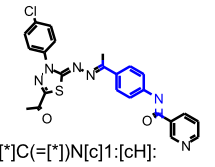
Feature Contribution

Top features for positive contribution

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

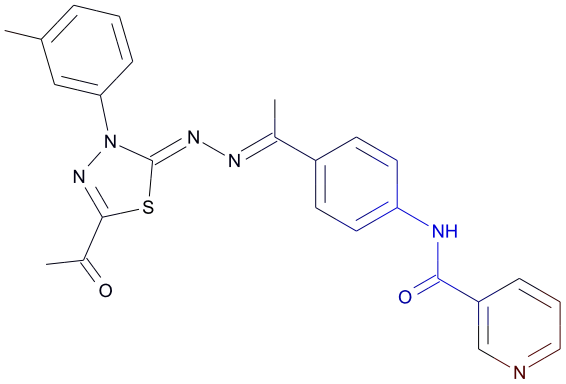
FCFP_10	547884906	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	0.317	4 out of 4
FCFP_10	-745491832	 Cl[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.304	29 out of 32

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1175232969	 [*]:[cH]:[c](NC(=O)[c]:([*]):[*]):[cH]:[*]	-1.29	0 out of 4
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4
FCFP_10	-790336137	 [*]C(=[*])N[c]1:[cH]:[cH]:[c](:[cH]:[cH]:1)C(=[*])[*]	-0.507	0 out of 1

7c

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.664

Enrichment: 0.963

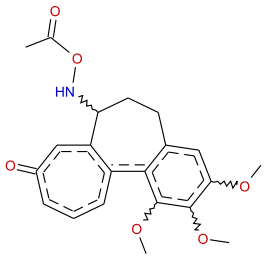
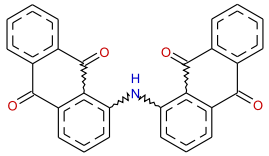
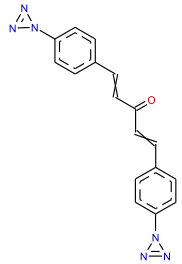
Bayesian Score: -4.12

Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 0.015

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.726	0.801	0.828
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

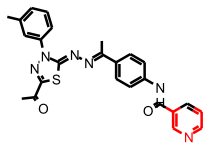
Model Applicability

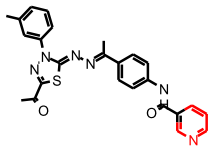
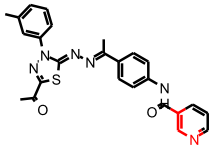
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

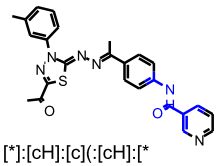
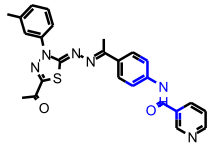
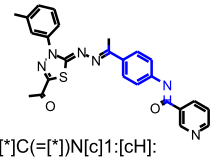
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	547884906	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	0.317	4 out of 4

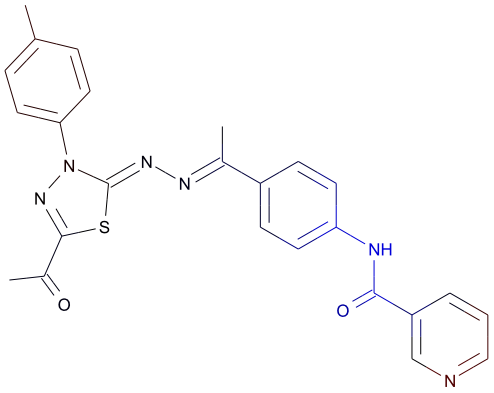
FCFP_10	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.285	10 out of 11
FCFP_10	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.259	14 out of 16

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4
FCFP_10	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-1.29	0 out of 4
FCFP_10	-790336137	 [*]C(=[*])N[c]1:[cH]: [cH]:[c](:[cH]:[cH]: 1)C(=[*])[*]	-0.507	0 out of 1

7d

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.717

Enrichment: 1.04

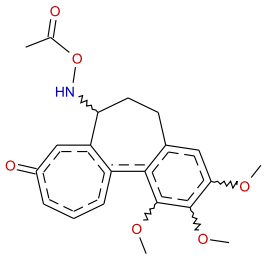
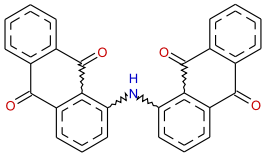
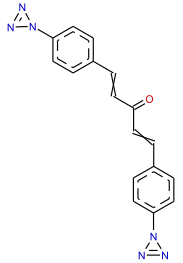
Bayesian Score: -3.2

Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 0.015

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.729	0.804	0.827
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

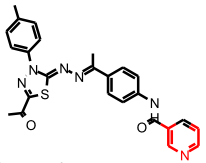
Model Applicability

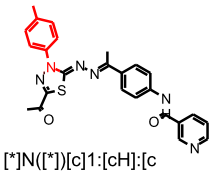
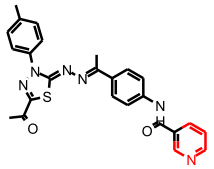
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

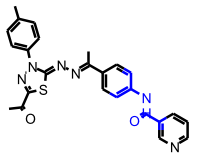
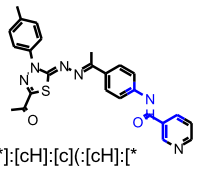
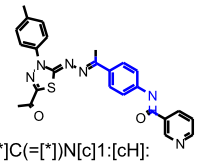
Feature Contribution

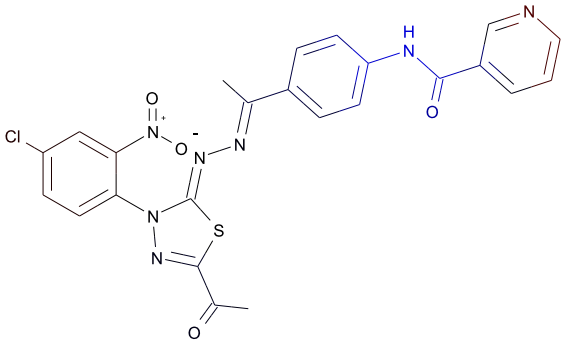
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	547884906	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	0.317	4 out of 4

FCFP_10	633795852	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.294	3 out of 3
FCFP_10	-1695756380	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	0.285	10 out of 11

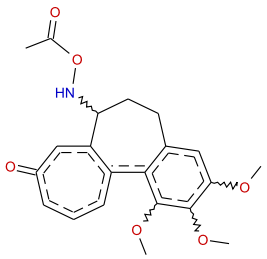
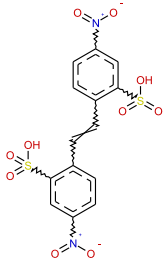
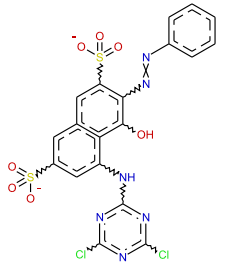
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-1.29	0 out of 4
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4
FCFP_10	-790336137	 [*]C(=[*])N[c]1:[cH]:[cH]:[c](:[cH]:[cH]:1)C(=[*])[*]	-0.507	0 out of 1



$C_{24}H_{18}ClN_7O_4S$
Molecular Weight: 535.96222
ALogP: 4.153
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Structural Similar Compounds

Name	COLCHICINE	2;2'-Stilbenedisulfonic acid; 4;4'-dinitro-	2;7-Naphthalenedisulfonic acid;
Structure			
Actual Endpoint	Moderate_Severe	Mild	Moderate_Severe
Predicted Endpoint	Moderate_Severe	Mild	Moderate_Severe
Distance	0.929	0.955	0.984
Reference	AJOPAA 31;837;48	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1062;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;1300;86

Model Prediction

Prediction: Mild
Probability: 0.709
Enrichment: 1.03
Bayesian Score: -3.36
Mahalanobis Distance: 12
Mahalanobis Distance p-value: 6.04e-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.

Model Applicability

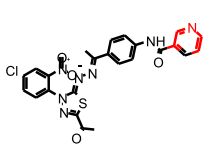
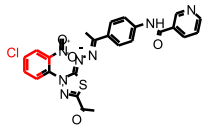
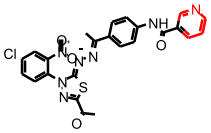
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
3. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
4. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
5. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

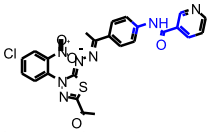
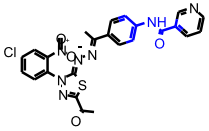
Feature Contribution

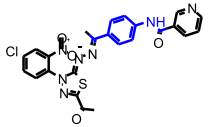
Top features for positive contribution

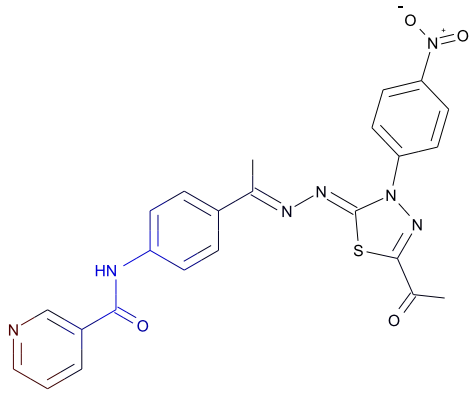
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set

FCFP_10	547884906	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	0.317	4 out of 4
FCFP_10	-745491832	 Cl[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.304	29 out of 32
FCFP_10	-1695756380	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	0.285	10 out of 11

Top Features for negative contribution

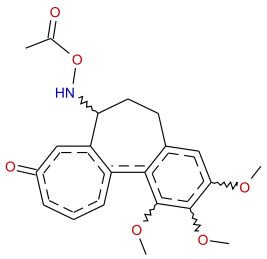
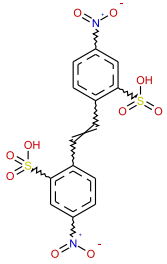
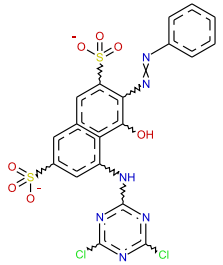
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4
FCFP_10	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-1.29	0 out of 4

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



$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Structural Similar Compounds

Name	COLCHICINE	2;2'-Stilbenedisulfonic acid; 4;4'-dinitro-	2;7-Naphthalenedisulfonic acid;
Structure			
Actual Endpoint	Moderate_Severe	Mild	Moderate_Severe
Predicted Endpoint	Moderate_Severe	Mild	Moderate_Severe
Distance	0.881	0.896	0.953
Reference	AJOPAA 31;837;48	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1062;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;1300;86

Model Prediction

Prediction: Mild
Probability: 0.67
Enrichment: 0.973
Bayesian Score: -4.02
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.00137

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

Model Applicability

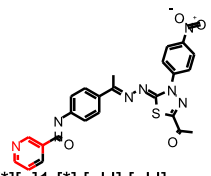
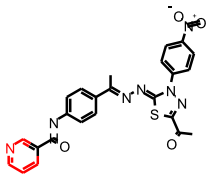
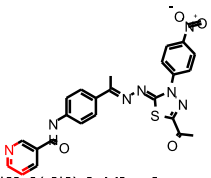
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
3. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
4. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
5. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

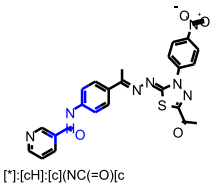
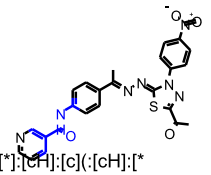
Feature Contribution

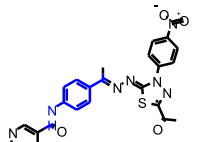
Top features for positive contribution

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

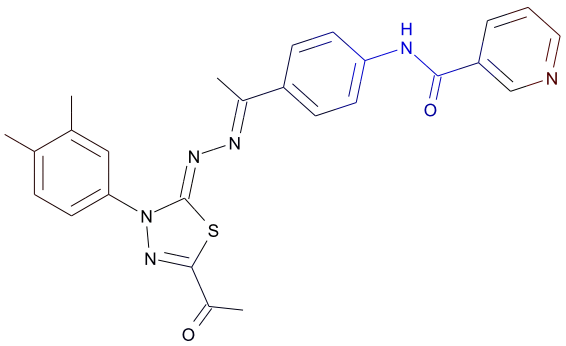
Top Features for negative contribution

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

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

7g

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



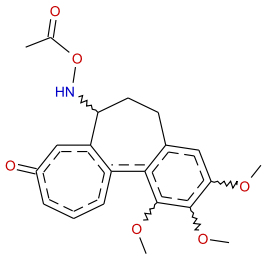
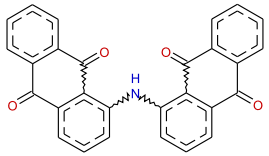
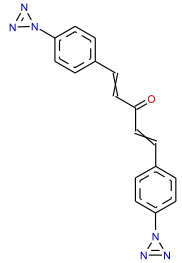
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.716
Enrichment: 1.04
Bayesian Score: -3.23
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 0.00942

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.772	0.793	0.843
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

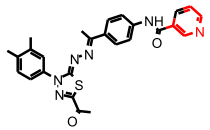
Model Applicability

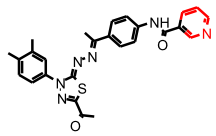
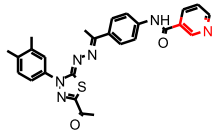
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

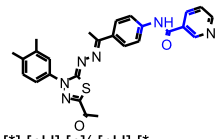
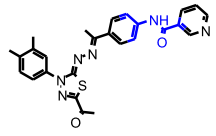
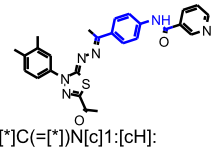
Feature Contribution

Top features for positive contribution

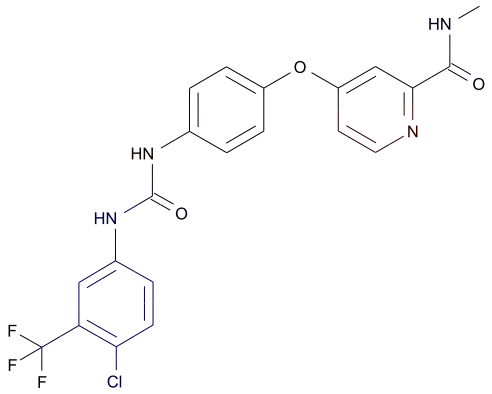
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	547884906	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	0.317	4 out of 4

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	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-1.29	0 out of 4
FCFP_10	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-1.29	0 out of 4
FCFP_10	-790336137	 [*]C(=[*])N[c]1:[cH]: [cH]:[c](:[cH]:[cH]: 1)C(=[*])[*]	-0.507	0 out of 1

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

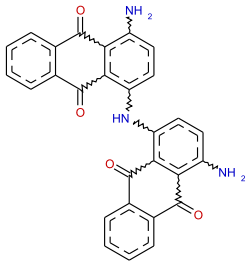
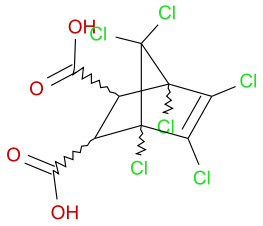
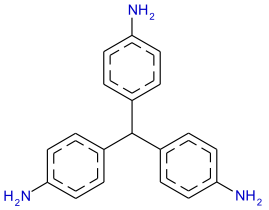
Mahalanobis Distance: 8.95

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

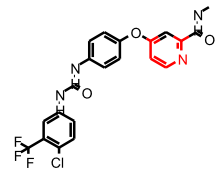
Model Applicability

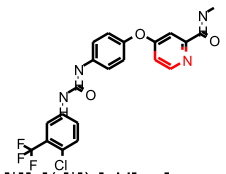
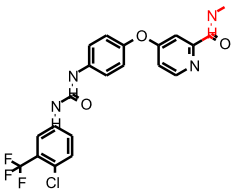
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

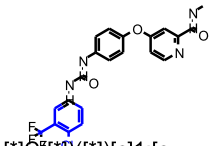
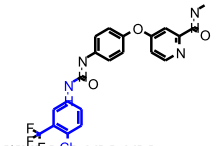
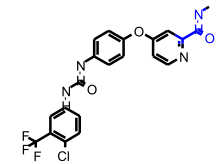
Feature Contribution

Top features for positive contribution

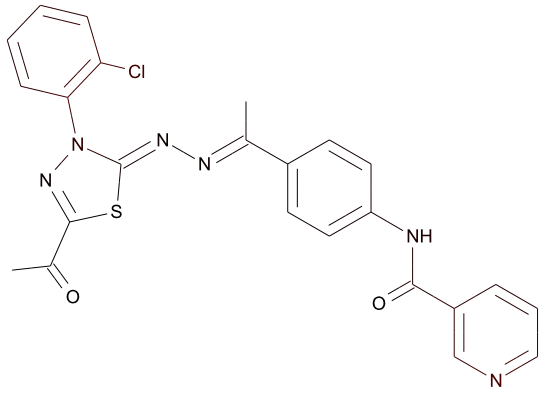
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.285	10 out of 11

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:[c H]:[*]:[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



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

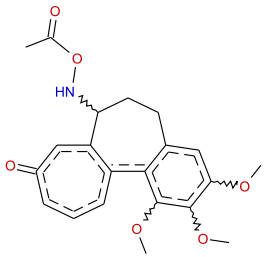
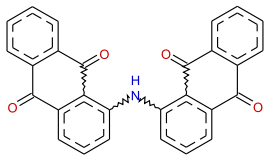
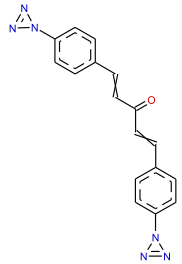
Bayesian Score: 3.96

Mahalanobis Distance: 8.11

Mahalanobis Distance p-value: 0.898

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.742	0.791	0.844
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

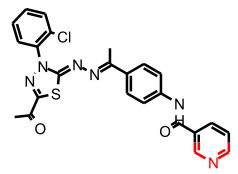
Model Applicability

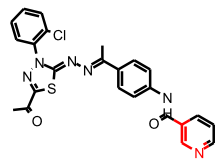
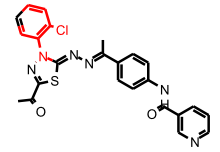
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

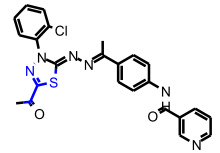
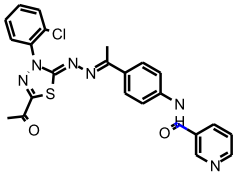
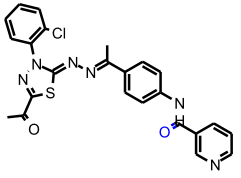
Feature Contribution

Top features for positive contribution

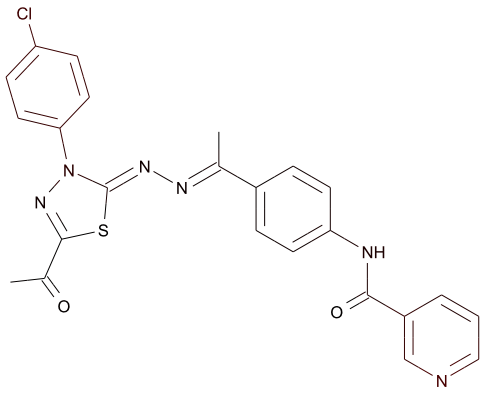
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*]:[cH]:n:[cH]:[*]	0.208	44 out of 44

FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-62776068	 [*]N([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	0.197	13 out of 13

Top Features for negative contribution

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

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

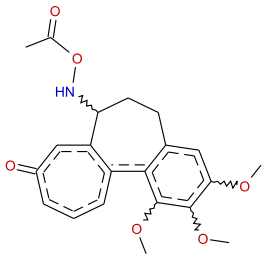
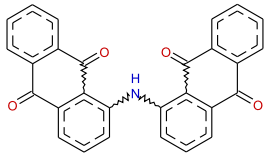
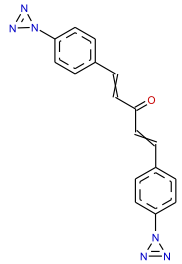
Bayesian Score: 4.06

Mahalanobis Distance: 8.11

Mahalanobis Distance p-value: 0.898

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.742	0.795	0.844
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

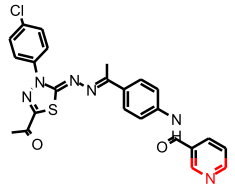
Model Applicability

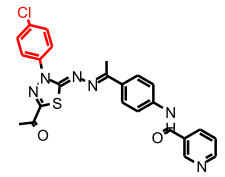
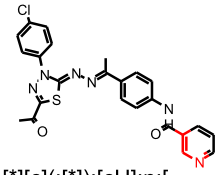
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

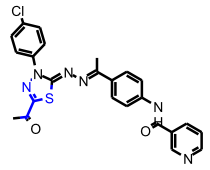
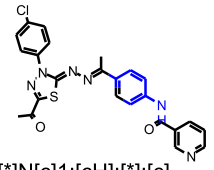
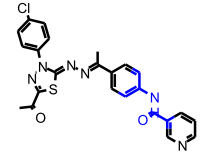
Feature Contribution

Top features for positive contribution

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

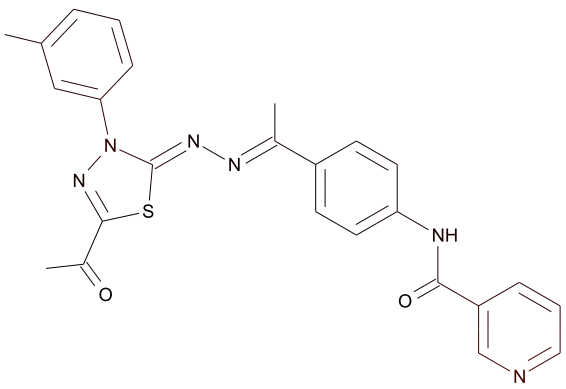
FCFP_12	-1508180856	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	0.2	17 out of 17
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.0662	198 out of 262
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0	102 out of 121
FCFP_12	1175232969	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	0	4 out of 5

7c

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

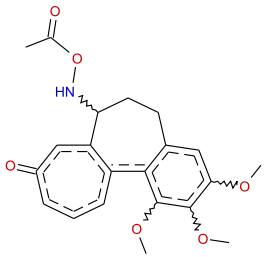
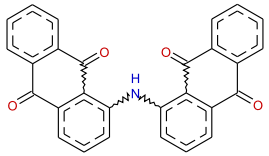
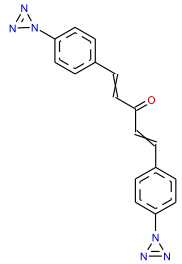
Bayesian Score: 3.34

Mahalanobis Distance: 8.04

Mahalanobis Distance p-value: 0.916

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.708	0.786	0.820
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

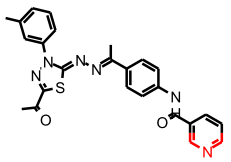
Model Applicability

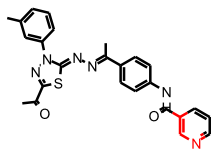
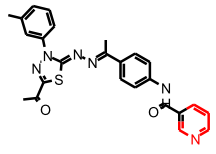
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

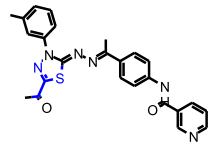
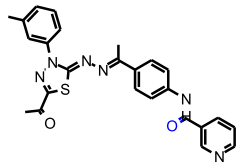
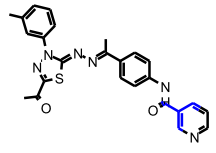
Feature Contribution

Top features for positive contribution

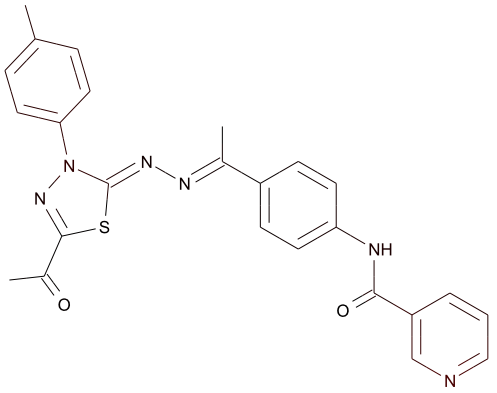
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*]:[cH]:n:[cH]:[*]	0.208	44 out of 44

FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	0.194	11 out of 11

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.0662	198 out of 262
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

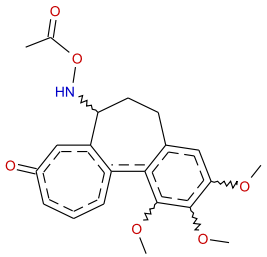
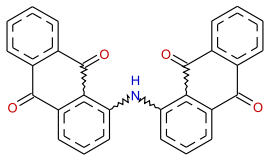
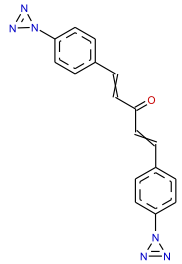
Bayesian Score: 3.45

Mahalanobis Distance: 8.04

Mahalanobis Distance p-value: 0.916

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	1;4-PENTADIENE-3-ONE;1;5-BIS(4-(2;3-DIDEHYDROTRIAZIRIDINYL)PHENYL)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.711	0.790	0.820
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	28ZPAK-;123;72

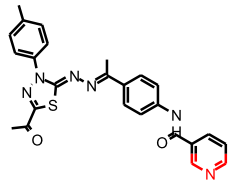
Model Applicability

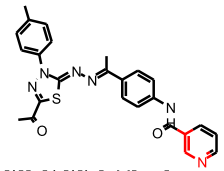
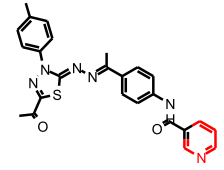
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

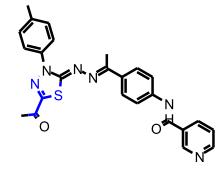
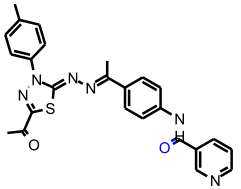
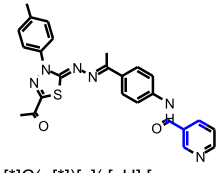
Feature Contribution

Top features for positive contribution

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

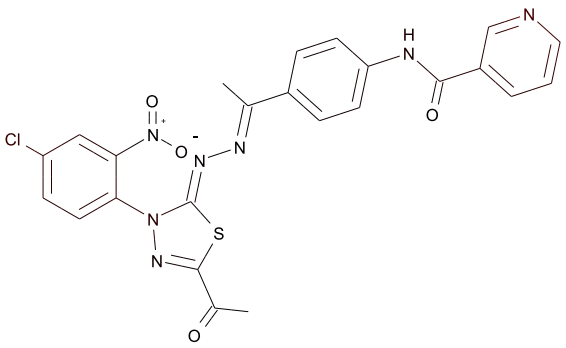
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	0.194	11 out of 11

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.0662	198 out of 262
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382

7e

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

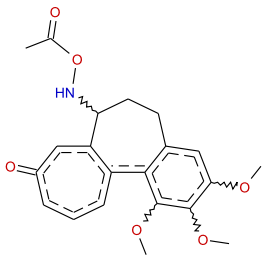
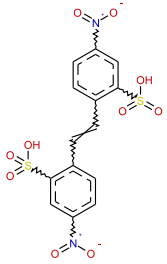
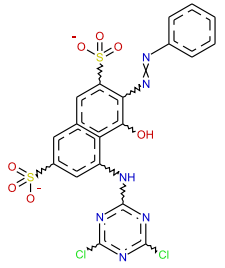
ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Structural Similar Compounds

Name	COLCHICINE	2;2'-Stilbenedisulfonic acid; 4;4'-dinitro-	2;7-Naphthalenedisulfonic acid;
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.909	0.947	0.958
Reference	AJOPAA 31;837;48	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 1062;86	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. - ;1300;86

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

Bayesian Score: 3.67

Mahalanobis Distance: 8.87

Mahalanobis Distance p-value: 0.585

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

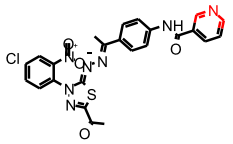
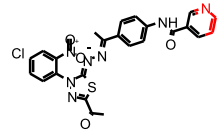
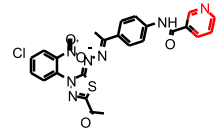
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
3. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
4. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
5. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

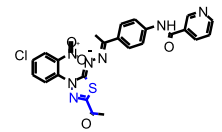
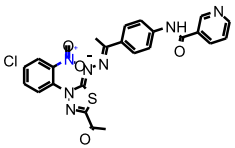
Feature Contribution

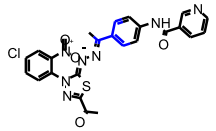
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

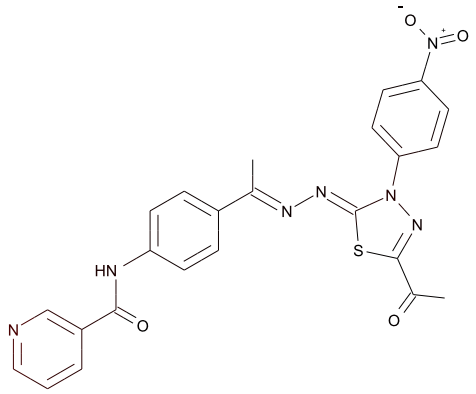
FCFP_12	1747237384	 [*]:[cH]:n:[cH]:[*]	0.208	44 out of 44
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.194	11 out of 11

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.0662	198 out of 262
FCFP_12	8	 [*][N+](=[*])[*]	-0.0561	3 out of 4

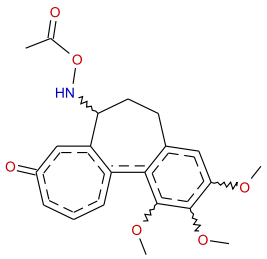
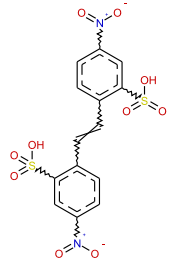
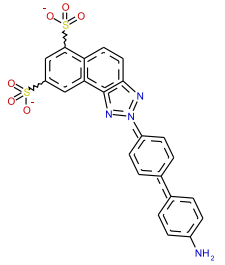
FCFP_12	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Structural Similar Compounds

Name	COLCHICINE	2;2'-Stilbenedisulfonic acid; 4;4'-dinitro-	2H-Naphtho(1;2-d)triazole-6;8-disulfonic acid;
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.869	0.895	0.905
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 Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 2.8
Mahalanobis Distance: 8.74
Mahalanobis Distance p-value: 0.65

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

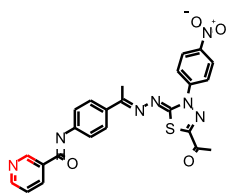
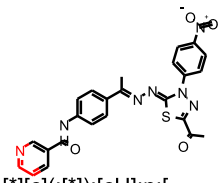
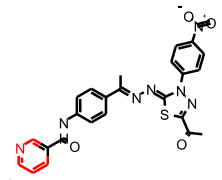
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
3. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
4. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
5. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

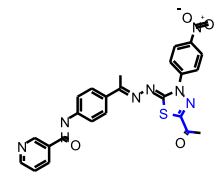
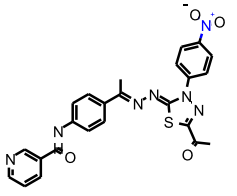
Feature Contribution

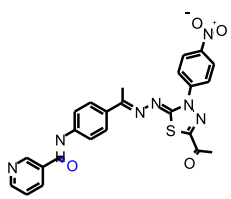
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	1747237384	 [*]:[cH]:n:[cH]:[*]	0.208	44 out of 44
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.194	11 out of 11

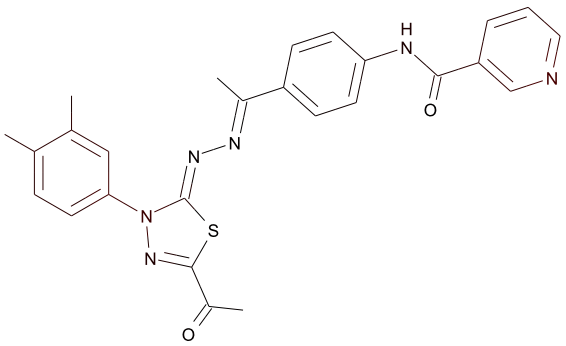
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.0662	198 out of 262
FCFP_12	8	 [*][N+](=[*])[*]	-0.0561	3 out of 4

FCFP_12	1	 [*]=O	0	872 out of 1051

7g

TOPKAT_Ocular_Irritancy_None_vs_Irritant



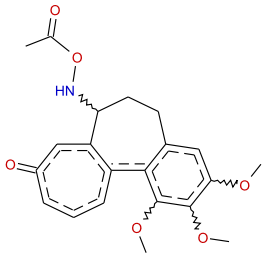
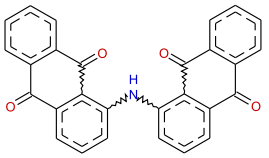
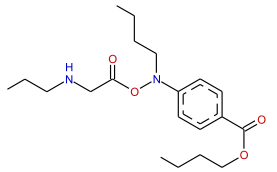
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 3.54
Mahalanobis Distance: 8.13
Mahalanobis Distance p-value: 0.895

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

Structural Similar Compounds

Name	COLCHICINE	ANTHRAQUINONE; 1;1'-IMINODI-	Benzoic acid; p-(N-butyl-2-(butylamino)acetamido)-; butyl ester;
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.748	0.781	0.830
Reference	AJOPAA 31;837;48	28ZPAK-;125;72	Arzneimittel-Forschung 8;609;58

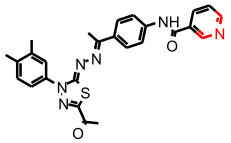
Model Applicability

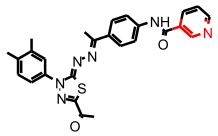
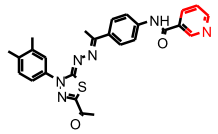
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

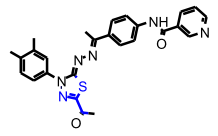
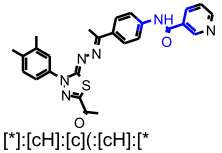
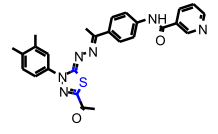
Feature Contribution

Top features for positive contribution

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

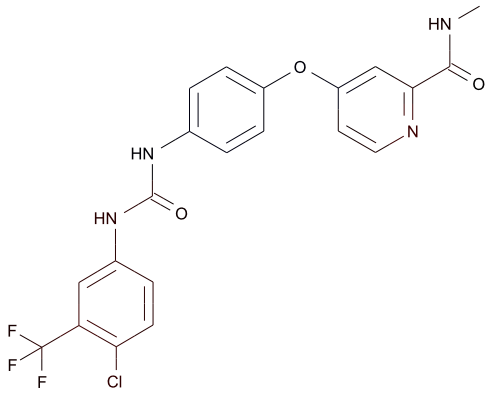
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.2	16 out of 16
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	0.194	11 out of 11

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.0662	198 out of 262
FCFP_12	-1925475824	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	0	4 out of 5
FCFP_12	-1143715940	 [*]C1=[*][*]C(=[*])S1	0	454 out of 590

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

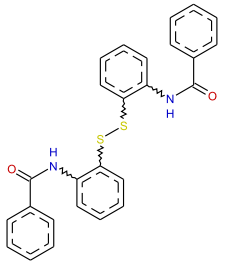
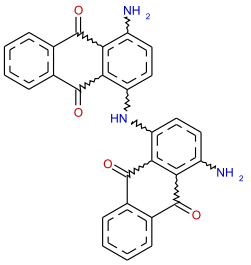
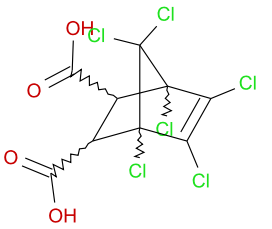
Bayesian Score: 3.04

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

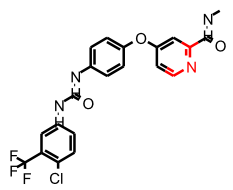
Model Applicability

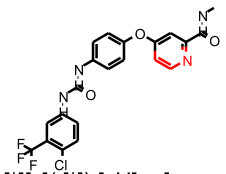
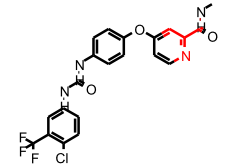
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

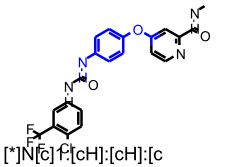
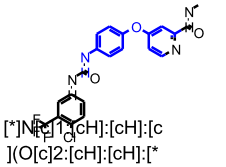
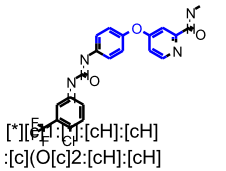
Feature Contribution

Top features for positive contribution

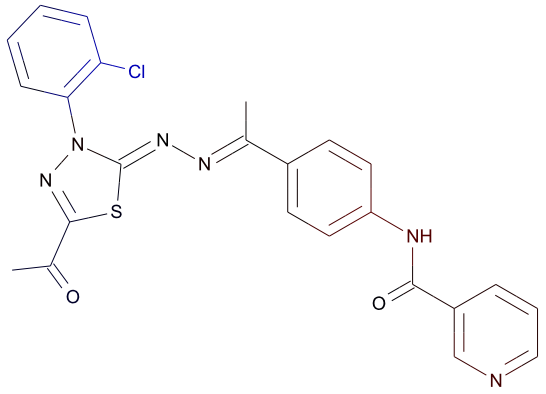
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*]:[cH]:n:[cH]:[*]	0.208	44 out of 44

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

Top Features for negative contribution

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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.255

Enrichment: 0.792

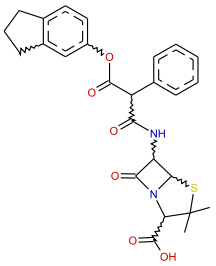
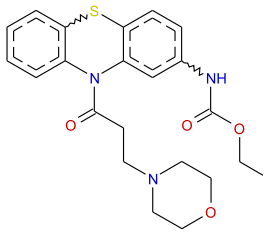
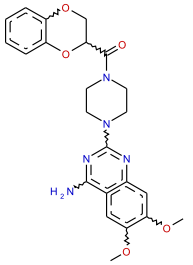
Bayesian Score: -2.44

Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 0.00042

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.666	0.698	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

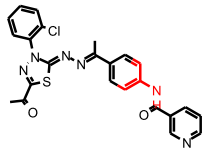
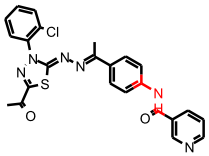
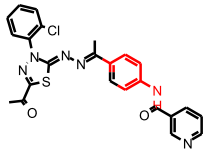
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

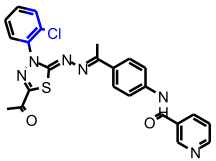
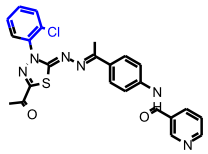
Feature Contribution

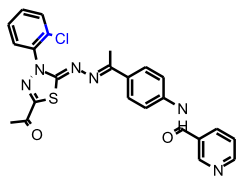
Top features for positive contribution

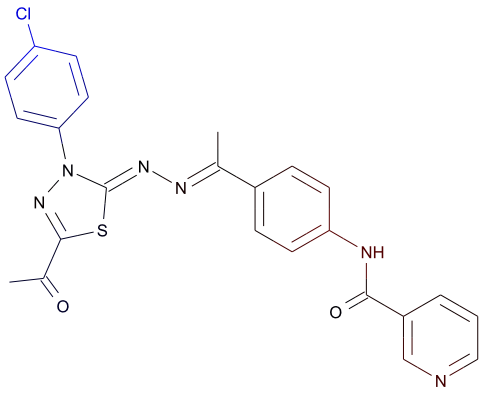
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

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	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

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



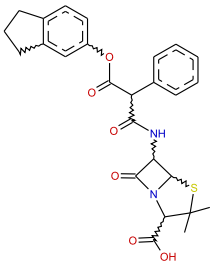
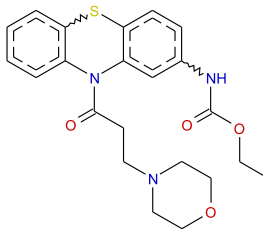
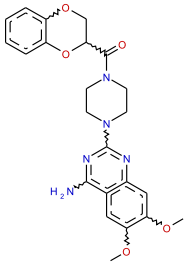
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96465
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.241
Enrichment: 0.749
Bayesian Score: -3.4
Mahalanobis Distance: 13.5
Mahalanobis Distance p-value: 2.05e-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	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.668	0.708	0.715
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

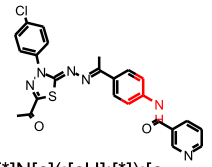
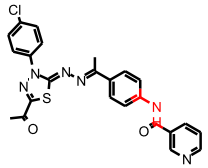
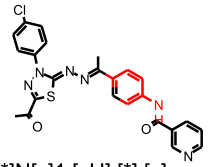
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

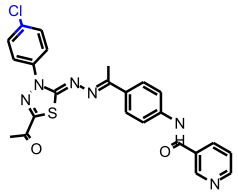
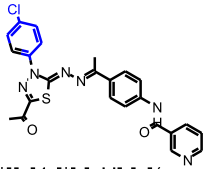
Feature Contribution

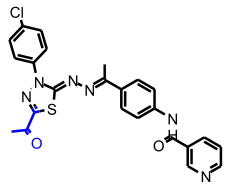
Top features for positive contribution

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

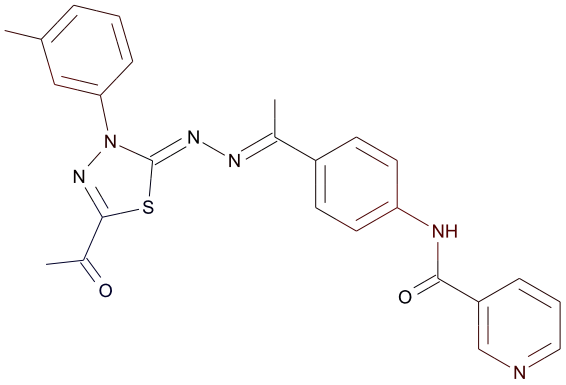
Top Features for negative contribution

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

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

7c

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



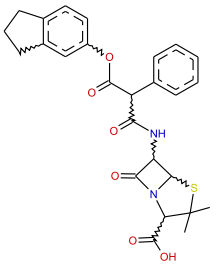
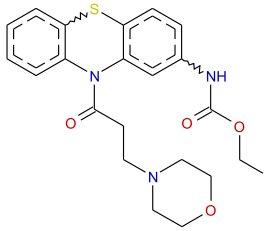
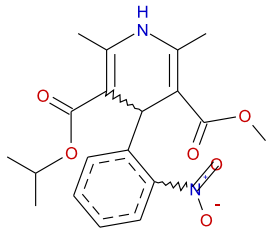
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.354
Enrichment: 1.1
Bayesian Score: 2.44
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 3.69e-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	Moricizine	Nisoldipine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.666	0.683	0.691
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

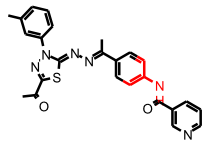
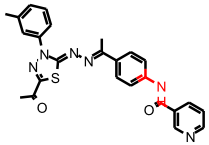
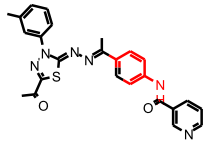
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

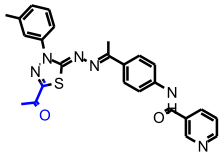
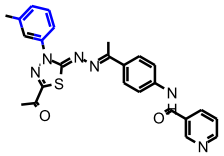
Feature Contribution

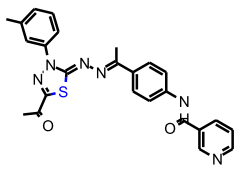
Top features for positive contribution

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

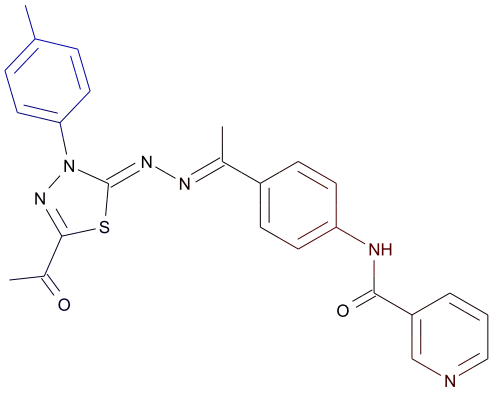
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4
ECFP_12	2007300961	 [*][c]1:[*]:[c]([*]):[cH]:[cH]:[cH]:1	-0.426	7 out of 36

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

7d

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



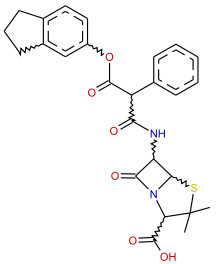
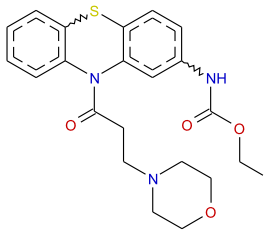
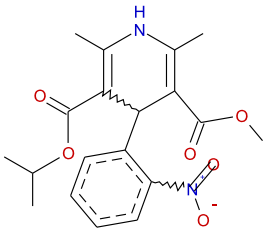
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.263
Enrichment: 0.818
Bayesian Score: -1.93
Mahalanobis Distance: 13.1
Mahalanobis Distance p-value: 8.3e-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	Moricizine	Nisoldipine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.668	0.686	0.691
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

Model Applicability

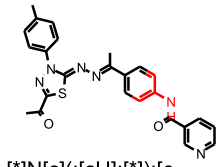
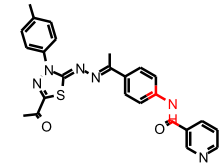
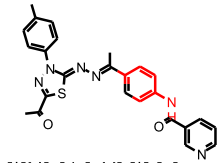
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

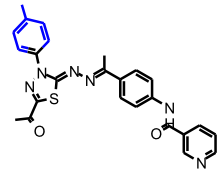
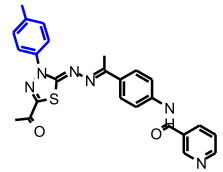
Feature Contribution

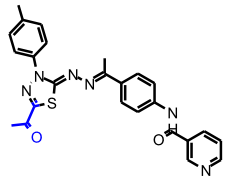
Top features for positive contribution

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

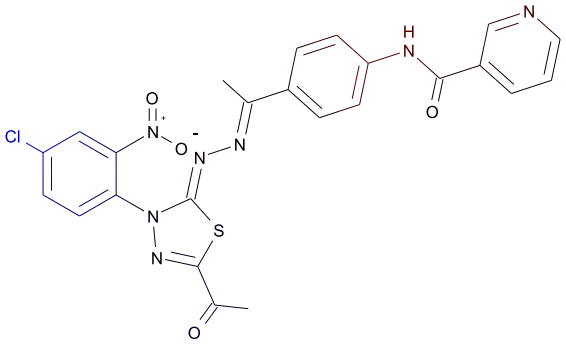
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-533780882	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.06	0 out of 6
ECFP_12	-1926229349	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.06	0 out of 6

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

7e

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.233

Enrichment: 0.725

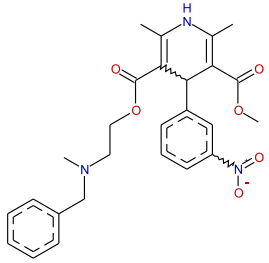
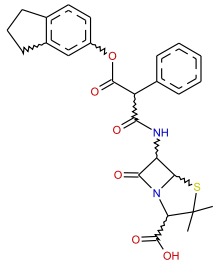
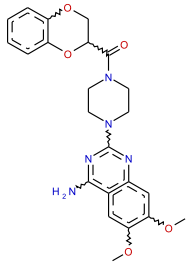
Bayesian Score: -3.98

Mahalanobis Distance: 14.1

Mahalanobis Distance p-value: 1.16e-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	Nicardipine	Carbenicillin	Doxazosin
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.784	0.793	0.794
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

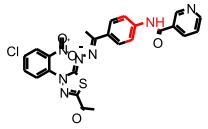
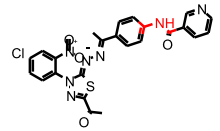
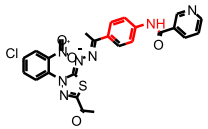
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1043790491: [*][N+](=[*])[*]
3. Unknown ECFP_2 feature: 781519895: [*][O-]
4. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
5. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
6. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
7. Unknown ECFP_2 feature: -1956535100: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
8. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
9. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
10. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
11. Unknown ECFP_2 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
12. Unknown ECFP_2 feature: 2104376220: [*][N+](=O)[*]
13. Unknown ECFP_2 feature: -659271057: [*][N+](=[*])[O-]

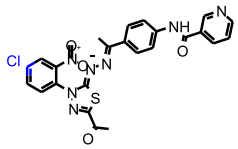
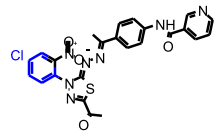
Feature Contribution

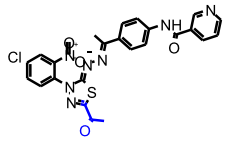
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

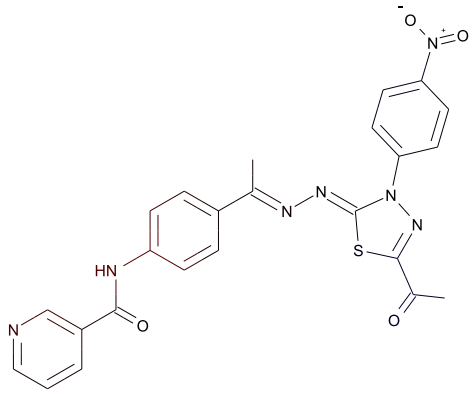
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17
ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9

Top Features for negative contribution

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

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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



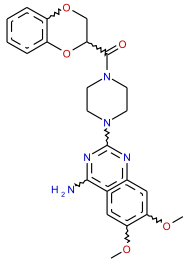
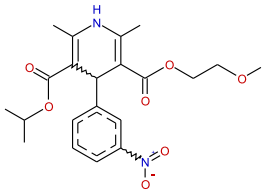
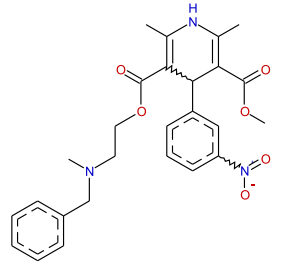
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.292
Enrichment: 0.906
Bayesian Score: -0.345
Mahalanobis Distance: 13.6
Mahalanobis Distance p-value: 1.21e-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	Doxazosin	Nimodipine	Nicardipine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.751	0.762	0.784
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

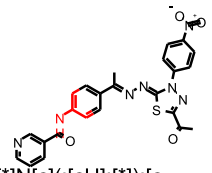
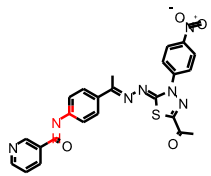
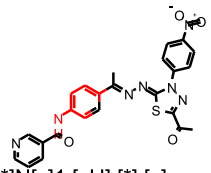
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1043790491: [*][N+](=[*])[*]
3. Unknown ECFP_2 feature: 781519895: [*][O-]
4. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
5. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
6. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
7. Unknown ECFP_2 feature: -179073144: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
8. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
9. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
10. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
11. Unknown ECFP_2 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
12. Unknown ECFP_2 feature: 2104376220: [*][N+](=O)[*]
13. Unknown ECFP_2 feature: -659271057: [*][N+](=[*])[O-]

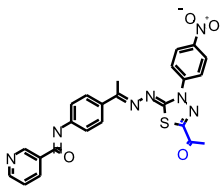
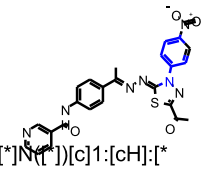
Feature Contribution

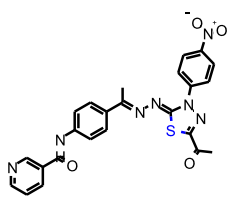
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17
ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9

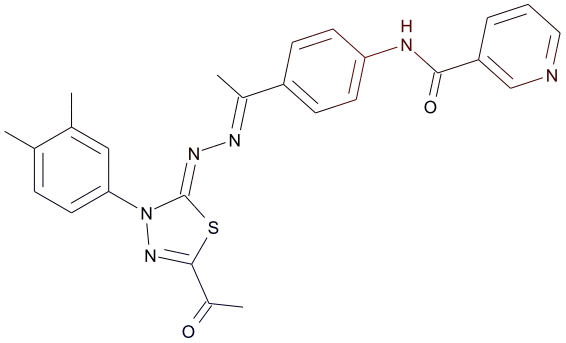
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4
ECFP_12	-2137232509	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.485	0 out of 2

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

7g

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



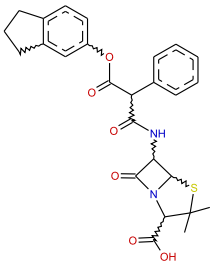
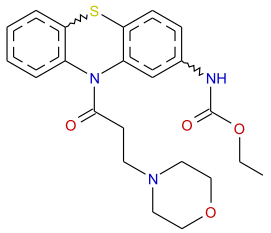
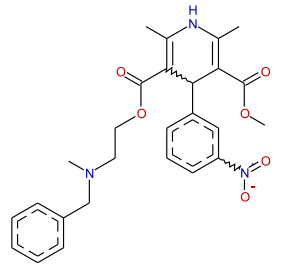
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.325
Enrichment: 1.01
Bayesian Score: 1.23
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 3.8e-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	Moricizine	Nicardipine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.682	0.714	0.728
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

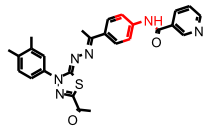
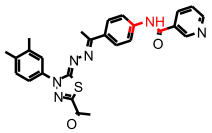
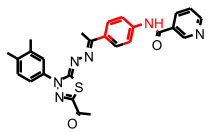
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

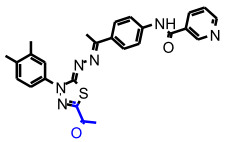
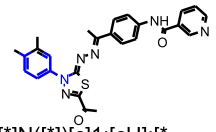
Feature Contribution

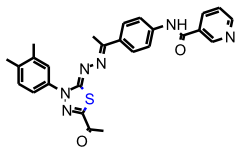
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set

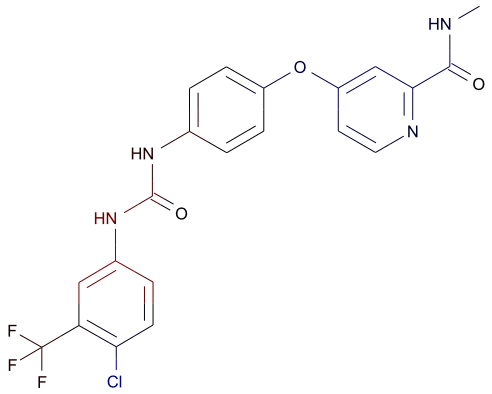
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17
ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	129482634	 [*]C(=[*])C(=O)C	-0.811	0 out of 4
ECFP_12	-2137232509	 [*]N([*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.485	0 out of 2

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

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

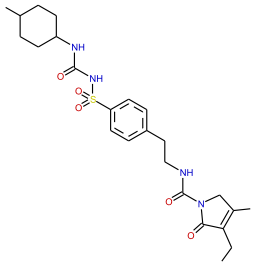
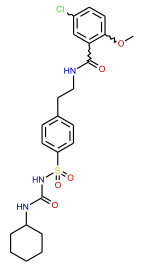
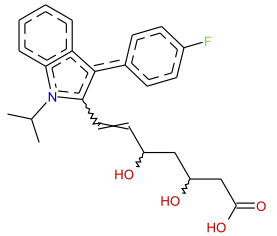
Mahalanobis Distance: 12.2

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

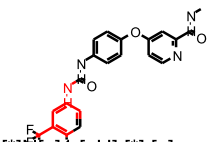
Model Applicability

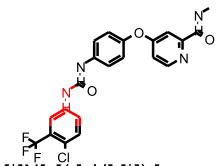
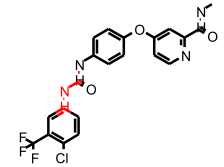
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

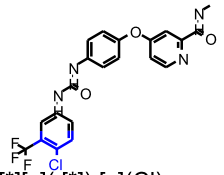
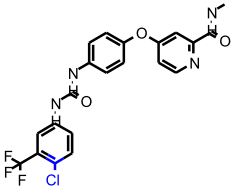
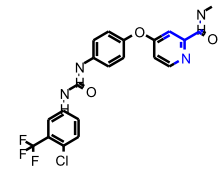
Feature Contribution

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

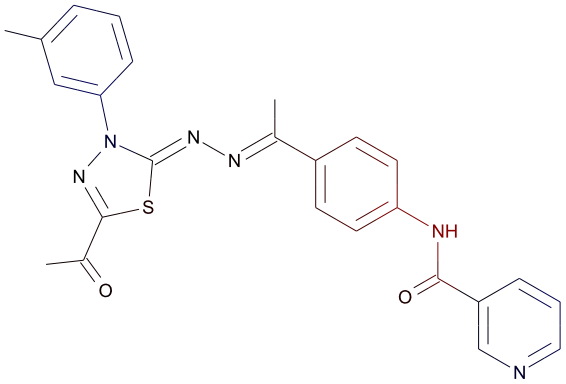
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10
ECFP_12	-1236483485	 [*]C(=[*])N[c](:[*]):[*]	0.46	9 out of 17

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

7c

TOPKAT_Rat_Female_FDA_Single_vs_Multiple



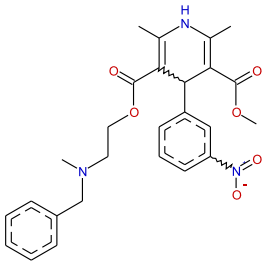
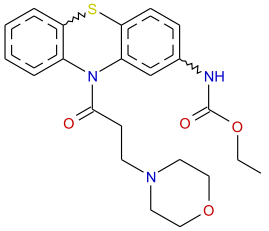
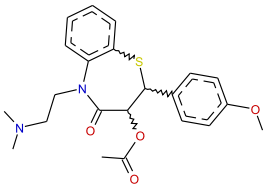
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.537
Enrichment: 1.44
Bayesian Score: 1.89
Mahalanobis Distance: 17.3
Mahalanobis Distance p-value: 1.75e-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	Nicardipine	Moricizine	Diltiazem
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.660	0.660	0.738
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

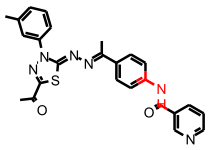
Model Applicability

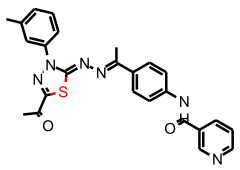
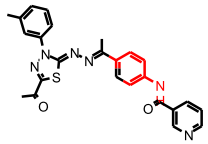
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: 3.5752. Training min, max, SD, explained variance: -2.8991, 3.0113, 1.313, 0.0255.

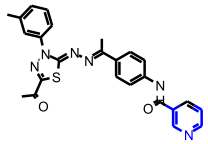
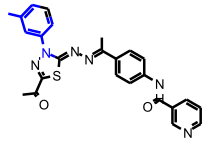
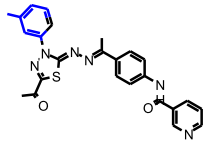
Feature Contribution

Top features for positive contribution

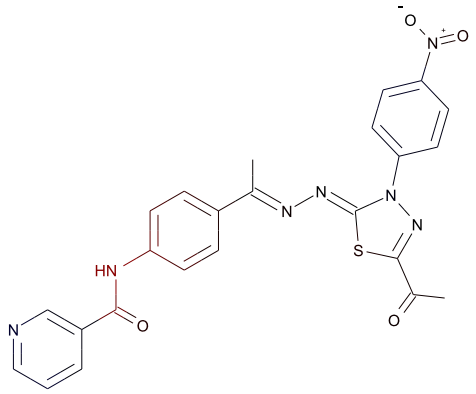
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	1631845520	 [*]C(=[*])N[c](:[*]):[*]	0.601	6 out of 9

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	1188429584	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	-0.666	0 out of 3
SCFP_4	-352263424	 [*]N([*])[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	-0.489	0 out of 2
SCFP_4	-1212387212	 [*][c]1:[*]:[cH]:[cH]:[c](C):[cH]:1	-0.489	0 out of 2

TOPKAT_Rat_Female_FDA_Single_vs_Multiple



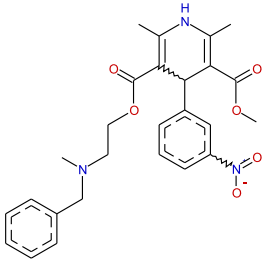
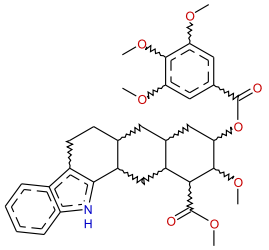
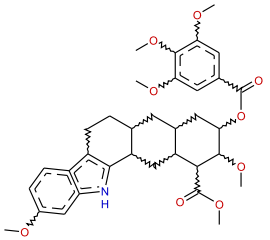
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.632
Enrichment: 1.69
Bayesian Score: 3.92
Mahalanobis Distance: 23.4
Mahalanobis Distance p-value: 5.51e-014

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

Structural Similar Compounds

Name	Nicardipine	Deserpidine	Reserpine
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.746	0.837	0.839
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

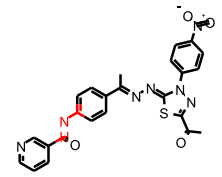
Model Applicability

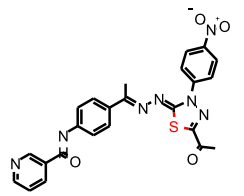
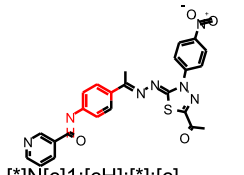
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Num_H_Acceptors out of range. Value: 10. Training min, max, mean, SD: 0, 9, 3.9512, 2.048.

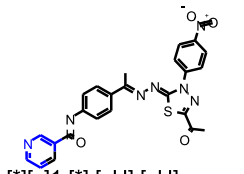
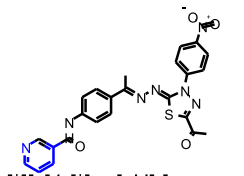
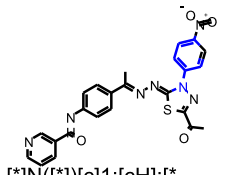
Feature Contribution

Top features for positive contribution

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

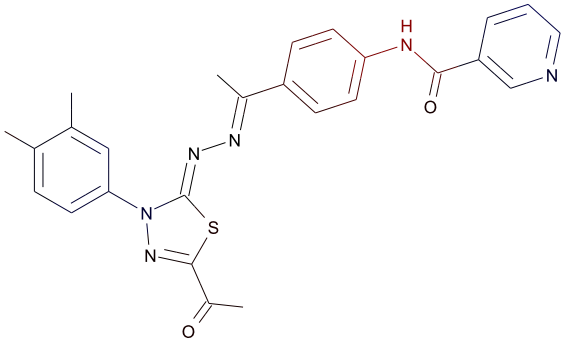
SCFP_4	17	 [*]S[*]	0.548	10 out of 17
SCFP_4	-1375926917	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.522	6 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	1188429584	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	-0.666	0 out of 3
SCFP_4	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.489	0 out of 2
SCFP_4	-1380909229	 [*]N([*])[c]1:[cH]:[*]:[cH]:[cH]:[cH]:1	-0.413	3 out of 16

7g

TOPKAT_Rat_Female_FDA_Single_vs_Multiple



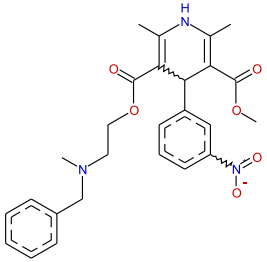
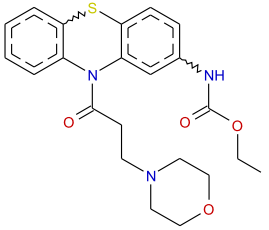
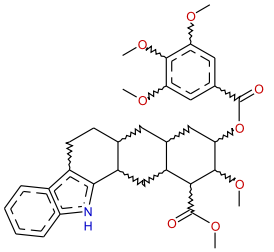
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.565
Enrichment: 1.51
Bayesian Score: 2.47
Mahalanobis Distance: 17
Mahalanobis Distance p-value: 3.51e-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	Nicardipine	Moricizine	Deserpidine
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.669	0.693	0.719
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

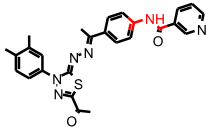
Model Applicability

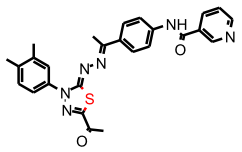
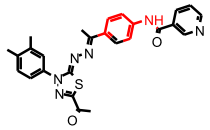
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: 3.5196. Training min, max, SD, explained variance: -2.8991, 3.0113, 1.313, 0.0255.

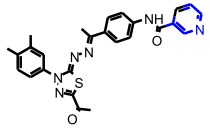
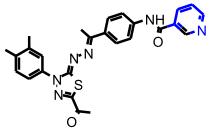
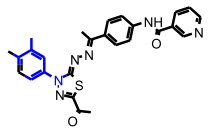
Feature Contribution

Top features for positive contribution

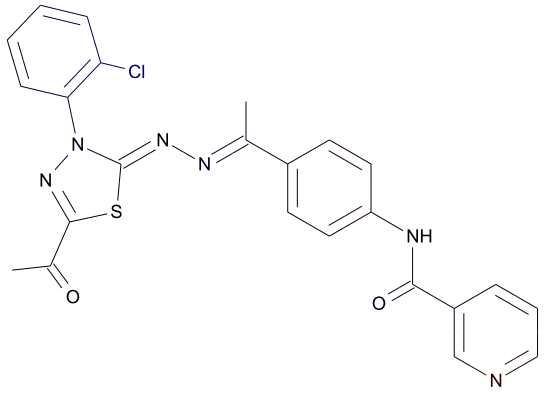
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	1631845520	 [*]C(=[*])N[c](:[*]):[*]	0.601	6 out of 9

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	1188429584	 [*][c]1:[*]:[cH]:[cH]:n:[cH]:1	-0.666	0 out of 3
SCFP_4	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-0.489	0 out of 2
SCFP_4	-352263424	 [*]N([*])[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	-0.489	0 out of 2

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.333

Enrichment: 0.996

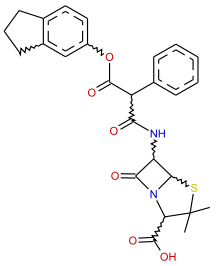
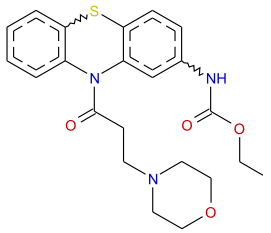
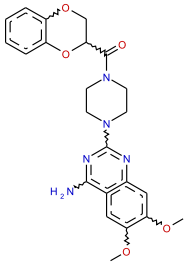
Bayesian Score: -0.84

Mahalanobis Distance: 17.1

Mahalanobis Distance p-value: 1.78e-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	Carbenicillin	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.658	0.690	0.695
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

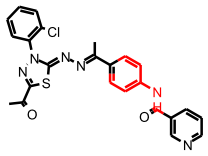
Model Applicability

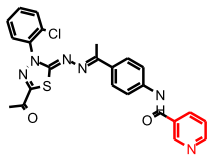
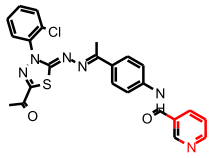
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC1 out of range. Value: -5.8913. Training min, max, SD, explained variance: -5.694, 8.348, 2.949, 0.0811.

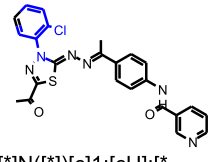
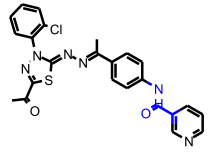
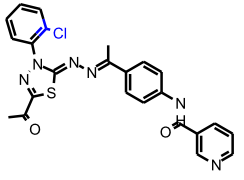
Feature Contribution

Top features for positive contribution

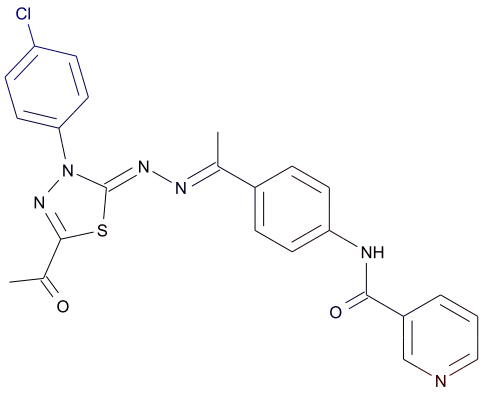
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.615	5 out of 7

SCFP_6	-105808146	 [*][c]1:[cH]:[cH]:[cH]:n:[cH]:1	0.415	1 out of 1
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	0.355	5 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1557430596	 [*]N([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	-0.825	0 out of 4
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	-0.436	4 out of 21
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.322

Enrichment: 0.963

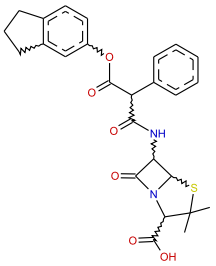
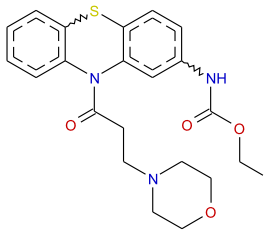
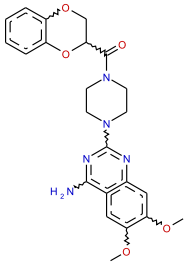
Bayesian Score: -1.25

Mahalanobis Distance: 16.5

Mahalanobis Distance p-value: 4.3e-010

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

Structural Similar Compounds

Name	Carbenicillin	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.661	0.694	0.698
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

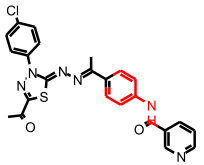
Model Applicability

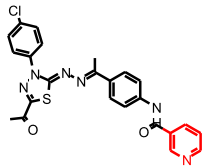
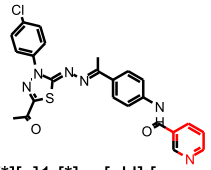
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC1 out of range. Value: -6.1431. Training min, max, SD, explained variance: -5.694, 8.348, 2.949, 0.0811.

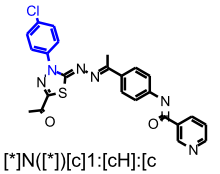
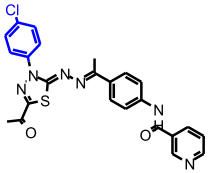
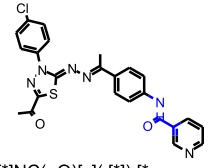
Feature Contribution

Top features for positive contribution

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

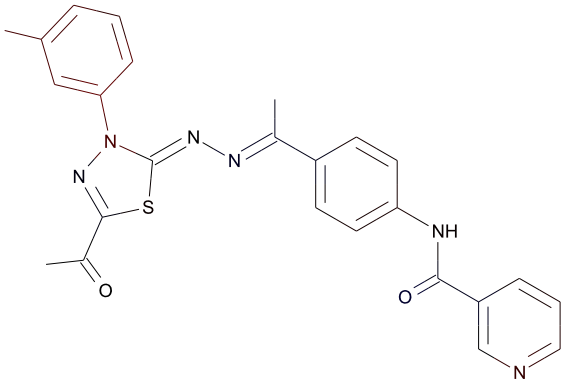
SCFP_6	-105808146	 [*][c]1:[cH]:[cH]:[cH]:n:[cH]:1	0.415	1 out of 1
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	0.355	5 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	341480432	 [*]N([*])[c]1:[cH]:[cH]:[c]([Cl]):[cH]:[cH]:1	-0.674	0 out of 3
SCFP_6	1905487031	 [*][c]1:[cH]:[cH]:[c]([Cl]):[cH]:[cH]:1	-0.48	2 out of 12
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	-0.436	4 out of 21

7c

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



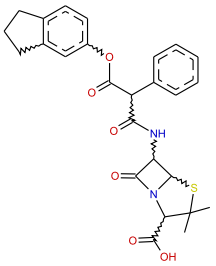
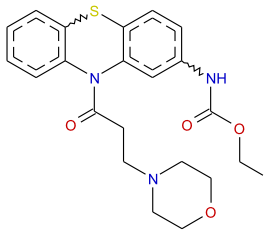
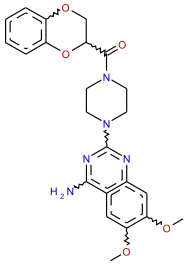
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.423
Enrichment: 1.27
Bayesian Score: 2.15
Mahalanobis Distance: 17.3
Mahalanobis Distance p-value: 7.26e-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	Carbenicillin	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.650	0.668	0.675
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

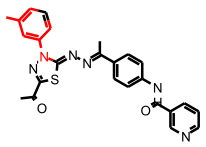
Model Applicability

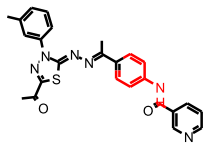
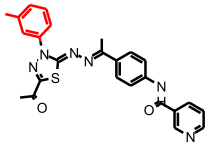
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: -4.4024. Training min, max, SD, explained variance: -3.9196, 6.4101, 1.581, 0.0233.

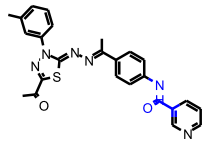
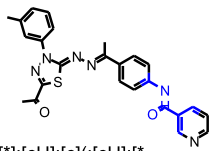
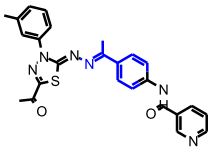
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-352263424	 [*]N([*])[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	0.712	3 out of 3

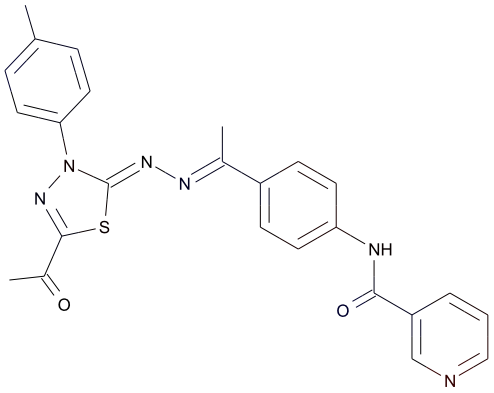
SCFP_6	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.615	5 out of 7
SCFP_6	2052999617	 [*][c]1:[cH]:[cH]:[cH]:[c](C):[cH]:1	0.415	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	-0.436	4 out of 21
SCFP_6	124026986	 [*]:[cH]:[c](:[cH]:[*])C(=O)N[c](:[*]):[*]	-0.278	0 out of 1
SCFP_6	-537745313	 [*]\N=C(/C)\[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.278	0 out of 1

7d

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.363

Enrichment: 1.09

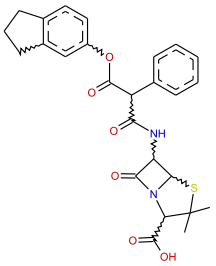
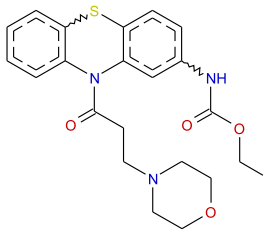
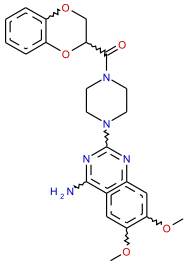
Bayesian Score: 0.226

Mahalanobis Distance: 17.1

Mahalanobis Distance p-value: 1.71e-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	Carbenicillin	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.653	0.671	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

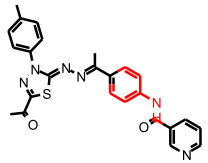
Model Applicability

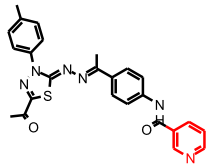
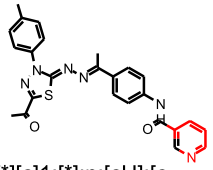
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: -4.169. Training min, max, SD, explained variance: -3.9196, 6.4101, 1.581, 0.0233.

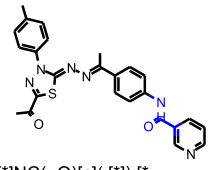
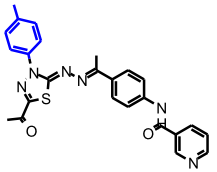
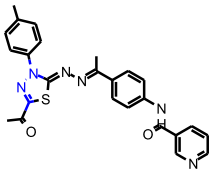
Feature Contribution

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

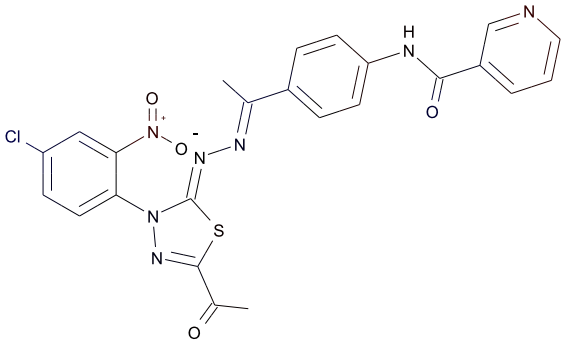
SCFP_6	-105808146	 [*][c]1:[cH]:[cH]:[cH]:n:[cH]:1	0.415	1 out of 1
SCFP_6	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	0.355	5 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	-0.436	4 out of 21
SCFP_6	795925860	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.38	1 out of 6
SCFP_6	-1325991669	 [*]N1[*][*]C(=N1)[*]	-0.278	0 out of 1

7e

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.319

Enrichment: 0.955

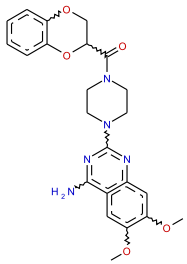
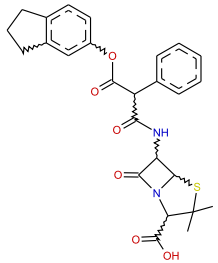
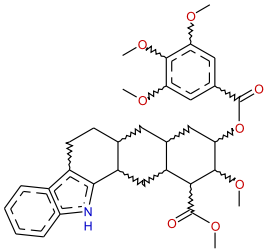
Bayesian Score: -1.36

Mahalanobis Distance: 19.8

Mahalanobis Distance p-value: 3e-018

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

Structural Similar Compounds

Name	Doxazosin	Carbenicillin	Deserpidine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.780	0.787	0.794
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

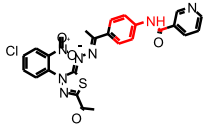
Model Applicability

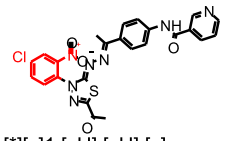
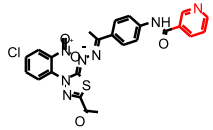
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC1 out of range. Value: -6.2502. Training min, max, SD, explained variance: -5.694, 8.348, 2.949, 0.0811.

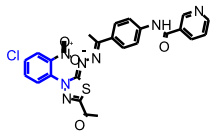
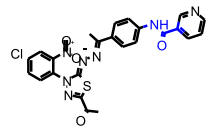
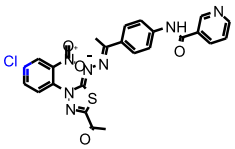
Feature Contribution

Top features for positive contribution

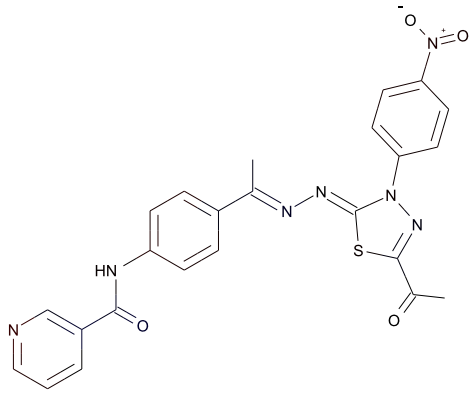
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.615	5 out of 7

SCFP_6	1907765814	 [*][c]1:[cH]:[cH]:[c](Cl):[cH]:[c]:1[N+](=[*])[*]	0.603	2 out of 2
SCFP_6	-105808146	 [*][c]1:[cH]:[cH]:[cH]:n:[cH]:1	0.415	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	341480432	 [*]N([*])[c]1:[cH]:[cH]:[c](Cl):[cH]:[cH]:1	-0.674	0 out of 3
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	-0.436	4 out of 21
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



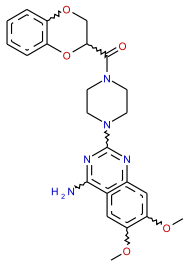
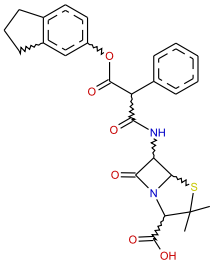
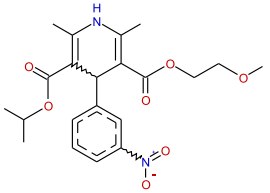
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.345
Enrichment: 1.03
Bayesian Score: -0.393
Mahalanobis Distance: 20.1
Mahalanobis Distance p-value: 6.54e-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	Doxazosin	Carbenicillin	Nimodipine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.734	0.776	0.783
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

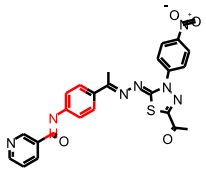
Model Applicability

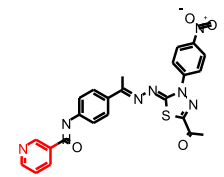
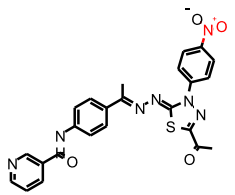
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: -4.384. Training min, max, SD, explained variance: -3.9196, 6.4101, 1.581, 0.0233.

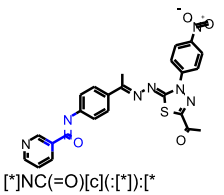
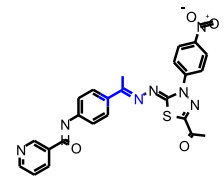
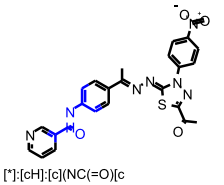
Feature Contribution

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

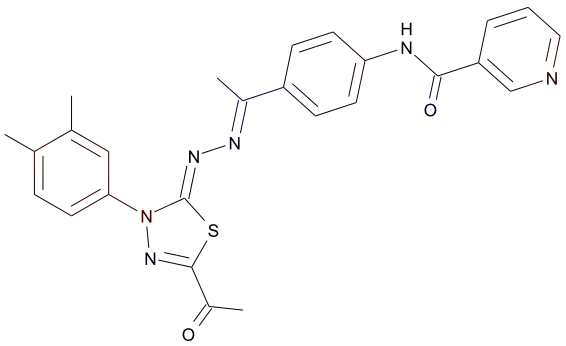
SCFP_6	-105808146	 [*][c]1:[cH]:[cH]:[cH]:n:[cH]:1	0.415	1 out of 1
SCFP_6	1311339974	 [*][N+](=O)[*]	0.405	13 out of 26

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	-0.436	4 out of 21
SCFP_6	-331724199	 [*]\N=C(/C)\[c](:[*]):[*]	-0.278	0 out of 1
SCFP_6	2096901122	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-0.278	0 out of 1

7g

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



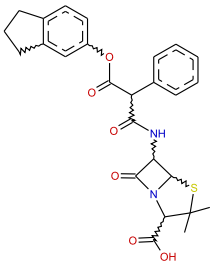
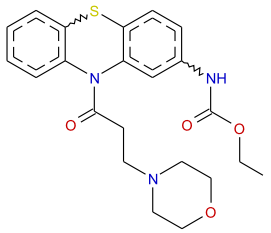
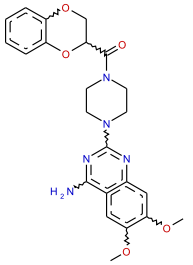
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.409
Enrichment: 1.22
Bayesian Score: 1.72
Mahalanobis Distance: 17.1
Mahalanobis Distance p-value: 1.48e-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	Carbenicillin	Moricizine	Doxazosin
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.669	0.702	0.716
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

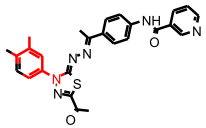
Model Applicability

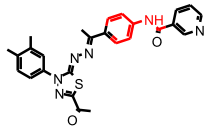
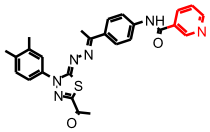
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC12 out of range. Value: -4.1817. Training min, max, SD, explained variance: -3.9196, 6.4101, 1.581, 0.0233.

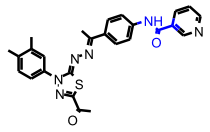
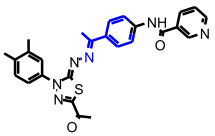
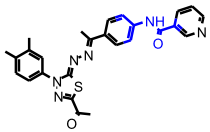
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-352263424	 [*]N([*])[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	0.712	3 out of 3

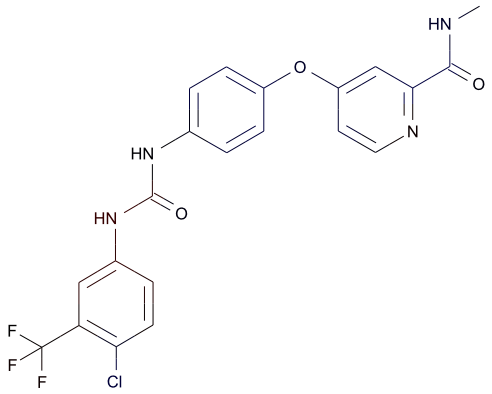
SCFP_6	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.615	5 out of 7
SCFP_6	-105808146	 [*][c]1:[cH]:[cH]:[cH]:n:[cH]:1	0.415	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1257084377	 [*]NC(=O)[c](:[*]):[*]	-0.436	4 out of 21
SCFP_6	-537745313	 [*]\N=C(/C)\[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.278	0 out of 1
SCFP_6	2096901122	 [*]:[cH]:[c](NC(=O)[c](:[*]):[*]):[cH]:[*]	-0.278	0 out of 1

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

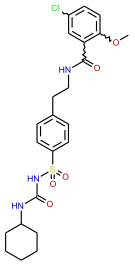
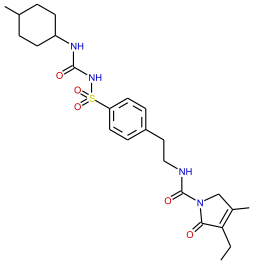
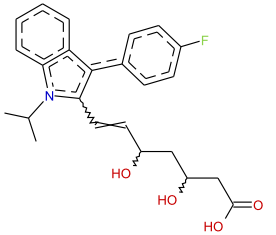
Bayesian Score: -2.4

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

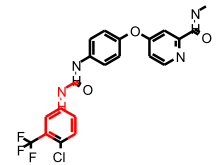
Model Applicability

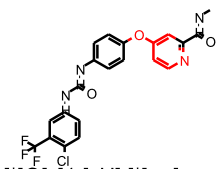
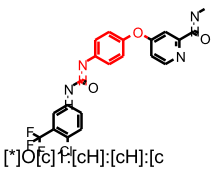
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

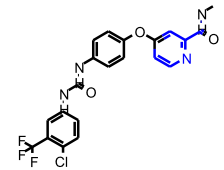
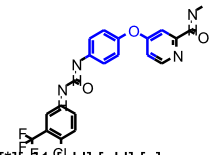
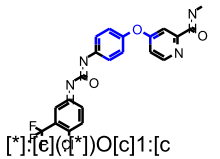
Feature Contribution

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

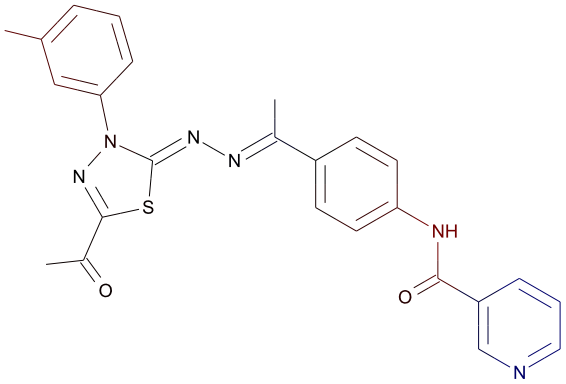
SCFP_6	-754059116	 [*]O[c]1:[cH]:[*]:n:[cH]:[cH]:1	0.415	1 out of 1
SCFP_6	-531283893	 [*]O[c]1:[cH]:[cH]:[c](NC(=[*])[*]):[cH]:[cH]:1	0.273	2 out of 4

Top Features for negative contribution

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

7c

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



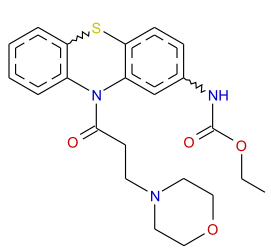
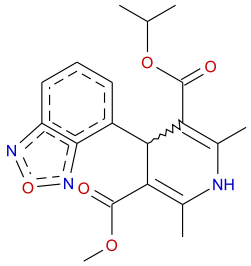
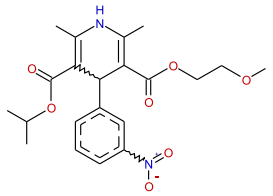
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.569
Enrichment: 1.37
Bayesian Score: 1.5
Mahalanobis Distance: 15.6
Mahalanobis Distance p-value: 3.18e-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	Moricizine	Isradipine	Nimodipine
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.692	0.745	0.783
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

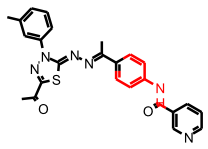
Model Applicability

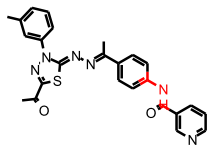
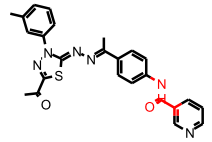
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: 3.6156. Training min, max, SD, explained variance: -3.4626, 3.2359, 1.286, 0.0225.

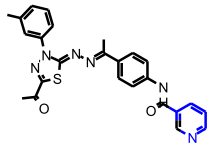
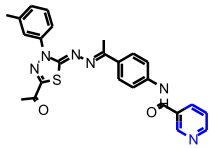
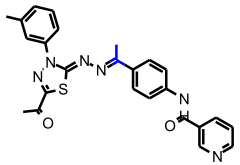
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.574	4 out of 5

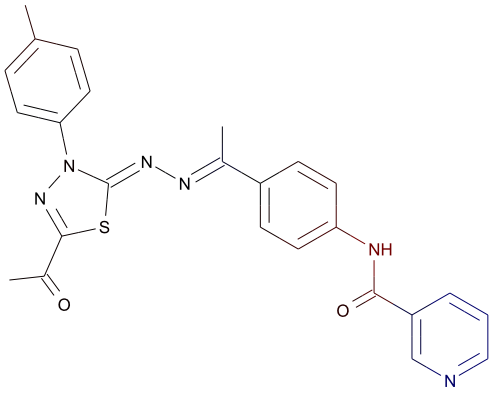
SCFP_8	1631845520	 [*]C(=[*])N[c](:[*]):[*]	0.495	6 out of 9
SCFP_8	1257084377	 [*]NC(=O)[c](:[*]):[*]	0.489	3 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-1.04	0 out of 5
SCFP_8	-937094999	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.463	1 out of 6
SCFP_8	136627117	 [*]C(=[*])C	-0.41	4 out of 18

7d

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.568

Enrichment: 1.37

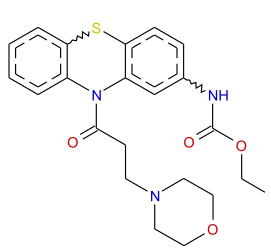
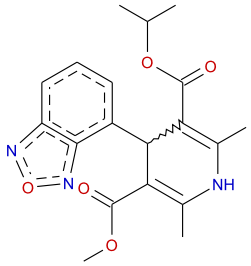
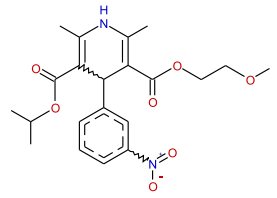
Bayesian Score: 0.701

Mahalanobis Distance: 17.1

Mahalanobis Distance p-value: 1.28e-007

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

Structural Similar Compounds

Name	Moricizine	Isradipine	Nimodipine
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.694	0.749	0.790
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

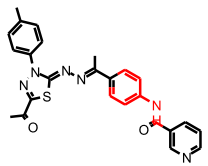
Model Applicability

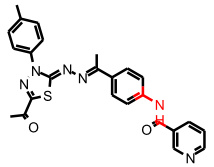
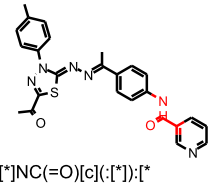
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: 3.4879. Training min, max, SD, explained variance: -3.4626, 3.2359, 1.286, 0.0225.

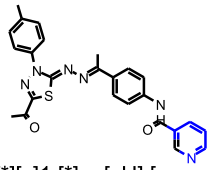
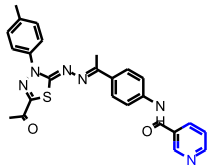
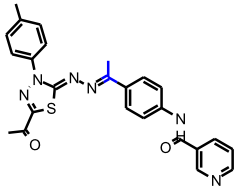
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.574	4 out of 5

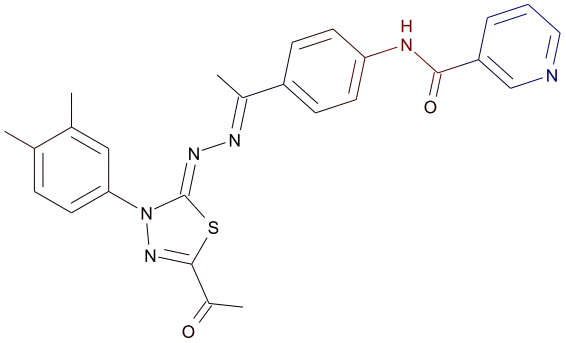
SCFP_8	1631845520	 [*]C(=[*])N[c](:[*]):[*]	0.495	6 out of 9
SCFP_8	1257084377	 [*]NC(=O)[c](:[*]):[*]	0.489	3 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:[cH]:1	-1.04	0 out of 5
SCFP_8	-937094999	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.463	1 out of 6
SCFP_8	136627117	 [*]C(=[*])C	-0.41	4 out of 18

7g

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



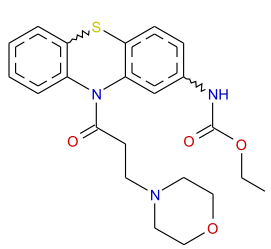
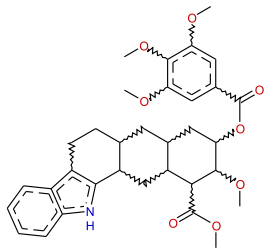
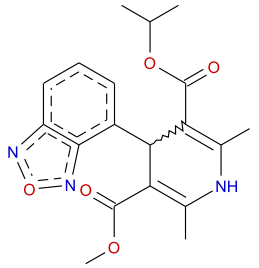
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.568
Enrichment: 1.37
Bayesian Score: 0.803
Mahalanobis Distance: 17.2
Mahalanobis Distance p-value: 1.13e-007

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

Structural Similar Compounds

Name	Moricizine	Deserpidine	Isradipine
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.728	0.765	0.801
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

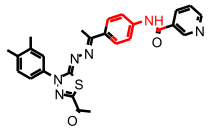
Model Applicability

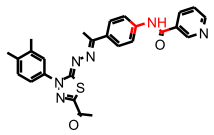
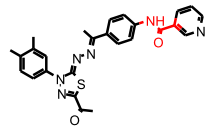
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC14 out of range. Value: 3.561. Training min, max, SD, explained variance: -3.4626, 3.2359, 1.286, 0.0225.

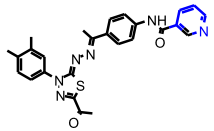
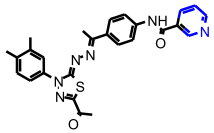
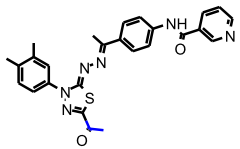
Feature Contribution

Top features for positive contribution

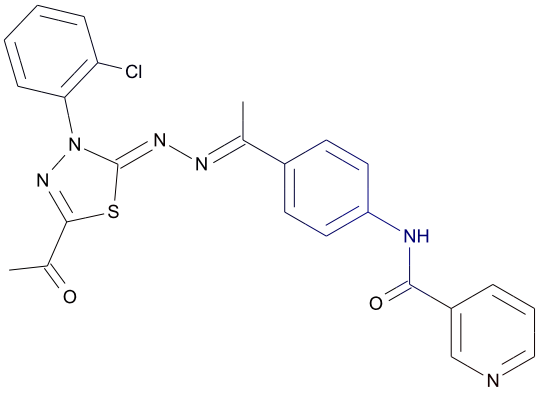
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-347048986	 [*]C(=[*])N[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.574	4 out of 5

SCFP_8	1631845520	 [*]C(=[*])N[c](:[*]):[*]	0.495	6 out of 9
SCFP_8	1257084377	 [*]NC(=O)[c](:[*]):[*]	0.489	3 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-758850909	 [*][c]1:[*]:n:[cH]:[cH]:[cH]:1	-1.04	0 out of 5
SCFP_8	-937094999	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	-0.463	1 out of 6
SCFP_8	136627117	 [*]C(=[*])C	-0.41	4 out of 18

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.9

Enrichment: 0.978

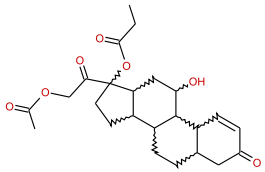
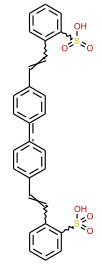
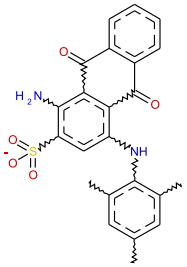
Bayesian Score: -2.75

Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.0042

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

Structural Similar Compounds

Name	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.757	0.802	0.834
Reference	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2-5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1-1972- Volume(issue) /page/year: 19,3103,1991	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986

Model Applicability

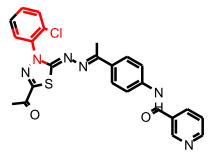
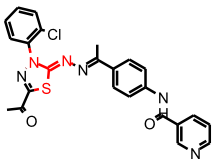
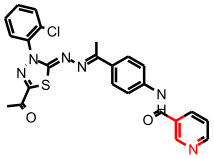
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

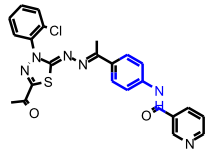
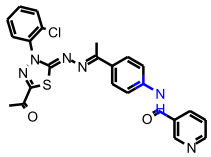
Feature Contribution

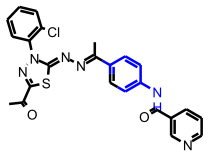
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

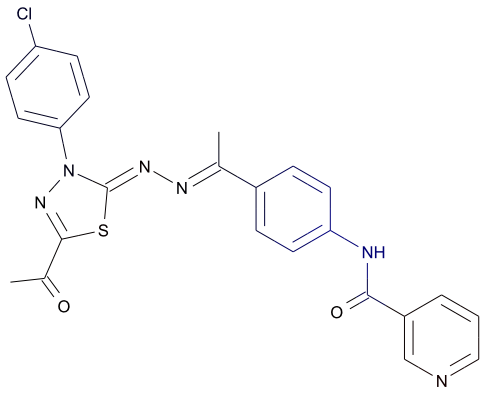
FCFP_12	-62776068	 [*]N([*])[c]1:[cH]:[*]:[cH]:[cH]:[c]:1Cl	0.0821	13 out of 13
FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.0821	13 out of 13

Top Features for negative contribution

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

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

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.864

Enrichment: 0.938

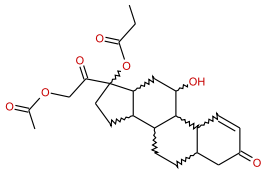
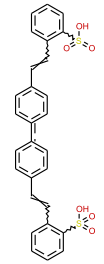
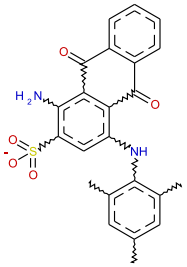
Bayesian Score: -3.1

Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.0042

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

Structural Similar Compounds

Name	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.758	0.806	0.837
Reference	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2-5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1-1972- Volume(issue) /page/year: 19,3103,1991	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986

Model Applicability

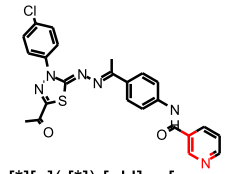
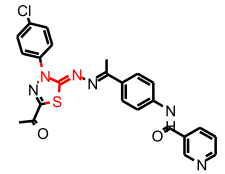
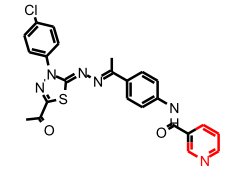
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

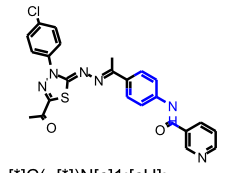
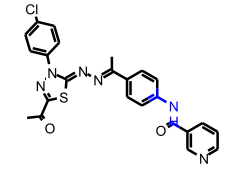
Feature Contribution

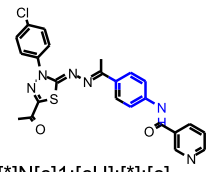
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

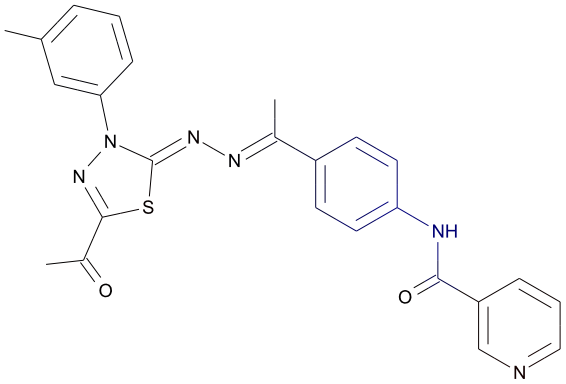
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	0.0772	7 out of 7

Top Features for negative contribution

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

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

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.928

Enrichment: 1.01

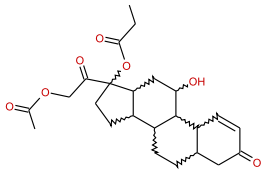
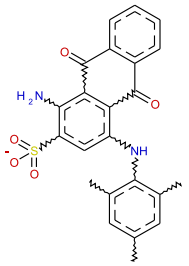
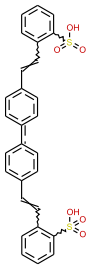
Bayesian Score: -2.37

Mahalanobis Distance: 10.4

Mahalanobis Distance p-value: 0.0209

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

Structural Similar Compounds

Name	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.729	0.807	0.819
Reference	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2-5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1-1972- Volume(issue) /page/year: 19,3103,1991	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973

Model Applicability

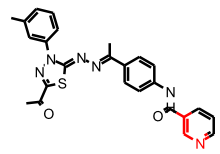
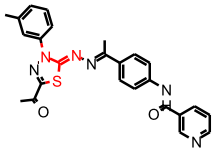
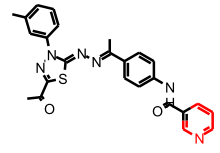
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

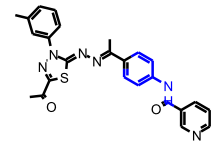
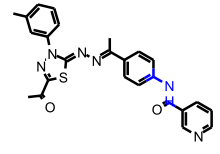
Feature Contribution

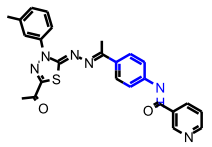
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	0.0772	7 out of 7

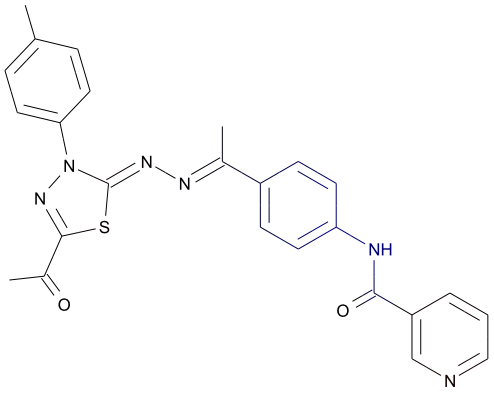
Top Features for negative contribution

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

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

7d

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.932

Enrichment: 1.01

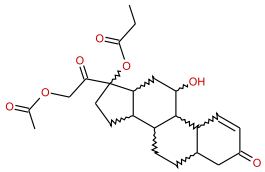
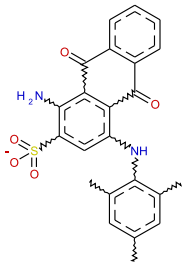
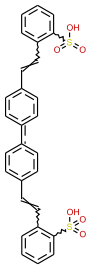
Bayesian Score: -2.29

Mahalanobis Distance: 10.4

Mahalanobis Distance p-value: 0.0209

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

Structural Similar Compounds

Name	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.729	0.811	0.823
Reference	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2-5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1-1972- Volume(issue) /page/year: 19,3103,1991	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volume(issue)/page/year: 2,193,1973

Model Applicability

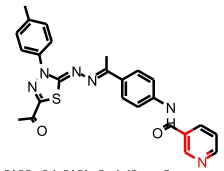
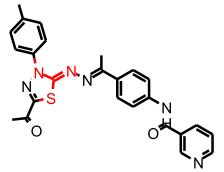
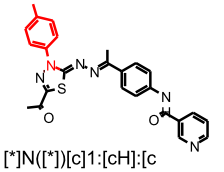
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

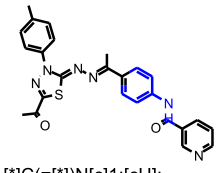
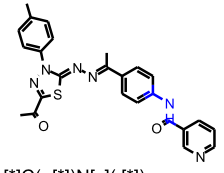
Feature Contribution

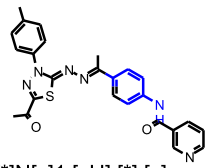
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

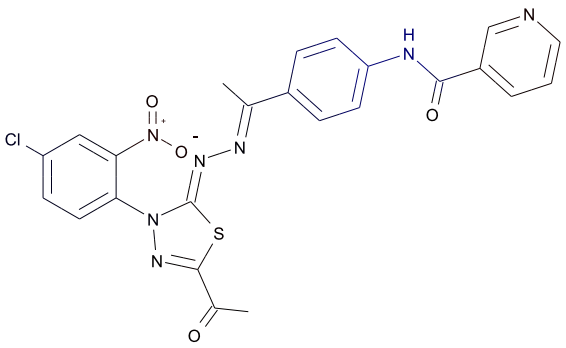
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	633795852	 [*]N([*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.0785	8 out of 8

Top Features for negative contribution

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

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

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.922

Enrichment: 1

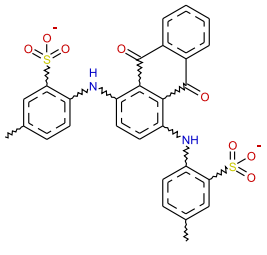
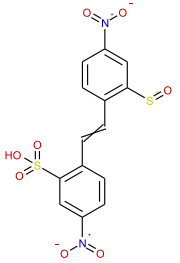
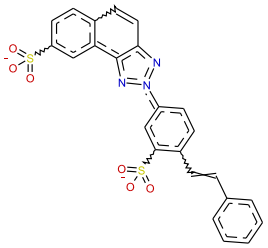
Bayesian Score: -2.46

Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 5.59e-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	Benzenesulfonic acid, 2,2'-(1,4-anthraquinonylenediimino)bis(5-meth yl-, disodium salt	2,2'-Stilbenedisulfonic acid, 4,4'-dinitro-	3-Picoline
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.847	0.949	0.954
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1330,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1062,1986	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: OTS0534823

Model Applicability

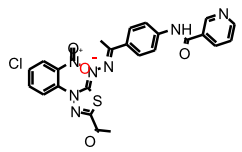
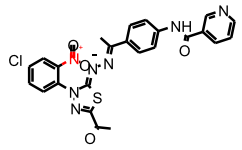
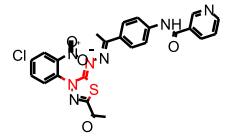
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

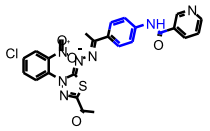
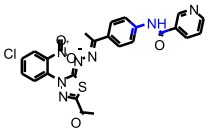
Feature Contribution

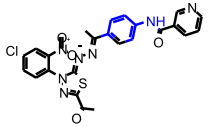
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

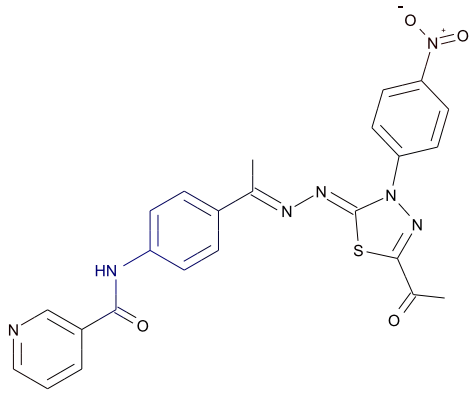
FCFP_12	5	 [*][O-]	0.0854	27 out of 27
FCFP_12	8	 [*][N+](=[*])[*]	0.0843	20 out of 20
FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13

Top Features for negative contribution

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

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

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{24}H_{19}N_7O_4S$

Molecular Weight: 501.51715

ALogP: 3.488

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.954

Enrichment: 1.04

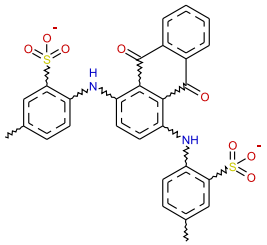
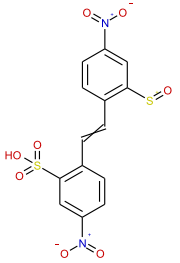
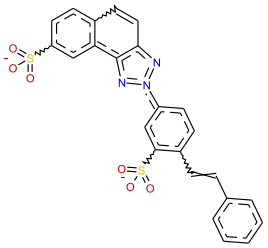
Bayesian Score: -1.76

Mahalanobis Distance: 11.4

Mahalanobis Distance p-value: 0.000572

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

Structural Similar Compounds

Name	Benzenesulfonic acid, 2,2'-(1,4-anthraquinonylenediimino)bis(5-meth yl-, disodium salt	2,2'-Stilbenedisulfonic acid, 4,4'-dinitro-	3-Picoline
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Irritant
Distance	0.825	0.886	0.903
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1330,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1062,1986	NTIS** National Technical Information Service. (Springfield, VA 22161) Forme rly U.S. Clearinghouse for Scientific & Technical Information. Volume(issue)/pag e/year: OTS0534823

Model Applicability

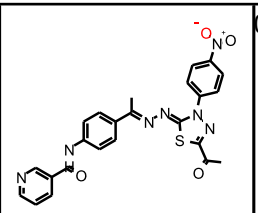
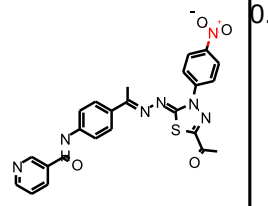
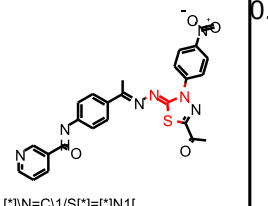
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

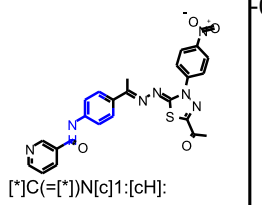
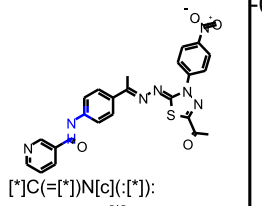
Feature Contribution

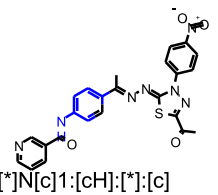
Top features for positive contribution

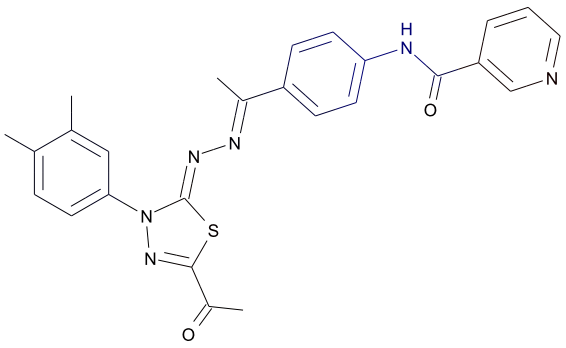
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	5	 [*][O-]	0.0854	27 out of 27
FCFP_12	8	 [*][N+](=[*])[*]	0.0843	20 out of 20
FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13

Top Features for negative contribution

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

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



$C_{26}H_{24}N_6O_2S$

Molecular Weight: 484.57275

ALogP: 4.566

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.917

Enrichment: 0.996

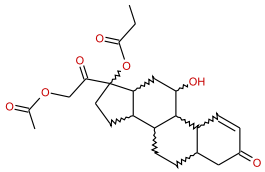
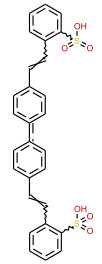
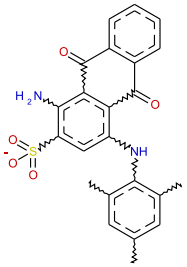
Bayesian Score: -2.54

Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.015

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

Structural Similar Compounds

Name	Pregna-1,4-diene-3,20-dione, 21-(acetyloxy)-11-hydroxy-6-methyl-17- (1-oxopropoxy)-, (6-alpha,11-beta)-	Benzenesulfonic acid, 2,2'-(4,4'-biphenylylenedivinylene)di-, disod ium salt	2-Anthracenesulfonic acid, 1-amino-9,10-dihydro-9,10-dioxo-4-(2,4,6 -trimethylanilino)-, monosodium salt
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.774	0.796	0.850
Reference	YACHDS Yakuri to Chiryo. Pharmacology and Therapeutics. (Raifu Saiensu Shup pan K.K., 2-5-13, Yaesu, Chuo-ku, Tokyo 104, Japan) V.1-1972- Volume(issue) /page/year: 19,3103,1991	MVCRB3 MVC-Report. (Stockholm, Sweden) No.1-2, 1972-73. Discontinued. Volu me(issue)/page/year: 2,193,1973	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,1327,1986

Model Applicability

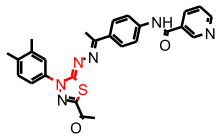
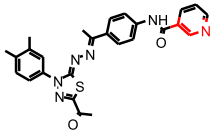
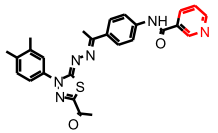
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

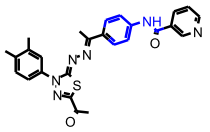
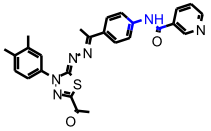
Feature Contribution

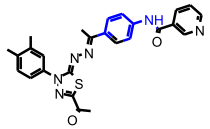
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1986158408	 [*]\N=C\1/S[*]=[*]N1[*]	0.0821	13 out of 13
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n :[cH]:1	0.0772	7 out of 7

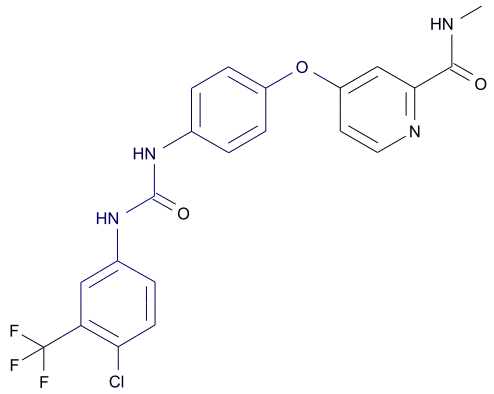
Top Features for negative contribution

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

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

Sorafenib

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Irritant

Probability: 0.264

Enrichment: 0.287

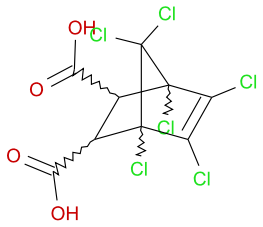
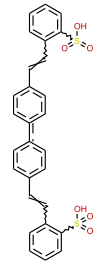
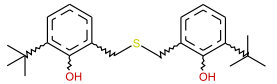
Bayesian Score: -5.23

Mahalanobis Distance: 8.27

Mahalanobis Distance p-value: 0.791

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

Structural Similar Compounds

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

Model Applicability

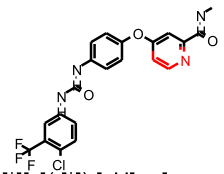
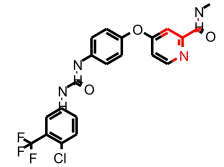
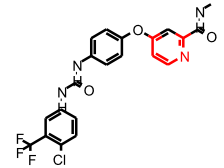
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

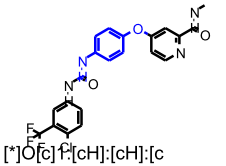
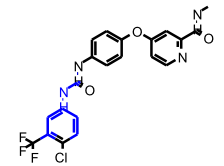
Feature Contribution

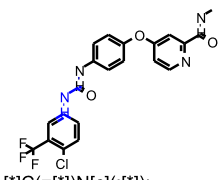
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

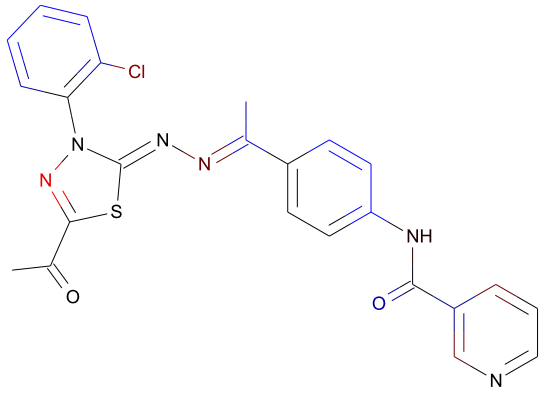
FCFP_12	-124655670	 [*][c](:[*]):[cH]:n:[*]	0.0821	13 out of 13
FCFP_12	-1539132615	 [*]C(=[*])[c](:[cH]:[*]):n:[*]	0.0795	9 out of 9
FCFP_12	-1695756380	 [*]1:[cH]:[cH]:[cH]:n:[cH]:1	0.0772	7 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-789307649	 [*]O[c]1:[cH]:[cH]:[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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



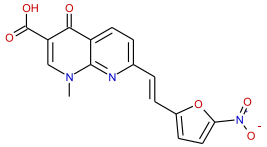
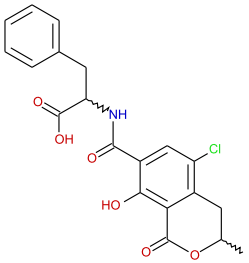
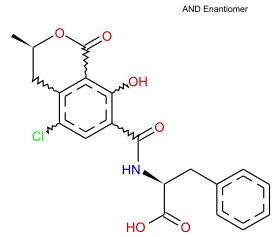
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96466
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 11
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.9
Mahalanobis Distance p-value: 5.16e-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	470	Ochratoxin A	542
Structure			
Actual Endpoint (-log C)	4.62839	4.79932	4.79932
Predicted Endpoint (-log C)	3.93264	3.6353	3.6353
Distance	0.805	0.825	0.825
Reference	CPDB	CPDB	CPDB

Model Applicability

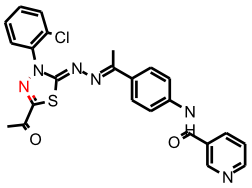
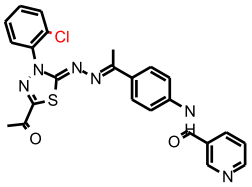
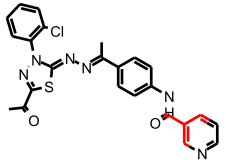
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
7. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
8. Unknown ECFP_2 feature: -1236953626: [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]

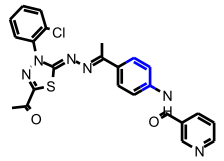
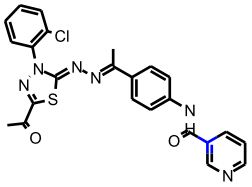
Feature Contribution

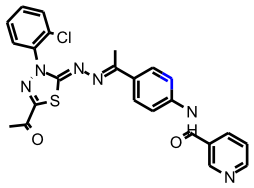
Top features for positive contribution

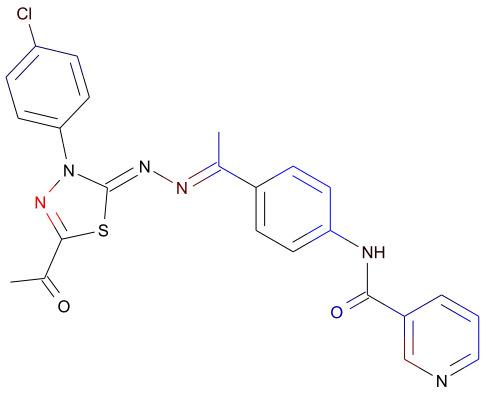
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

Top Features for negative contribution

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

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



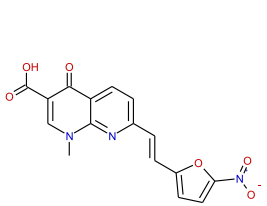
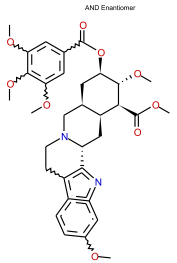
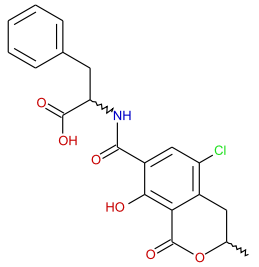
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96465
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 8.97
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14
Mahalanobis Distance p-value: 3.32e-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	470	223	Ochratoxin A
Structure			
Actual Endpoint (-log C)	4.62839	5.08368	4.79932
Predicted Endpoint (-log C)	3.93264	5.08273	3.6353
Distance	0.804	0.829	0.832
Reference	CPDB	CPDB	CPDB

Model Applicability

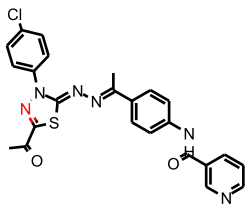
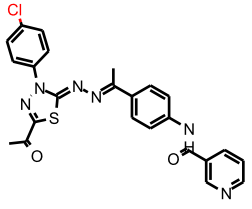
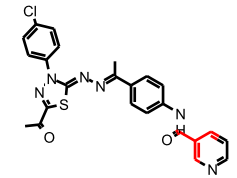
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC20 out of range. Value: 3.6093. Training min, max, SD, explained variance: -4.3384, 3.4394, 1.14, 0.0162.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
7. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1

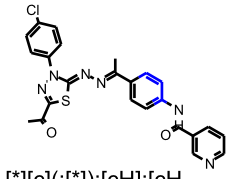
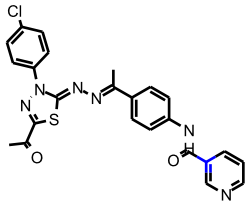
Feature Contribution

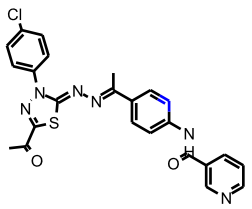
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

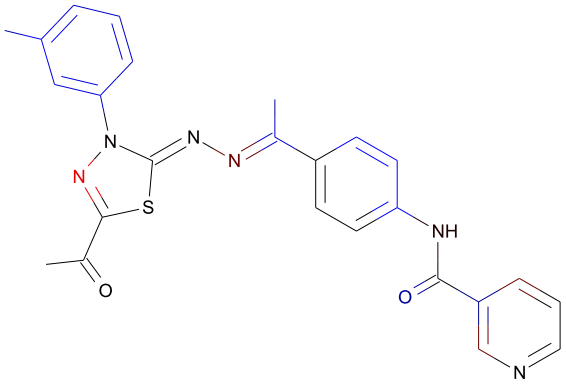
Top Features for negative contribution

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

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

7c

TOPKAT_Carcinogenic_Potency_TD50_Mouse



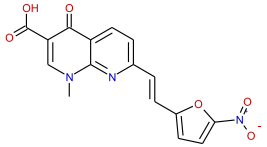
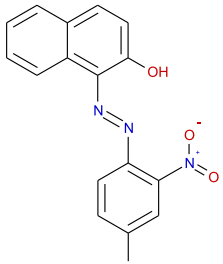
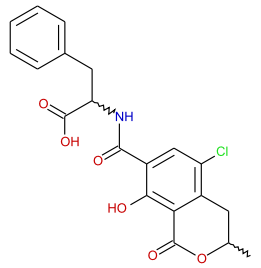
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 27.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.6
Mahalanobis Distance p-value: 4.29e-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	470	646	Ochratoxin A
Structure			
Actual Endpoint (-log C)	4.62839	0.937339	4.79932
Predicted Endpoint (-log C)	3.93264	3.26294	3.6353
Distance	0.788	0.817	0.828
Reference	CPDB	CPDB	CPDB

Model Applicability

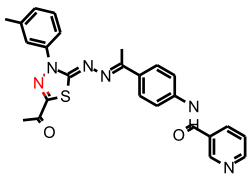
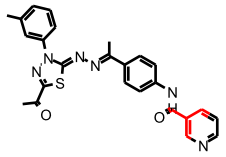
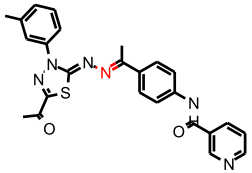
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

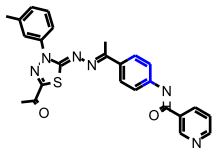
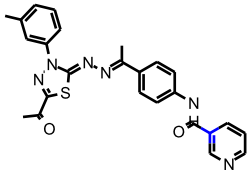
Feature Contribution

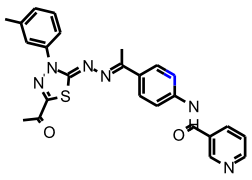
Top features for positive contribution

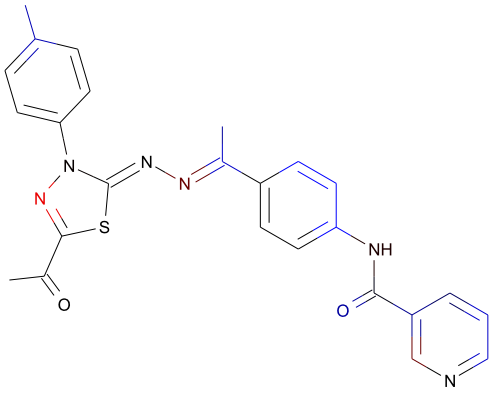
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

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

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



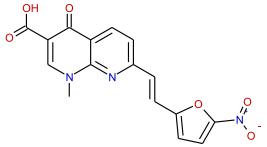
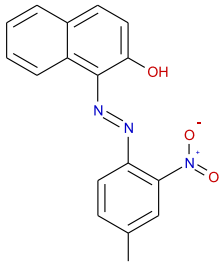
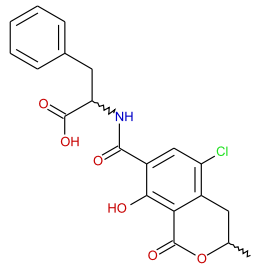
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 16.8
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.1
Mahalanobis Distance p-value: 6.91e-008

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

Structural Similar Compounds

Name	470	646	Ochratoxin A
Structure			
Actual Endpoint (-log C)	4.62839	0.937339	4.79932
Predicted Endpoint (-log C)	3.93264	3.26294	3.6353
Distance	0.787	0.820	0.831
Reference	CPDB	CPDB	CPDB

Model Applicability

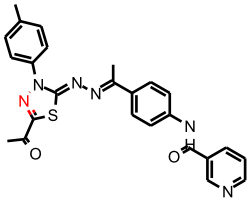
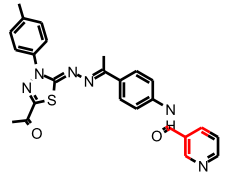
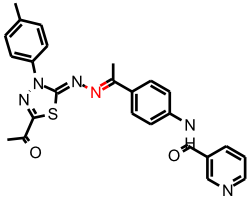
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

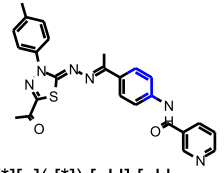
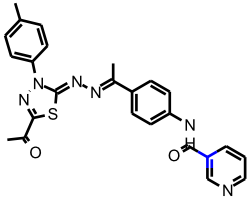
Feature Contribution

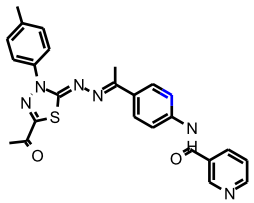
Top features for positive contribution

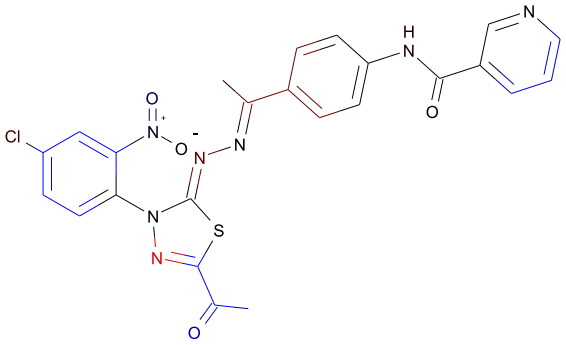
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

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

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



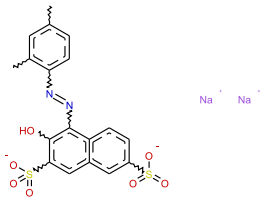
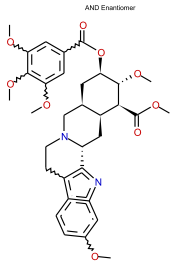
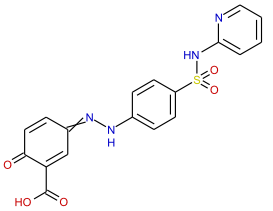
$C_{24}H_{18}ClN_7O_4S$
Molecular Weight: 535.96222
ALogP: 4.153
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 15.8
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15.4
Mahalanobis Distance p-value: 2.64e-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	420	223	Salicylazosulfapyridine
Structure			
Actual Endpoint (-log C)	2.78302	5.08368	2.5034
Predicted Endpoint (-log C)	3.31546	5.08273	3.54214
Distance	0.856	0.857	0.870
Reference	CPDB	CPDB	CPDB

Model Applicability

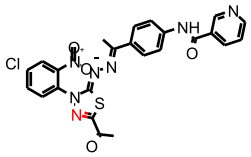
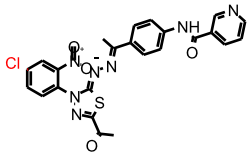
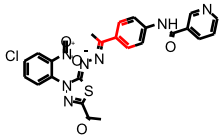
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
3. Unknown ECFP_2 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
4. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
5. Unknown ECFP_2 feature: -1236953626: [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]
6. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
7. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
8. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]

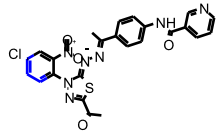
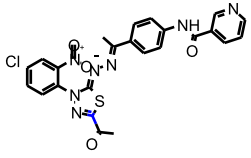
Feature Contribution

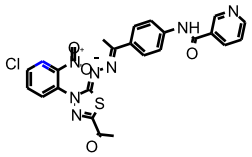
Top features for positive contribution

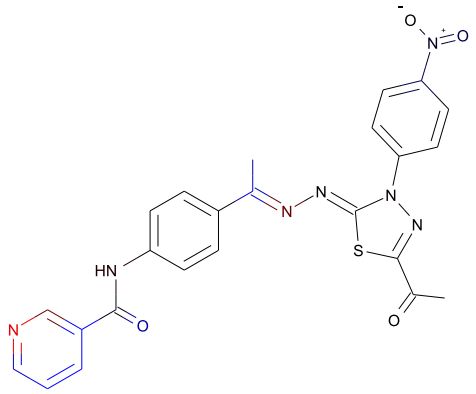
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107

Top Features for negative contribution

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

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



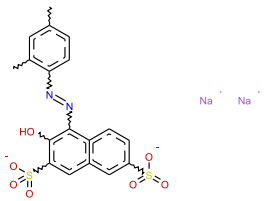
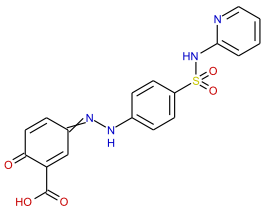
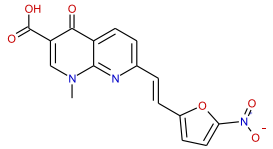
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 13.9
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.8
Mahalanobis Distance p-value: 1.3e-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	420	Salicylazosulfapyridine	470
Structure			
Actual Endpoint (-log C)	2.78302	2.5034	4.62839
Predicted Endpoint (-log C)	3.31546	3.54214	3.93264
Distance	0.813	0.838	0.848
Reference	CPDB	CPDB	CPDB

Model Applicability

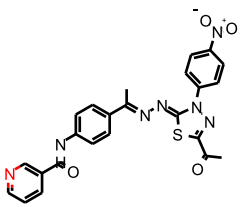
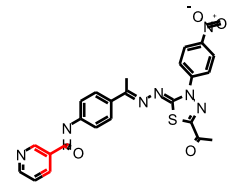
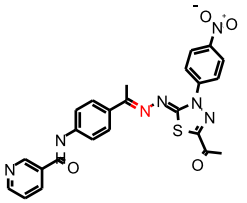
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

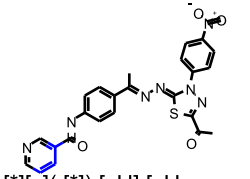
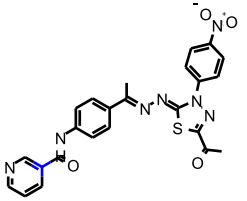
Feature Contribution

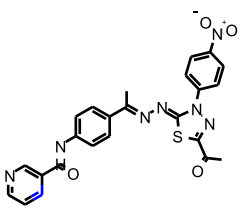
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

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

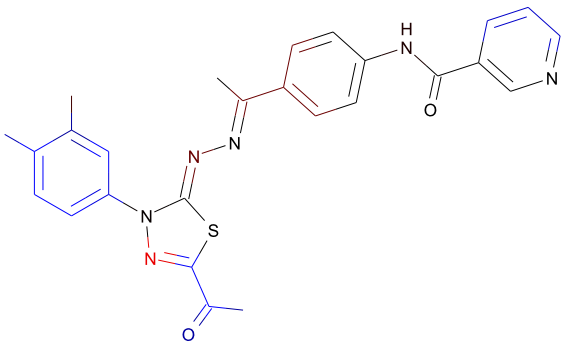
Top Features for negative contribution

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

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

7g

TOPKAT_Carcinogenic_Potency_TD50_Mouse



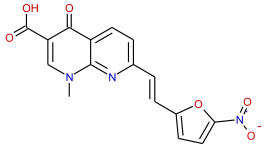
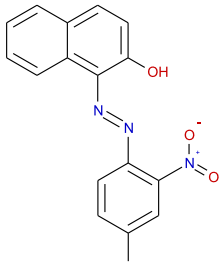
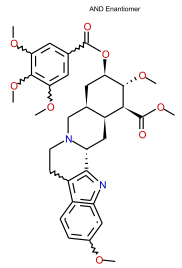
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 20.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.6
Mahalanobis Distance p-value: 3.88e-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	470	646	223
Structure			
Actual Endpoint (-log C)	4.62839	0.937339	5.08368
Predicted Endpoint (-log C)	3.93264	3.26294	5.08273
Distance	0.810	0.828	0.830
Reference	CPDB	CPDB	CPDB

Model Applicability

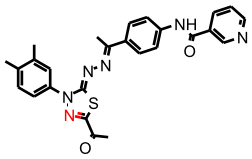
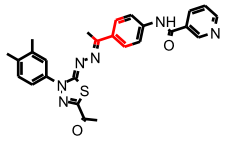
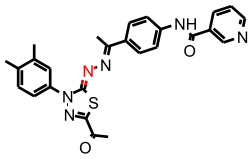
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

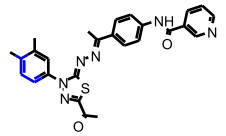
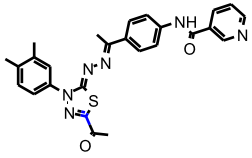
Feature Contribution

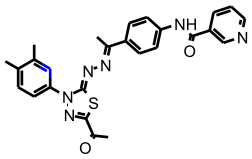
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

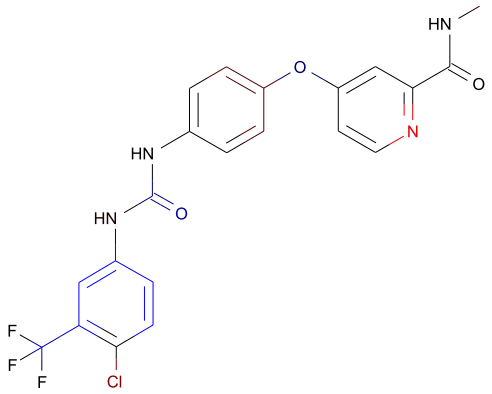
ECFP_6	655739385	 [*]N=[*]	0.229
ECFP_6	-175146122	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	0.107
ECFP_6	-1087070950	 [*]N=[*]	0.104

Top Features for negative contribution

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

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

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

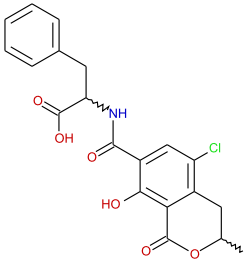
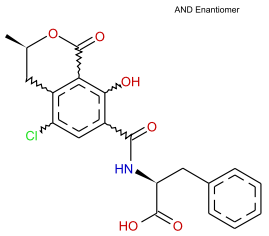
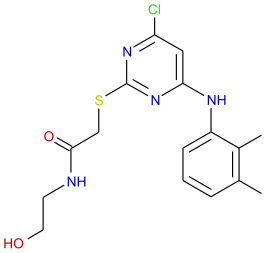
Mahalanobis Distance: 12.4

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

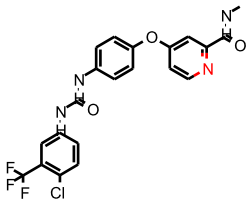
Model Applicability

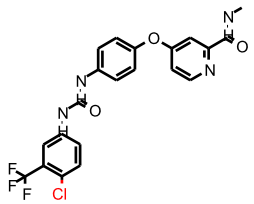
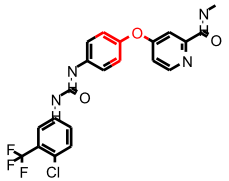
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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:[*]

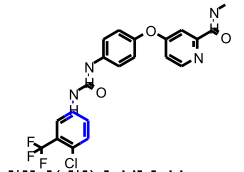
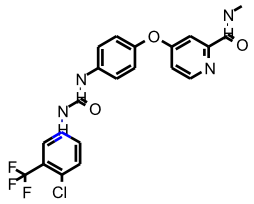
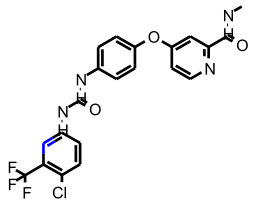
Feature Contribution

Top features for positive contribution

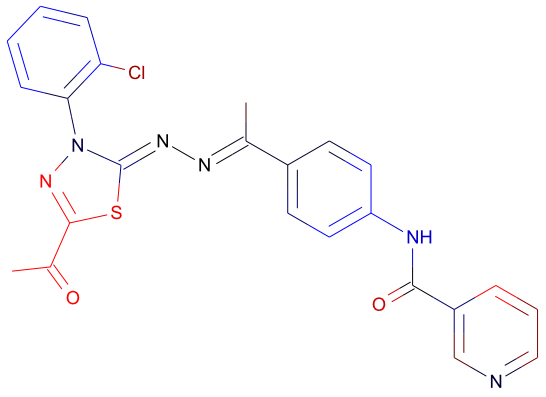
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]N=[*]	0.229

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



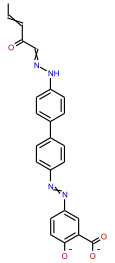
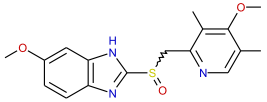
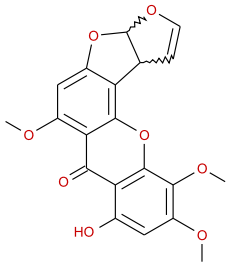
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96466
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 3.64
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 16.9
Mahalanobis Distance p-value: 2.28e-017

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	Omeprazole	5,6-Dimethoxysterigmatocystin
Structure			
Actual Endpoint (-log C)	5.31387	3.4628	6.02361
Predicted Endpoint (-log C)	4.30266	4.7324	4.98771
Distance	0.609	0.764	0.764
Reference	CPDB	CPDB	CPDB

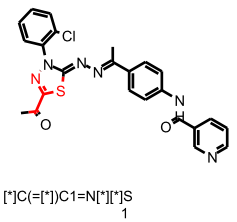
Model Applicability

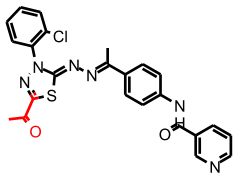
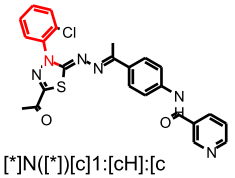
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

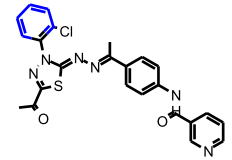
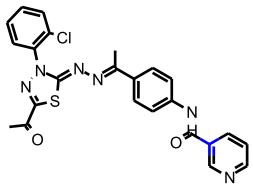
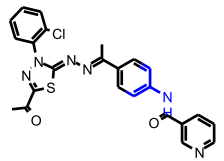
Feature Contribution

Top features for positive contribution

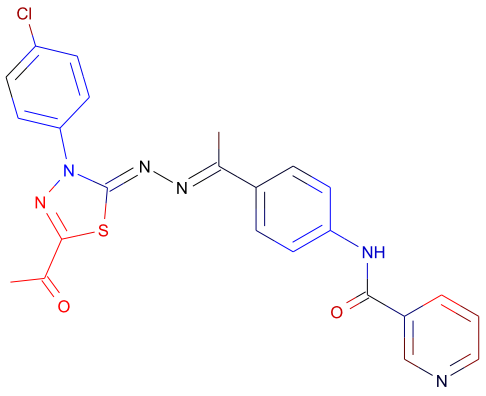
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	-2090462286	 [*]N([*])[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1[*]	0.245

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	 [*][c](:[*]):[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.323

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: 2.42

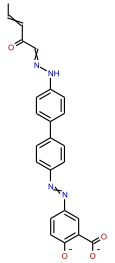
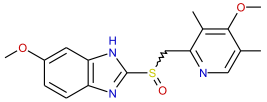
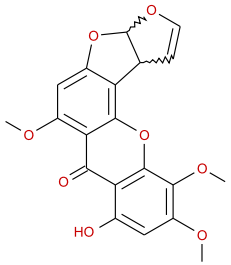
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 9.02e-013

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

Structural Similar Compounds

Name	C.I. direct brown 95	Omeprazole	5,6-Dimethoxysterigmatocystin
Structure			
Actual Endpoint (-log C)	5.31387	3.4628	6.02361
Predicted Endpoint (-log C)	4.30266	4.7324	4.98771
Distance	0.608	0.763	0.764
Reference	CPDB	CPDB	CPDB

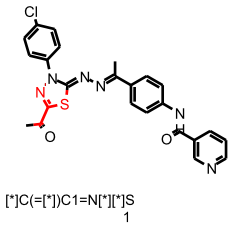
Model Applicability

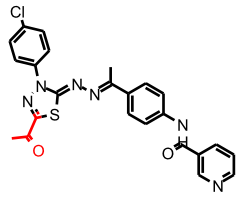
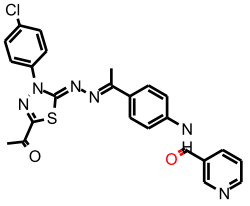
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

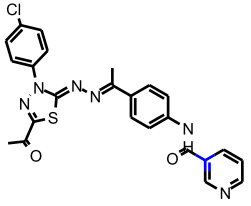
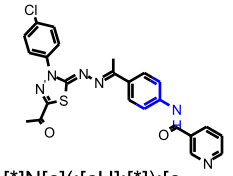
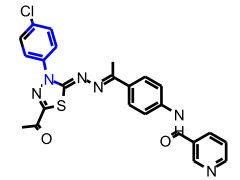
1. OPS PC24 out of range. Value: -3.1965. Training min, max, SD, explained variance: -3.0088, 7.4204, 1.076, 0.0111.

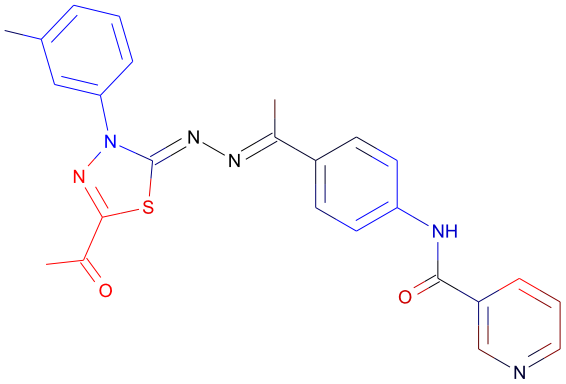
Feature Contribution

Top features for positive contribution

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

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

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



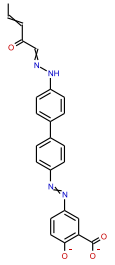
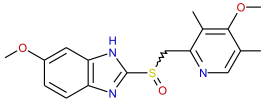
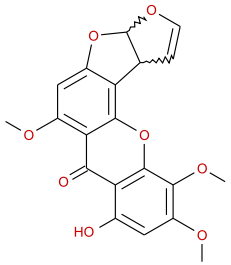
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 11.5
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 16.3
Mahalanobis Distance p-value: 2.97e-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	Omeprazole	5,6-Dimethoxysterigmatocystin
Structure			
Actual Endpoint (-log C)	5.31387	3.4628	6.02361
Predicted Endpoint (-log C)	4.30266	4.7324	4.98771
Distance	0.595	0.746	0.752
Reference	CPDB	CPDB	CPDB

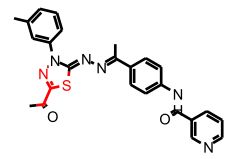
Model Applicability

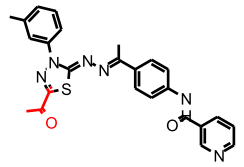
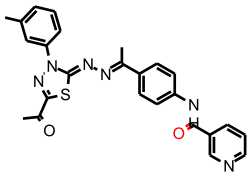
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC24 out of range. Value: -3.7507. Training min, max, SD, explained variance: -3.0088, 7.4204, 1.076, 0.0111.

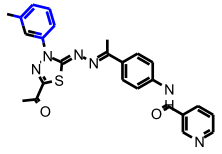
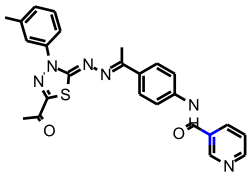
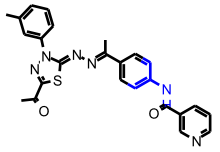
Feature Contribution

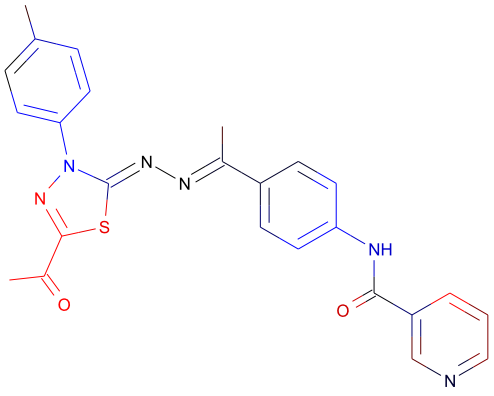
Top features for positive contribution

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

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

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



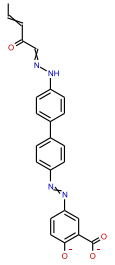
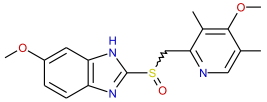
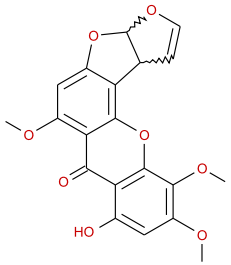
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 3.12
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 15.6
Mahalanobis Distance p-value: 4.93e-013

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

Structural Similar Compounds

Name	C.I. direct brown 95	Omeprazole	5,6-Dimethoxysterigmatocystin
Structure			
Actual Endpoint (-log C)	5.31387	3.4628	6.02361
Predicted Endpoint (-log C)	4.30266	4.7324	4.98771
Distance	0.598	0.744	0.751
Reference	CPDB	CPDB	CPDB

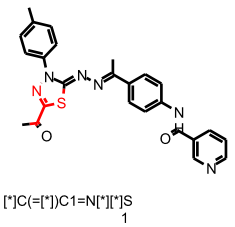
Model Applicability

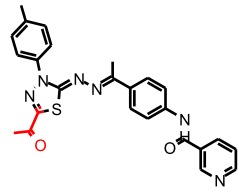
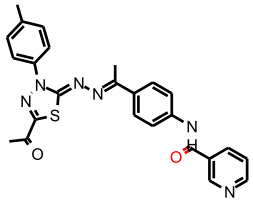
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC24 out of range. Value: -3.0664. Training min, max, SD, explained variance: -3.0088, 7.4204, 1.076, 0.0111.

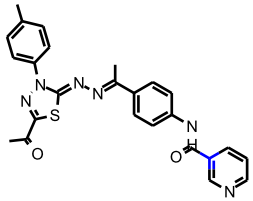
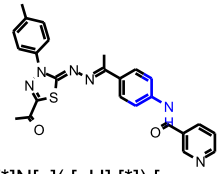
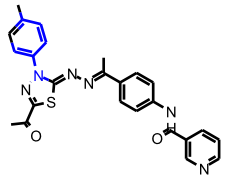
Feature Contribution

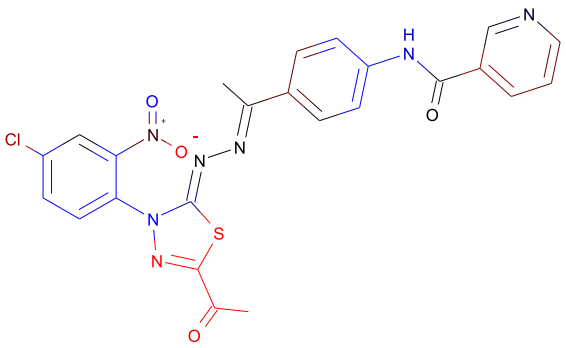
Top features for positive contribution

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

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

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



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: 1.2

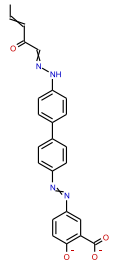
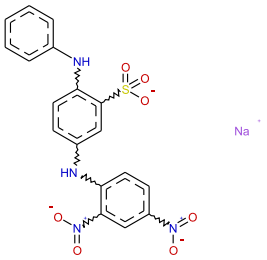
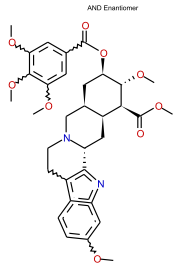
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 17.7

Mahalanobis Distance p-value: 4.87e-020

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	623	223
Structure			
Actual Endpoint (-log C)	5.31387	2.39985	6.29867
Predicted Endpoint (-log C)	4.30266	3.4177	7.5657
Distance	0.718	0.729	0.811
Reference	CPDB	CPDB	CPDB

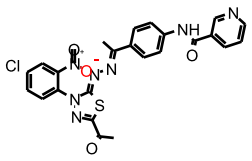
Model Applicability

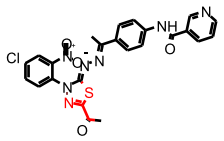
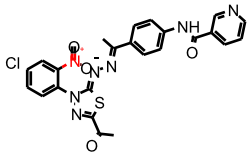
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

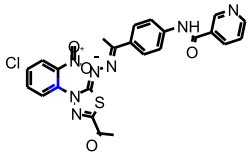
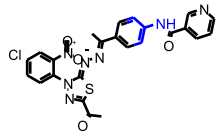
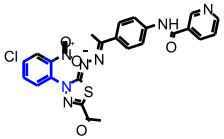
1. OPS PC17 out of range. Value: -4.2325. Training min, max, SD, explained variance: -4.122, 4.7538, 1.355, 0.0176.
2. OPS PC24 out of range. Value: -4.0628. Training min, max, SD, explained variance: -3.0088, 7.4204, 1.076, 0.0111.

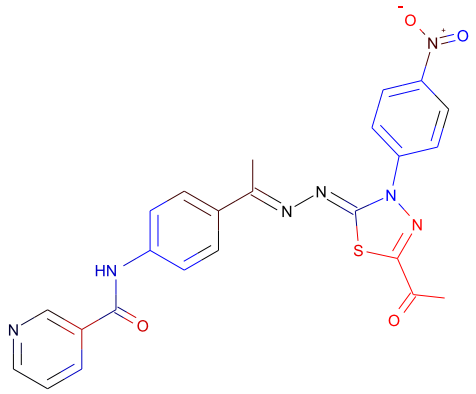
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	5	 [*][O-]	0.431

FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357
FCFP_6	8	 [*][N+](=[*])[*]	0.336

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



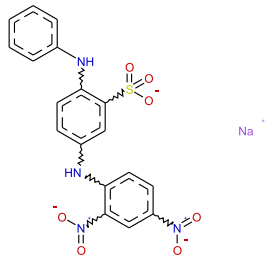
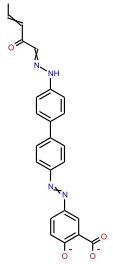
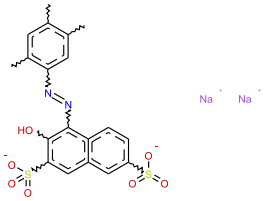
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 1.91
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 16.7
Mahalanobis Distance p-value: 1.49e-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	623	C.I. direct brown 95	418
Structure			
Actual Endpoint (-log C)	2.39985	5.31387	2.9349
Predicted Endpoint (-log C)	3.4177	4.30266	3.45907
Distance	0.676	0.698	0.768
Reference	CPDB	CPDB	CPDB

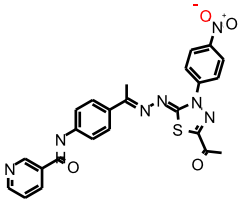
Model Applicability

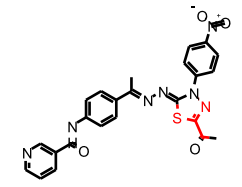
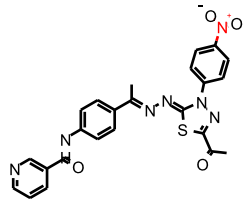
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC24 out of range. Value: -3.64. Training min, max, SD, explained variance: -3.0088, 7.4204, 1.076, 0.0111.

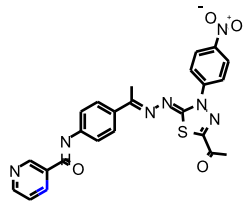
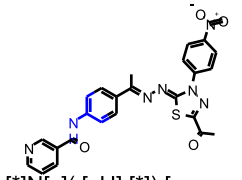
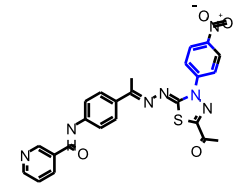
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	5	 [*][O-]	0.431

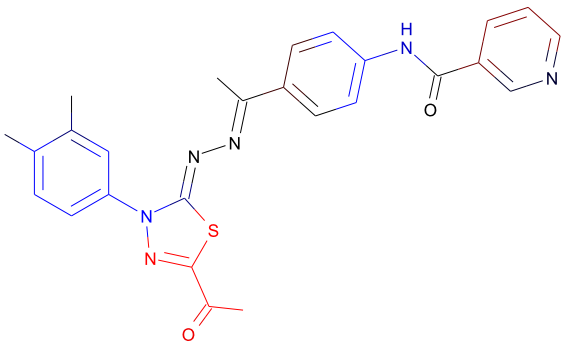
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357
FCFP_6	8	 [*][N+](=[*])[*]	0.336

Top Features for negative contribution

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

7g

TOPKAT_Carcinogenic_Potency_TD50_Rat



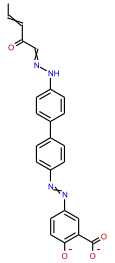
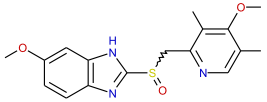
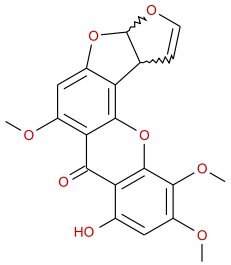
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 4.03
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 16.3
Mahalanobis Distance p-value: 3.36e-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	Omeprazole	5,6-Dimethoxysterigmatocystin
Structure			
Actual Endpoint (-log C)	5.31387	3.4628	6.02361
Predicted Endpoint (-log C)	4.30266	4.7324	4.98771
Distance	0.609	0.769	0.775
Reference	CPDB	CPDB	CPDB

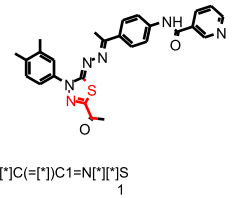
Model Applicability

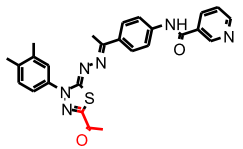
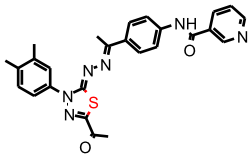
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC24 out of range. Value: -3.787. Training min, max, SD, explained variance: -3.0088, 7.4204, 1.076, 0.0111.

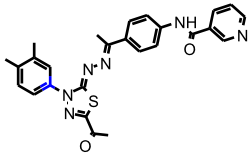
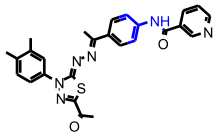
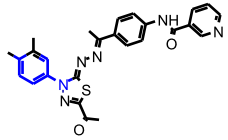
Feature Contribution

Top features for positive contribution

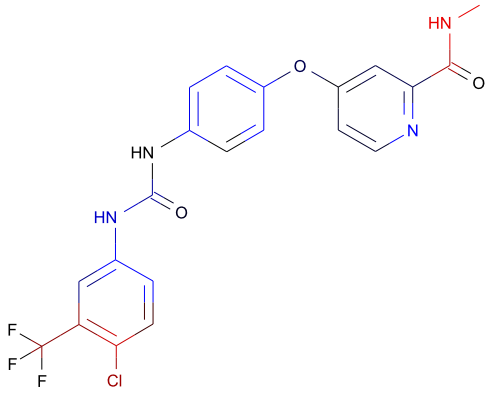
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	565998553	 [*]C(=[*])C1=N[*][*]S 1	0.357

FCFP_6	565968762	 [*]C(=[*])C(=O)C	0.266
FCFP_6	1	 [*]=O	0.234

Top Features for negative contribution

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

Sorafenib



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 14.2

Unit: mg/kg_body_weight/day

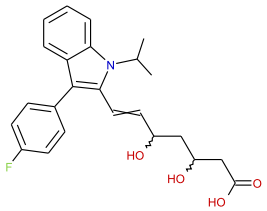
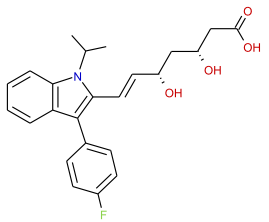
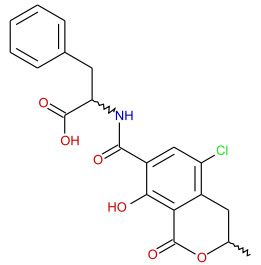
Mahalanobis Distance: 20.4

Mahalanobis Distance p-value: 9.56e-031

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

TOPKAT_Carcinogenic_Potency_TD50_Rat

Structural Similar Compounds

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

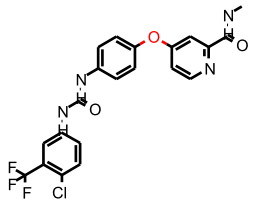
Model Applicability

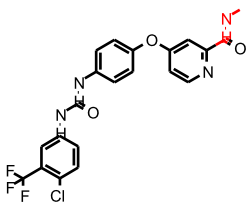
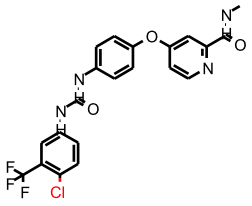
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

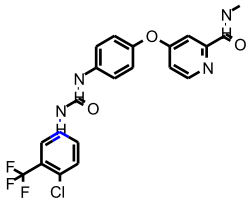
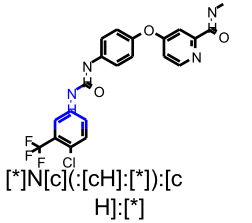
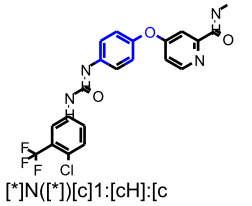
Feature Contribution

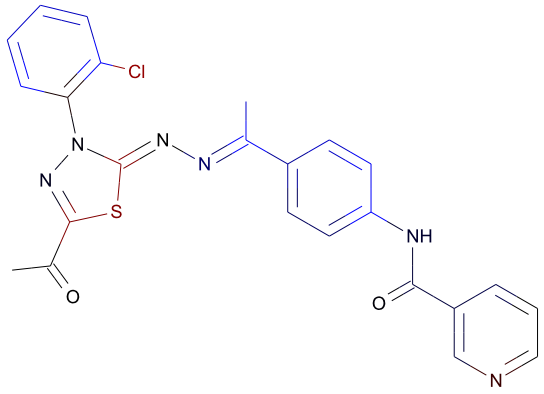
Top features for positive contribution

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

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

Top Features for negative contribution

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



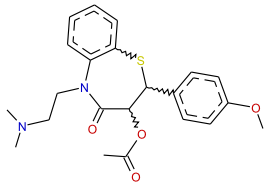
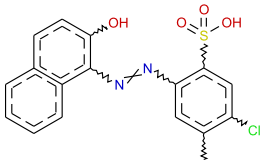
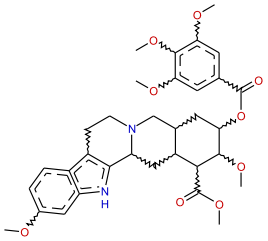
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96466
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0619
Unit: g/kg_body_weight
Mahalanobis Distance: 34.2
Mahalanobis Distance p-value: 3.23e-032

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

Structural Similar Compounds

Name	DILTIAZEM	D & C RED 9	RESERPINE
Structure			
Actual Endpoint (-log C)	4.21961	3.87715	6.38645
Predicted Endpoint (-log C)	4.005	3.6546	5.548
Distance	0.738	0.752	0.768
Reference	NDA-18602	NTP REPORT # 225	NTP 193 22

Model Applicability

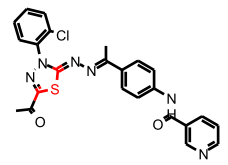
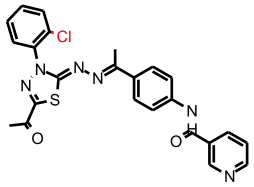
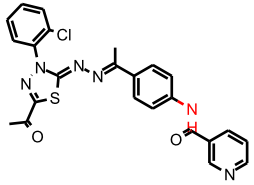
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC32 out of range. Value: 4.2981. Training min, max, SD, explained variance: -4.2021, 4.2975, 1.228, 0.0066.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
4. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
6. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
8. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
12. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
13. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
14. Unknown ECFP_6 feature: -1236953626: [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]
15. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
18. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]
19. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]

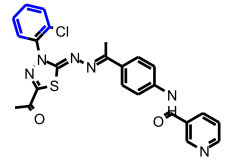
20. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl

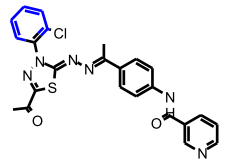
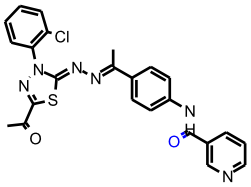
Feature Contribution

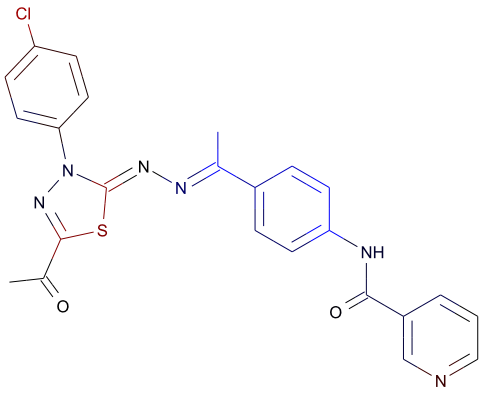
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

Top Features for negative contribution

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

ECFP_6	1564392544	 [*][c]1:[*]:[cH]:[cH] :[cH]:[cH]:1	-0.133
FCFP_6	1	 [*]=O	-0.102



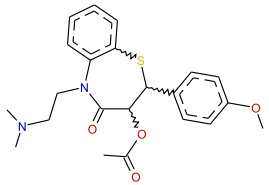
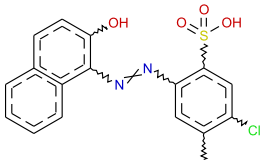
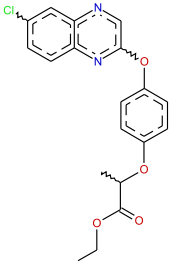
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96465
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0276
Unit: g/kg_body_weight
Mahalanobis Distance: 31.9
Mahalanobis Distance p-value: 4.44e-028

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

Structural Similar Compounds

Name	DILTIAZEM	D & C RED 9	ASSURE
Structure			
Actual Endpoint (-log C)	4.21961	3.87715	5.00328
Predicted Endpoint (-log C)	4.005	3.6546	4.27671
Distance	0.748	0.758	0.762
Reference	NDA-18602	NTP REPORT # 225	EPA COVER SHEET 0335;891001;(1)

Model Applicability

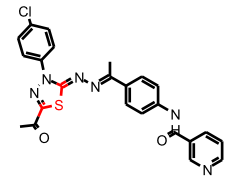
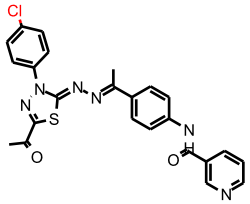
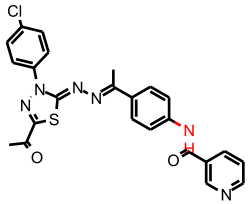
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
4. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
6. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
8. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
12. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
13. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
14. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
16. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
17. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
18. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
19. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]

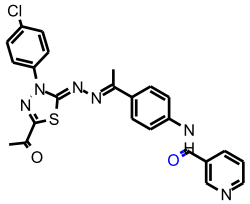
20. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]
21. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

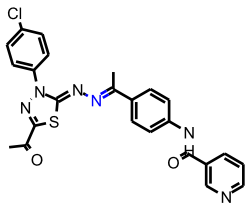
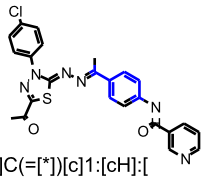
Feature Contribution

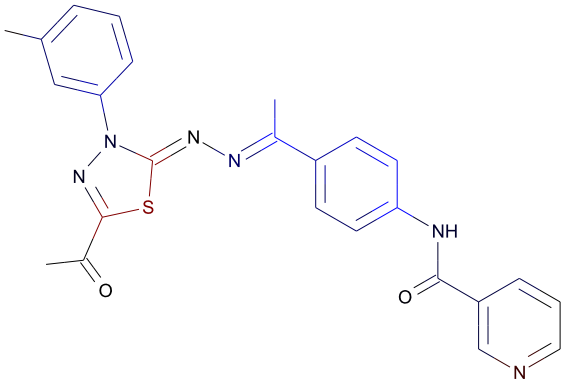
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

Top Features for negative contribution

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

ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.0906



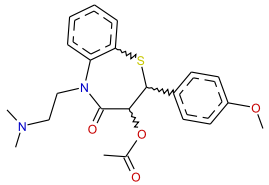
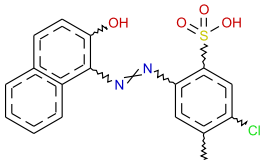
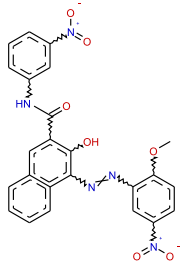
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0494
Unit: g/kg_body_weight
Mahalanobis Distance: 32.6
Mahalanobis Distance p-value: 2.34e-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	DILTIAZEM	D & C RED 9	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	4.21961	3.87715	2.28997
Predicted Endpoint (-log C)	4.005	3.6546	3.52921
Distance	0.732	0.760	0.771
Reference	NDA-18602	NTP REPORT # 225	NTP 411 146

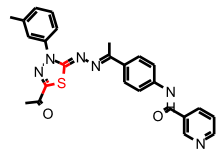
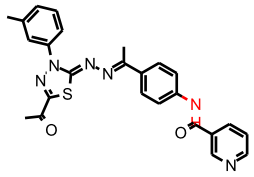
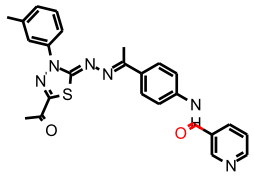
Model Applicability

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

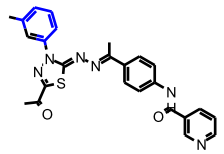
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
4. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
6. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
8. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
12. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
13. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
14. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
16. Unknown ECFP_6 feature: -179515162: [*]:[cH]:[c](C):[cH]:[*]
17. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
18. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
19. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]
20. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]

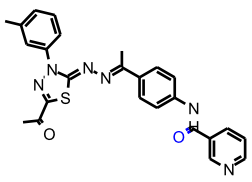
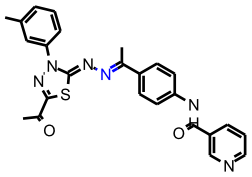
Feature Contribution

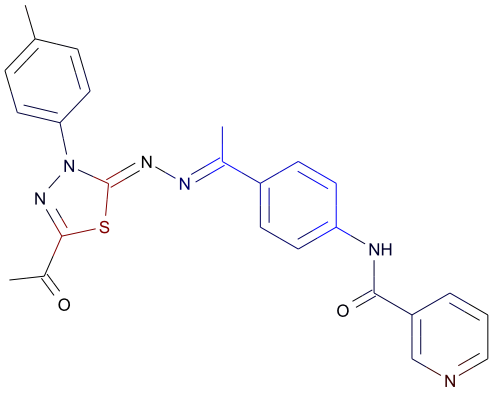
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	3	 [*]N[*]	0.0924
ECFP_6	2099970318	 [*]C(=O)[*]	0.0766

Top Features for negative contribution

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

FCFP_6	1	 [*]=O	-0.102
ECFP_6	-1087070950	 [*]N=[*]	-0.102



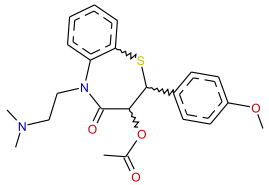
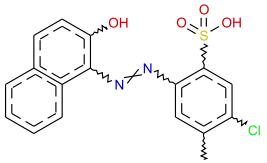
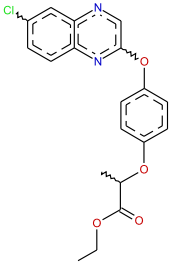
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0383
Unit: g/kg_body_weight
Mahalanobis Distance: 31.4
Mahalanobis Distance p-value: 3.26e-027

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

Structural Similar Compounds

Name	DILTIAZEM	D & C RED 9	ASSURE
Structure			
Actual Endpoint (-log C)	4.21961	3.87715	5.00328
Predicted Endpoint (-log C)	4.005	3.6546	4.27671
Distance	0.733	0.764	0.776
Reference	NDA-18602	NTP REPORT # 225	EPA COVER SHEET 0335;891001;(1)

Model Applicability

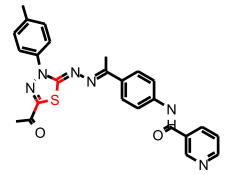
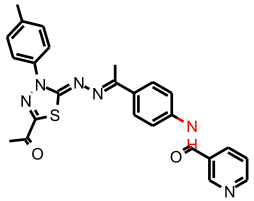
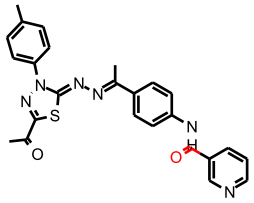
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
4. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
5. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
6. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
8. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
9. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
10. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
11. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
12. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
13. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
14. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: -179515162: [*]:[cH]:[c](C):[cH]:[*]
16. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
17. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
18. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]
19. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]

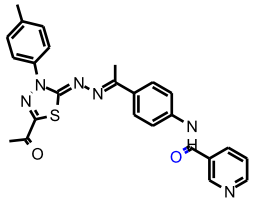
20. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

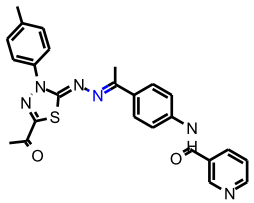
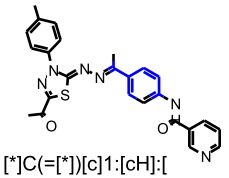
Feature Contribution

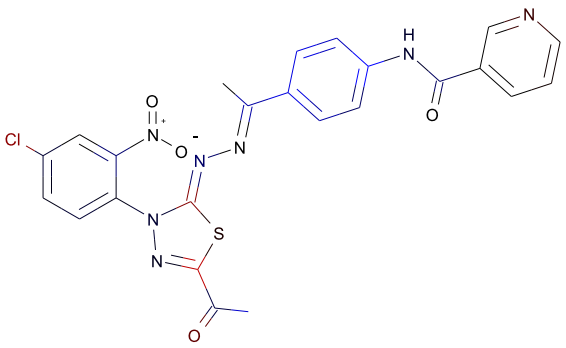
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	3	 [*]N[*]	0.0924
ECFP_6	2099970318	 [*]C(=O)[*]	0.0766

Top Features for negative contribution

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

ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.0906



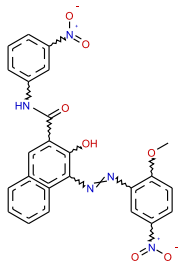
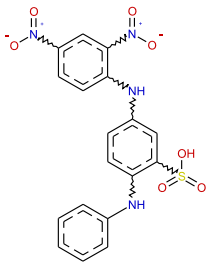
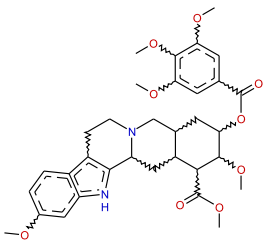
$C_{24}H_{18}ClN_7O_4S$
Molecular Weight: 535.96222
ALogP: 4.153
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 0.0304
Unit: g/kg_body_weight
Mahalanobis Distance: 36.6
Mahalanobis Distance p-value: 2.15e-036

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

Structural Similar Compounds

Name	C.I.PIGMENT RED 23	C.I. ACID ORANGE 3	RESERPINE
Structure			
Actual Endpoint (-log C)	2.28997	3.20573	6.38645
Predicted Endpoint (-log C)	3.52921	3.55956	5.548
Distance	0.641	0.765	0.775
Reference	NTP 411 146	NTP REPORT # 335	NTP 193 22

Model Applicability

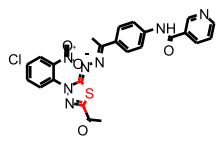
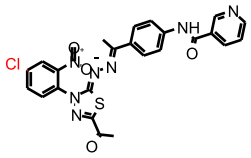
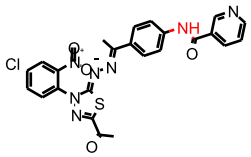
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 5: [*][O-]
3. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
4. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
5. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
6. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]
7. Unknown ECFP_6 feature: 912478223: [*]S[*]
8. Unknown ECFP_6 feature: 1043790491: [*][N+](=[*])[*]
9. Unknown ECFP_6 feature: 781519895: [*][O-]
10. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
11. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
12. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
13. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
14. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
15. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
16. Unknown ECFP_6 feature: -1236953626: [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]
17. Unknown ECFP_6 feature: -1956535100: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
18. Unknown ECFP_6 feature: -176494269: [*]:[cH]:[c](Cl):[cH]:[*]
19. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
20. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]

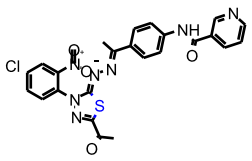
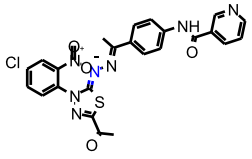
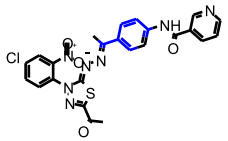
21. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
22. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
23. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
24. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
25. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
26. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
27. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]
28. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]
29. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
30. Unknown ECFP_6 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
31. Unknown ECFP_6 feature: 2104376220: [*][N+](=O)[*]
32. Unknown ECFP_6 feature: -659271057: [*][N+](=[*])[O-]

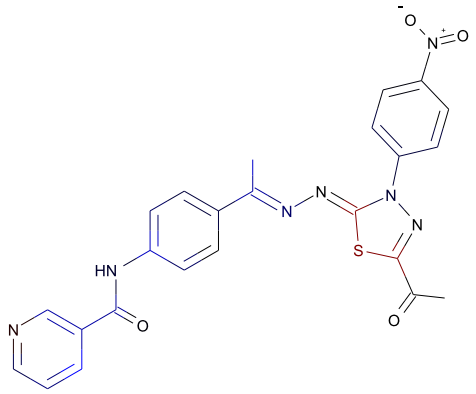
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	32	 [*]Cl	0.101
FCFP_6	3	 [*]N[*]	0.0924

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	-0.102
ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.0906



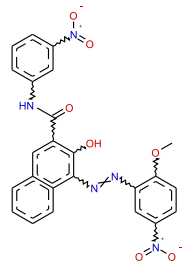
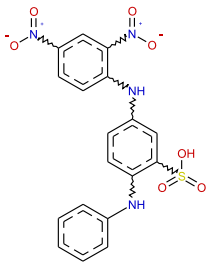
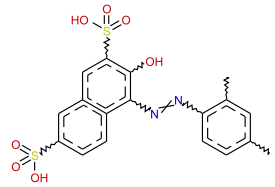
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 0.0345
Unit: g/kg_body_weight
Mahalanobis Distance: 34.4
Mahalanobis Distance p-value: 1.36e-032

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

Structural Similar Compounds

Name	C.I.PIGMENT RED 23	C.I. ACID ORANGE 3	3-HYDROXY-4-(2;4-XYLYLAZO)-2;7
Structure			
Actual Endpoint (-log C)	2.28997	3.20573	3.63994
Predicted Endpoint (-log C)	3.52921	3.55956	3.60721
Distance	0.656	0.743	0.784
Reference	NTP 411 146	NTP REPORT # 335	RQTOX ECAO CIN R638;9807;(4)

Model Applicability

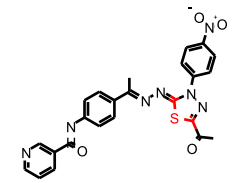
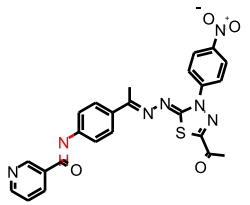
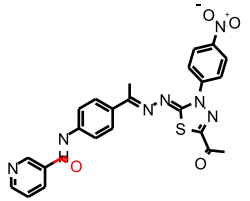
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 5: [*][O-]
3. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
4. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
5. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
6. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]
7. Unknown ECFP_6 feature: 912478223: [*]S[*]
8. Unknown ECFP_6 feature: 1043790491: [*][N+](=[*])[*]
9. Unknown ECFP_6 feature: 781519895: [*][O-]
10. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]
11. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]
12. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]
13. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
14. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
16. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
17. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
18. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
19. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]

20. Unknown ECFP_6 feature: -179073144: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
21. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
22. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
23. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
24. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
25. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
26. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
27. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
28. Unknown ECFP_6 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
29. Unknown ECFP_6 feature: 2104376220: [*][N+](=O)[*]
30. Unknown ECFP_6 feature: -659271057: [*][N+](=[*])[O-]

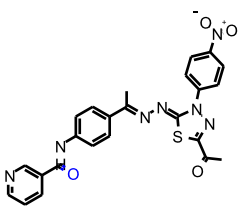
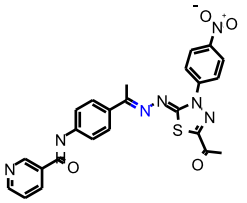
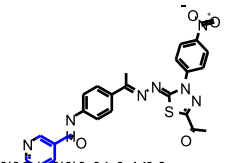
Feature Contribution

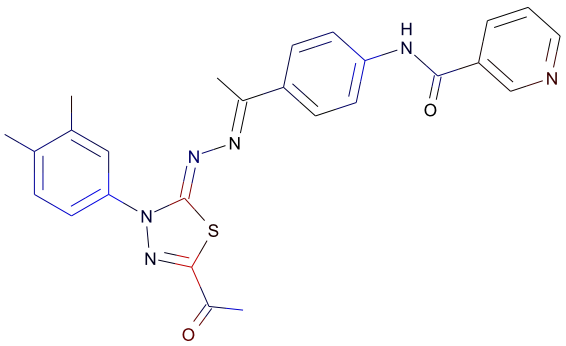
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
FCFP_6	3	 [*]N[*]	0.0924
ECFP_6	2099970318	 [*]C(=O)[*]	0.0766

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_6	1	 [*]=O	-0.102
ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.0906



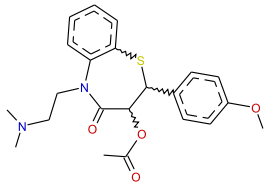
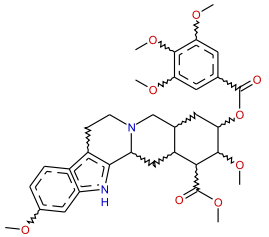
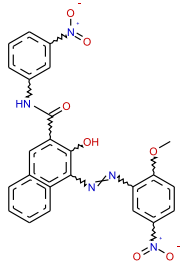
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0368
Unit: g/kg_body_weight
Mahalanobis Distance: 32.2
Mahalanobis Distance p-value: 1.26e-028

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

Structural Similar Compounds

Name	DILTIAZEM	RESERPINE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	4.21961	6.38645	2.28997
Predicted Endpoint (-log C)	4.005	5.548	3.52921
Distance	0.753	0.756	0.760
Reference	NDA-18602	NTP 193 22	NTP 411 146

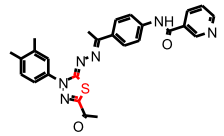
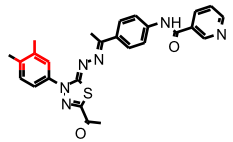
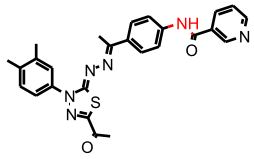
Model Applicability

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

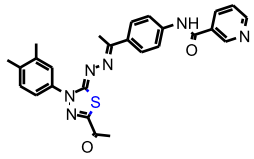
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: 912478223: [*]S[*]
3. Unknown ECFP_6 feature: 2122741631: [*]C1=[*][*]C(=[*])S1
4. Unknown ECFP_6 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
5. Unknown ECFP_6 feature: 2092245922: [*]N1[*][*]C(=N1)[*]
6. Unknown ECFP_6 feature: -1110911409: [*]=C1[*][*]=NN1[c](:[*]):[*]
7. Unknown ECFP_6 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
8. Unknown ECFP_6 feature: 129482634: [*]C(=[*])C(=O)C
9. Unknown ECFP_6 feature: -175021654: [*]N([*])[c](:[cH]:[*]):[cH]:[*]
10. Unknown ECFP_6 feature: -819426257: [*]C(=NN=[*])[*]
11. Unknown ECFP_6 feature: 562081661: [*]C(=NN=[*])[*]
12. Unknown ECFP_6 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
13. Unknown ECFP_6 feature: -175146122: [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]
14. Unknown ECFP_6 feature: -177077903: [*]N[c](:[cH]:[*]):[cH]:[*]
15. Unknown ECFP_6 feature: 1430169877: [*]NC(=O)[c](:[*]):[*]
16. Unknown ECFP_6 feature: -709633021: [*][c](:[*]):[cH]:n:[*]
17. Unknown ECFP_6 feature: -677055651: [*]:[cH]:n:[cH]:[*]
18. Unknown ECFP_6 feature: 1996163143: [*]:[cH]:[cH]:n:[*]
19. Unknown ECFP_6 feature: 1997021792: [*]:[cH]:[cH]:[cH]:[*]

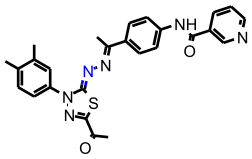
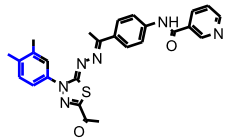
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.13
ECFP_6	2147419938	 [*][c](:[*]):[c](C):[cH]:[*]	0.098
FCFP_6	3	 [*]N[*]	0.0924

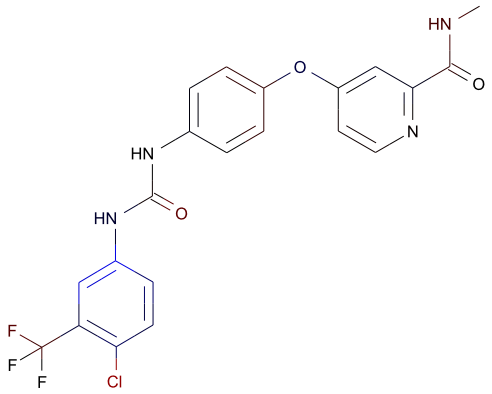
Top Features for negative contribution

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

ECFP_6	-1087070950	 [*]N=[*]	-0.102
FCFP_6	-453677277	 [*]C(=[*])[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.0906

Sorafenib

TOPKAT_Chronic_LOAEL



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.00483

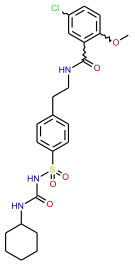
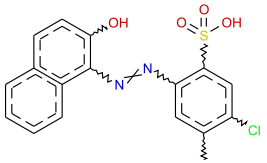
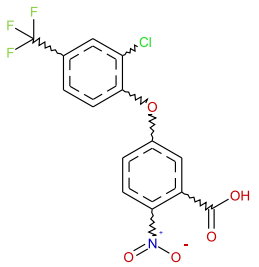
Unit: g/kg_body_weight

Mahalanobis Distance: 30

Mahalanobis Distance p-value: 1.21e-024

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

Structural Similar Compounds

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

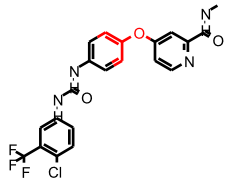
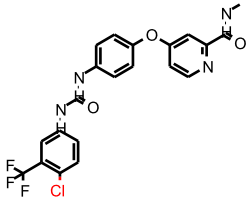
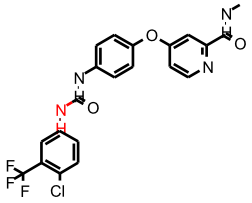
Model Applicability

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

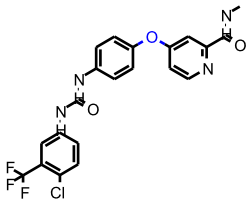
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_6 feature: -1046436026: [*]F
3. Unknown ECFP_6 feature: 99947387: [*]:[c](:[*])Cl
4. Unknown ECFP_6 feature: 226796801: [*]C([*])([*])F
5. Unknown ECFP_6 feature: 1305253718: [*]:[c](:[*])O[c](:[*]):[*]
6. Unknown ECFP_6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

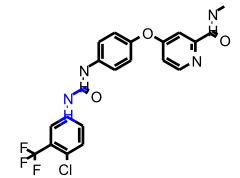
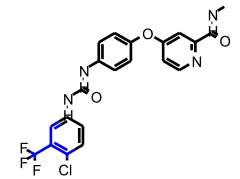
Feature Contribution

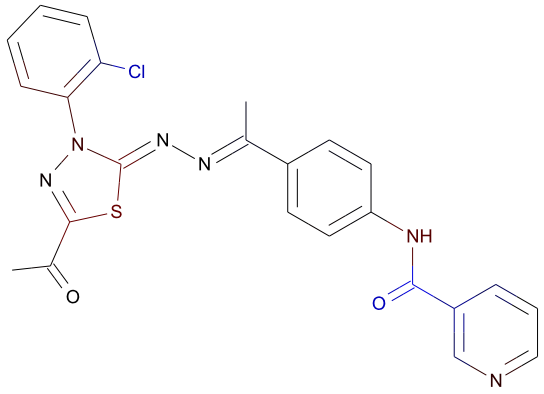
Top features for positive contribution

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

Top Features for negative contribution

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

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



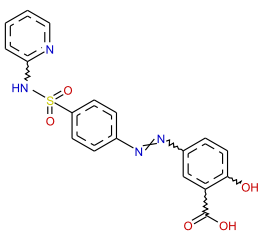
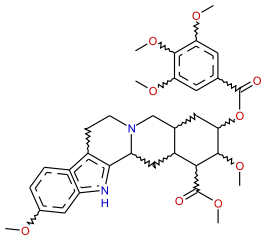
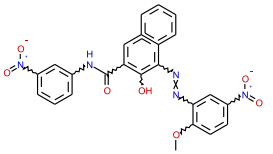
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96466
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0508
Unit: g/kg_body_weight
Mahalanobis Distance: 10.7
Mahalanobis Distance p-value: 3.02e-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	RESERPINE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	3.375	6.13118	2.30052
Predicted Endpoint (-log C)	2.80292	4.38304	3.55333
Distance	0.730	0.734	0.811
Reference	NCI/NTP TR-457	NCI/NTP TR-193	NCI/NTP TR-411

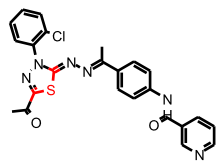
Model Applicability

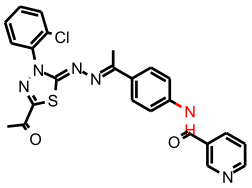
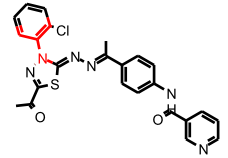
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

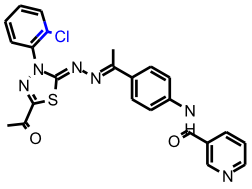
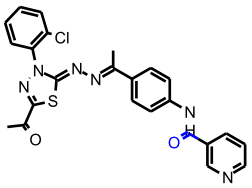
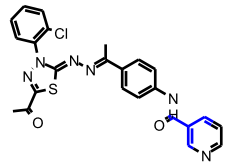
1. All properties and OPS components are within expected ranges.

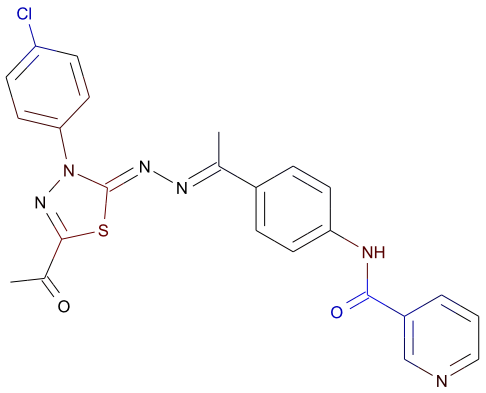
Feature Contribution

Top features for positive contribution

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

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



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: 0.0508

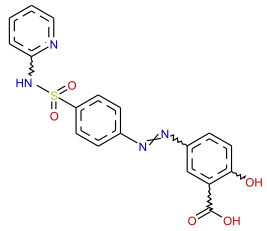
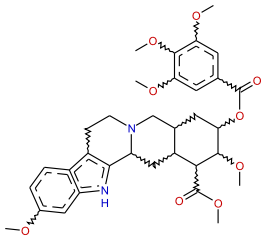
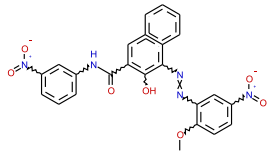
Unit: g/kg_body_weight

Mahalanobis Distance: 10.7

Mahalanobis Distance p-value: 3.02e-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	RESERPINE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	3.375	6.13118	2.30052
Predicted Endpoint (-log C)	2.80292	4.38304	3.55333
Distance	0.730	0.734	0.811
Reference	NCI/NTP TR-457	NCI/NTP TR-193	NCI/NTP TR-411

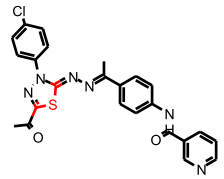
Model Applicability

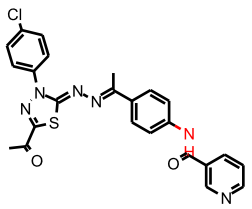
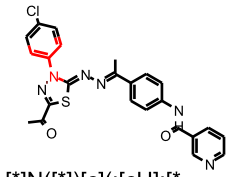
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

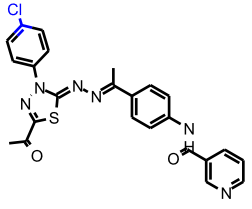
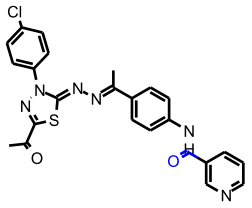
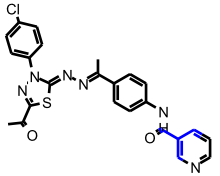
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

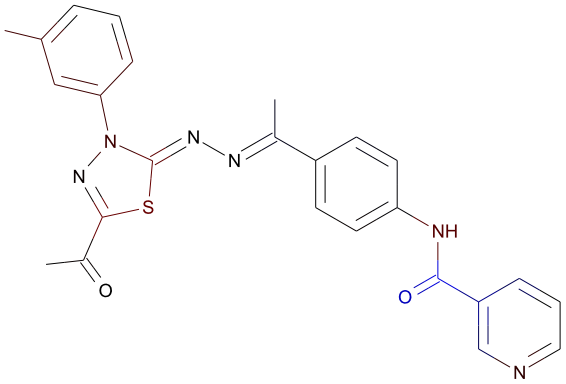
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095

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

7c

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{25}H_{22}N_6O_2S$

Molecular Weight: 470.54617

ALogP: 4.08

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: 0.0344

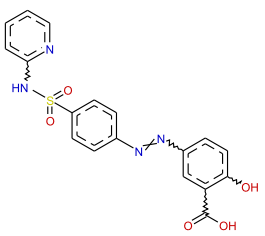
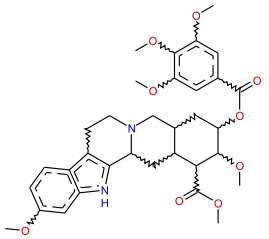
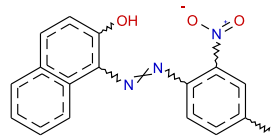
Unit: g/kg_body_weight

Mahalanobis Distance: 10.4

Mahalanobis Distance p-value: 9.12e-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	RESERPINE	C.I.PIGMENT RED 3
Structure			
Actual Endpoint (-log C)	3.375	6.13118	2.65635
Predicted Endpoint (-log C)	2.80292	4.38304	2.97957
Distance	0.704	0.749	0.802
Reference	NCI/NTP TR-457	NCI/NTP TR-193	NCI/NTP TR-407

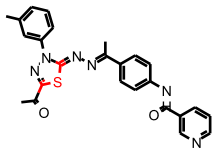
Model Applicability

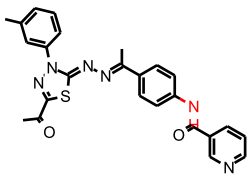
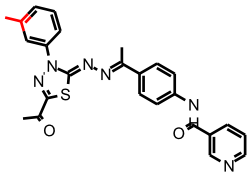
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

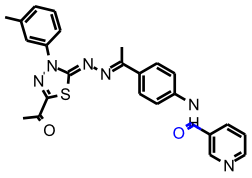
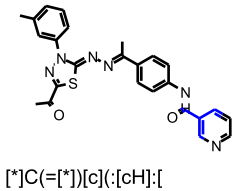
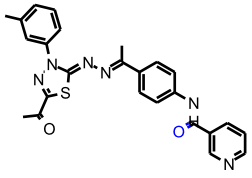
Feature Contribution

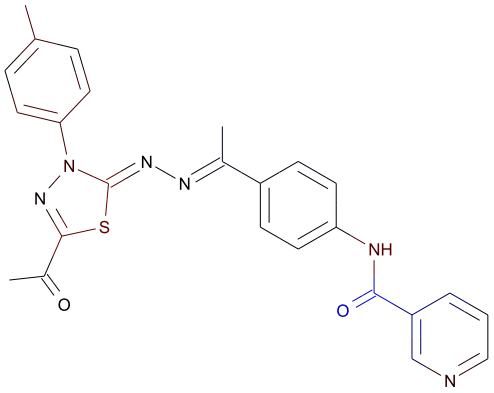
Top features for positive contribution

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

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

Top Features for negative contribution

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



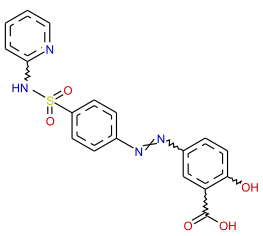
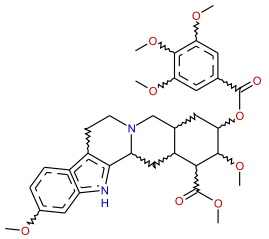
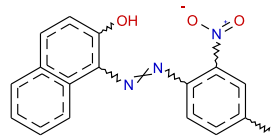
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0344
Unit: g/kg_body_weight
Mahalanobis Distance: 10.4
Mahalanobis Distance p-value: 9.12e-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	RESERPINE	C.I.PIGMENT RED 3
Structure			
Actual Endpoint (-log C)	3.375	6.13118	2.65635
Predicted Endpoint (-log C)	2.80292	4.38304	2.97957
Distance	0.704	0.749	0.802
Reference	NCI/NTP TR-457	NCI/NTP TR-193	NCI/NTP TR-407

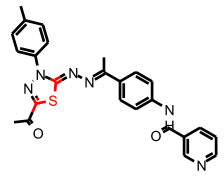
Model Applicability

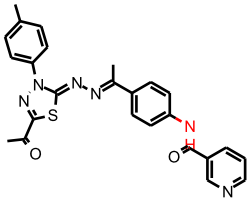
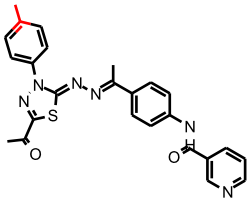
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

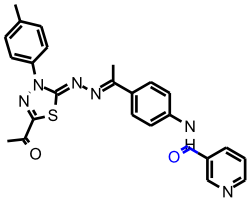
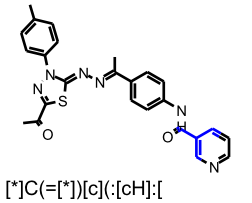
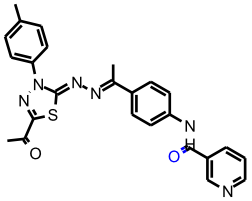
Feature Contribution

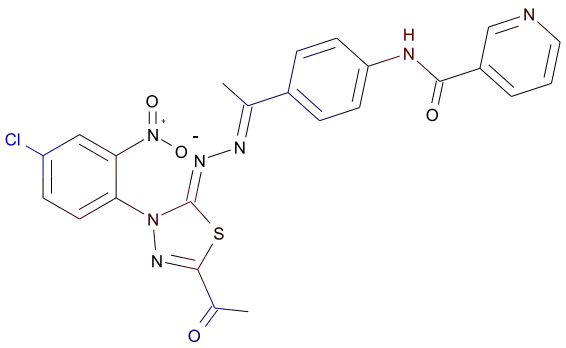
Top features for positive contribution

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

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

Top Features for negative contribution

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



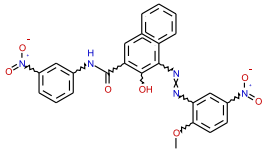
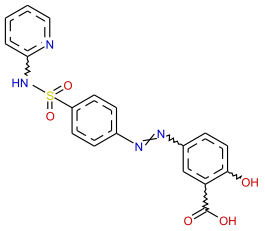
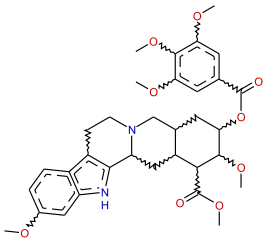
$C_{24}H_{18}ClN_7O_4S$
Molecular Weight: 535.96222
ALogP: 4.153
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 0.0369
Unit: g/kg_body_weight
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 7.44e-008

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

Structural Similar Compounds

Name	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRIDINE	RESERPINE
Structure			
Actual Endpoint (-log C)	2.30052	3.375	6.13118
Predicted Endpoint (-log C)	3.55333	2.80292	4.38304
Distance	0.681	0.769	0.805
Reference	NCI/NTP TR-411	NCI/NTP TR-457	NCI/NTP TR-193

Model Applicability

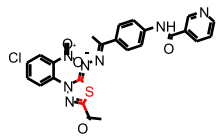
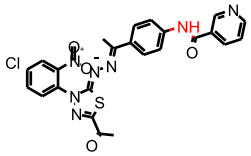
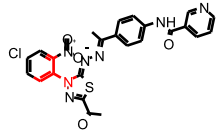
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC5 out of range. Value: 5.5211. Training min, max, SD, explained variance: -3.3892, 5.0834, 1.644, 0.0611.
2. Unknown FCFP_2 feature: 8: [*][N+](=[*])[*]
3. Unknown FCFP_2 feature: 5: [*][O-]
4. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
5. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
6. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
7. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

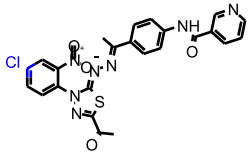
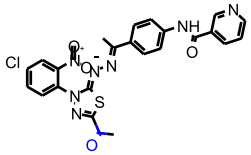
Feature Contribution

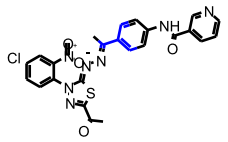
Top features for positive contribution

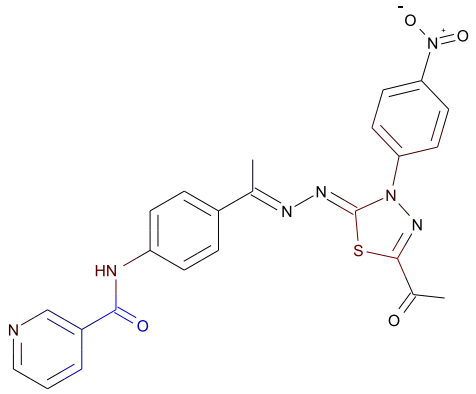
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095
FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]	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



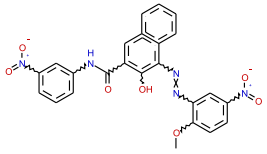
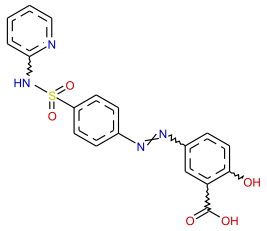
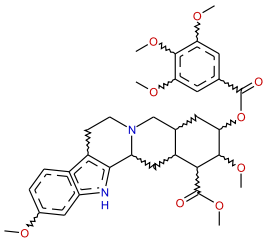
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 0.0307
Unit: g/kg_body_weight
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 5.08e-007

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

Structural Similar Compounds

Name	C.I.PIGMENT RED 23	SALICYLAZOSULFAPYRIDINE	RESERPINE
Structure			
Actual Endpoint (-log C)	2.30052	3.375	6.13118
Predicted Endpoint (-log C)	3.55333	2.80292	4.38304
Distance	0.686	0.723	0.825
Reference	NCI/NTP TR-411	NCI/NTP TR-457	NCI/NTP TR-193

Model Applicability

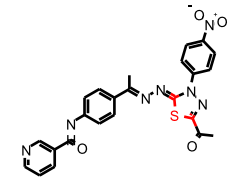
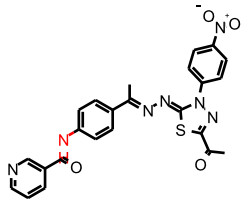
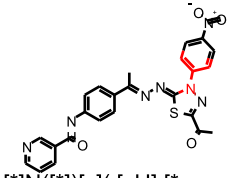
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 8: [*][N+](=[*])[*]
3. Unknown FCFP_2 feature: 5: [*][O-]
4. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
5. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
6. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
7. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

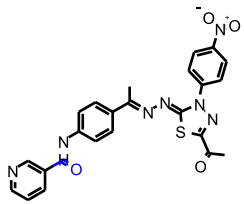
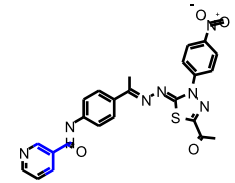
Feature Contribution

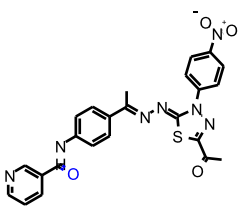
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095
FCFP_2	3	 [*]N[*]	0.0737
FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]	0.0611

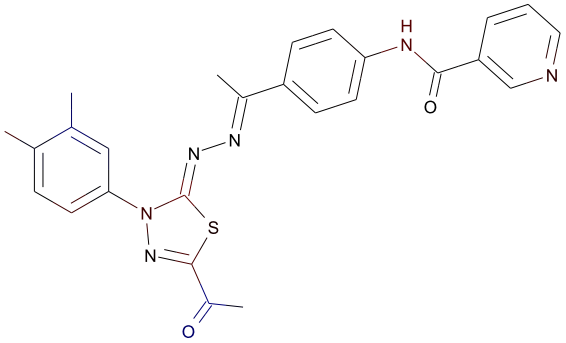
Top Features for negative contribution

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

FCFP_2	1	 [*]=O	-0.0796

7g

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



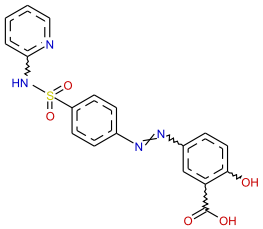
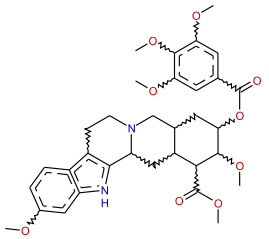
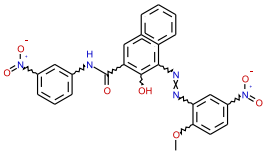
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.0326
Unit: g/kg_body_weight
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 1.22e-005

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

Structural Similar Compounds

Name	SALICYLAZOSULFAPYRIDINE	RESERPINE	C.I.PIGMENT RED 23
Structure			
Actual Endpoint (-log C)	3.375	6.13118	2.30052
Predicted Endpoint (-log C)	2.80292	4.38304	3.55333
Distance	0.720	0.735	0.795
Reference	NCI/NTP TR-457	NCI/NTP TR-193	NCI/NTP TR-411

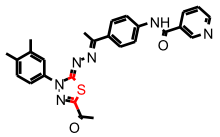
Model Applicability

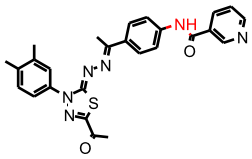
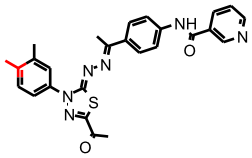
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

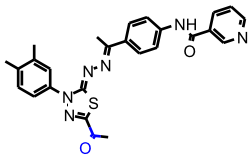
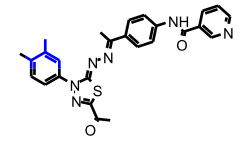
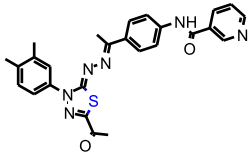
1. All properties and OPS components are within expected ranges.

Feature Contribution

Top features for positive contribution

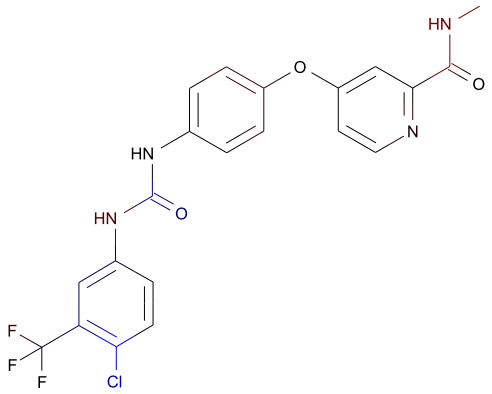
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	-1143715940	 [*]C1=[*][*]C(=[*])S1	0.095

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

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

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Feed



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.0885

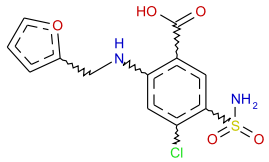
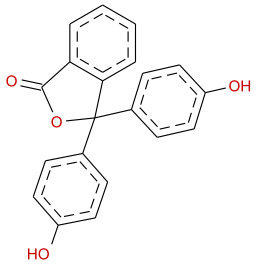
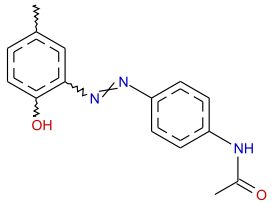
Unit: g/kg_body_weight

Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 1.76e-009

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

Structural Similar Compounds

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

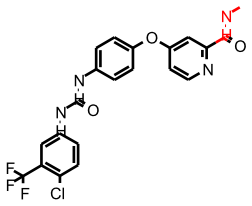
Model Applicability

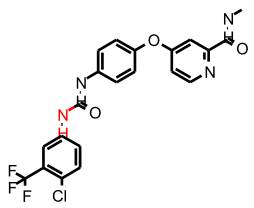
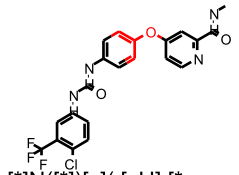
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

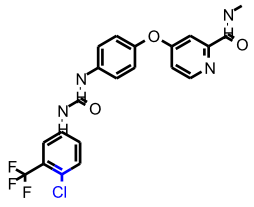
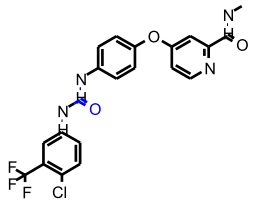
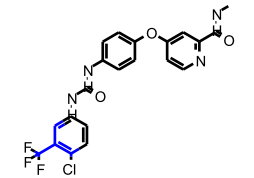
Feature Contribution

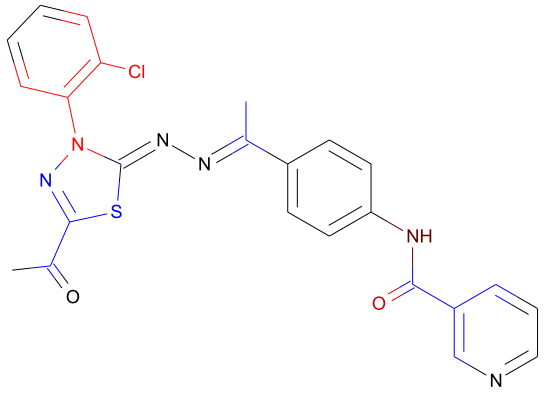
Top features for positive contribution

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

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



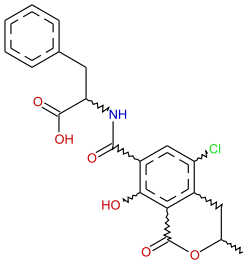
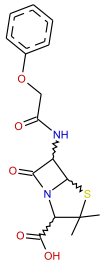
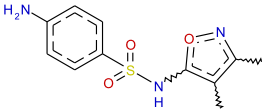
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96466
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 5.39e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 2.67e-006

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

Structural Similar Compounds

Name	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.917	1.129	1.178
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

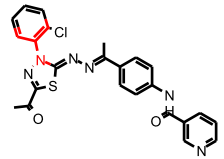
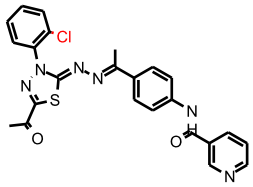
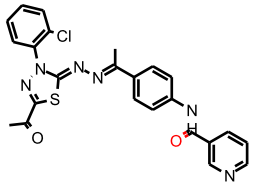
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 490.96. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC5 out of range. Value: -3.7095. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
5. OPS PC10 out of range. Value: 2.9353. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
6. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
8. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

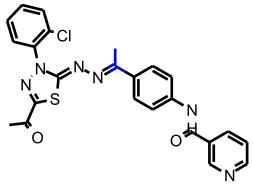
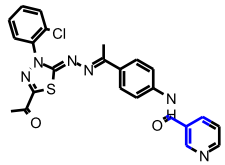
Feature Contribution

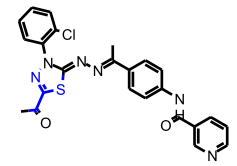
Top features for positive contribution

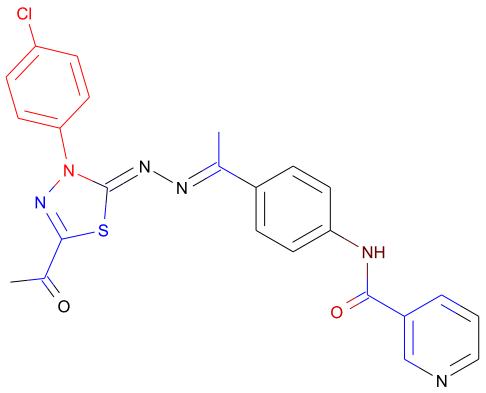
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

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

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348



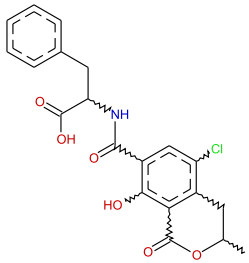
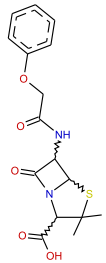
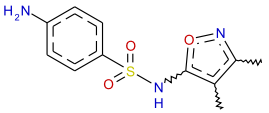
$C_{24}H_{19}ClN_6O_2S$
Molecular Weight: 490.96465
ALogP: 4.258
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 5.39e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 10.2
Mahalanobis Distance p-value: 2.67e-006

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

Structural Similar Compounds

Name	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.917	1.129	1.178
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

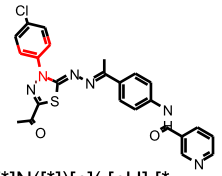
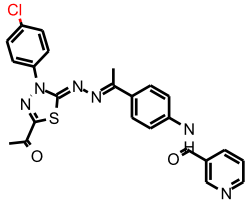
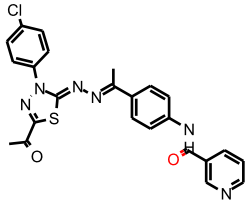
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 490.96. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC5 out of range. Value: -3.7095. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
5. OPS PC10 out of range. Value: 2.9353. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
6. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
8. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

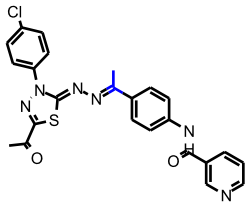
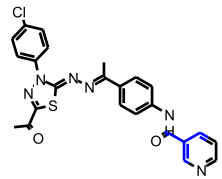
Feature Contribution

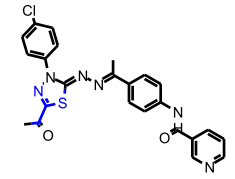
Top features for positive contribution

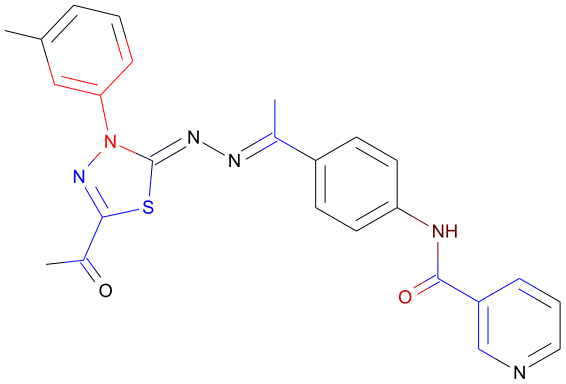
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

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

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348



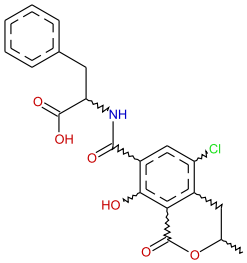
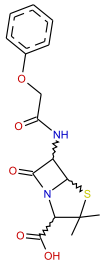
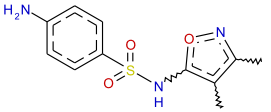
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.000513
Unit: g/kg_body_weight
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 3.18e-008

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

Structural Similar Compounds

Name	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.949	1.097	1.112
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

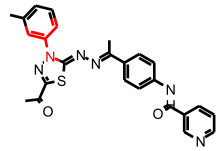
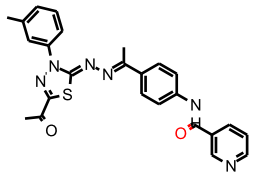
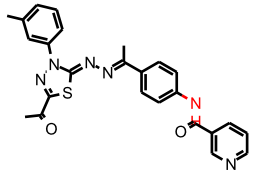
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 470.55. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC6 out of range. Value: -3.1418. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
5. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
7. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

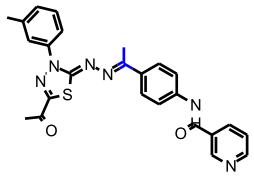
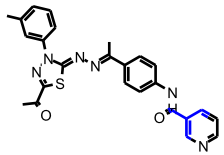
Feature Contribution

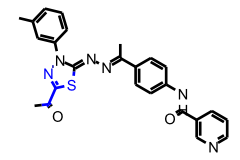
Top features for positive contribution

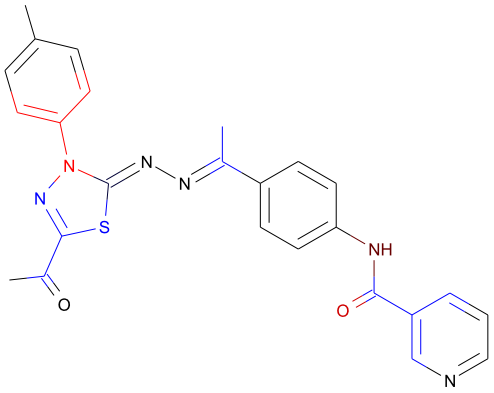
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]	0.672
FCFP_2	1	 [*]=O	0.511
FCFP_2	3	 [*]N[*]	0.104

Top Features for negative contribution

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

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348



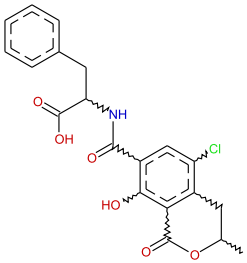
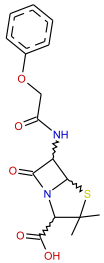
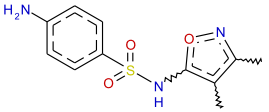
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.000513
Unit: g/kg_body_weight
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 3.18e-008

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

Structural Similar Compounds

Name	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.949	1.097	1.112
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

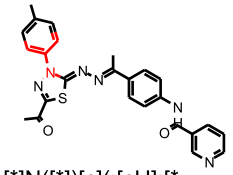
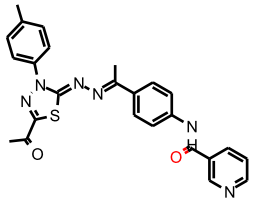
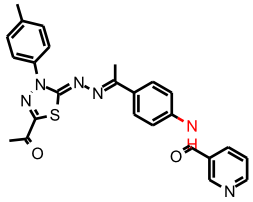
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 470.55. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC6 out of range. Value: -3.1418. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
5. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
6. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
7. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]

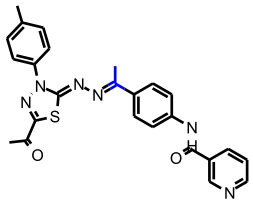
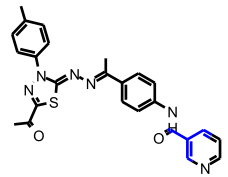
Feature Contribution

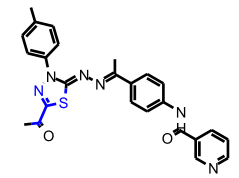
Top features for positive contribution

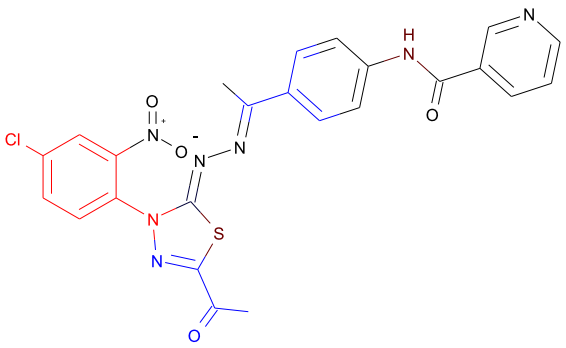
Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]	0.672
FCFP_2	1	 [*]=O	0.511
FCFP_2	3	 [*]N[*]	0.104

Top Features for negative contribution

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

FCFP_2	565998553		-0.348



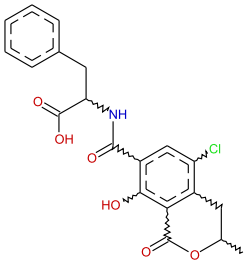
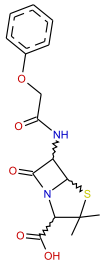
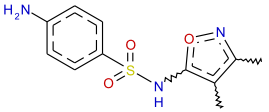
$C_{24}H_{18}ClN_7O_4S$
Molecular Weight: 535.96222
ALogP: 4.153
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 1.61e-005
Unit: g/kg_body_weight
Mahalanobis Distance: 12.9
Mahalanobis Distance p-value: 5.2e-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	1.148	1.323	1.451
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

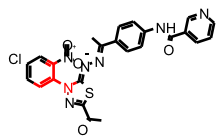
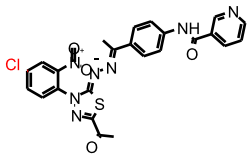
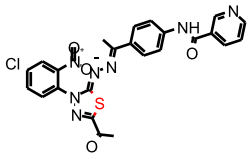
Model Applicability

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

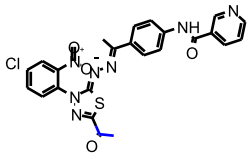
1. Molecular_Weight out of range. Value: 535.96. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 10. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. Molecular_PolarSurfaceArea out of range. Value: 170.49. Training min, max, mean, SD: 0, 138.03, 28.978, 32.1.
5. OPS PC1 out of range. Value: 8.2287. Training min, max, SD, explained variance: -4.0008, 7.9165, 2.861, 0.2531.
6. OPS PC5 out of range. Value: -4.3335. Training min, max, SD, explained variance: -3.4, 4.1587, 1.489, 0.0686.
7. OPS PC10 out of range. Value: 2.7221. Training min, max, SD, explained variance: -3.9696, 2.3971, 0.982, 0.0298.
8. Unknown FCFP_2 feature: 8: [*][N+](=[*])[*]
9. Unknown FCFP_2 feature: 5: [*][O-]
10. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
11. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]
12. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
13. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
14. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
15. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
16. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

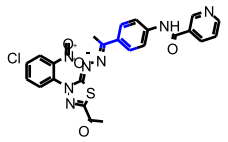
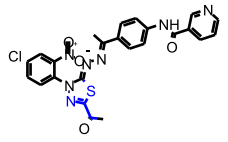
Feature Contribution

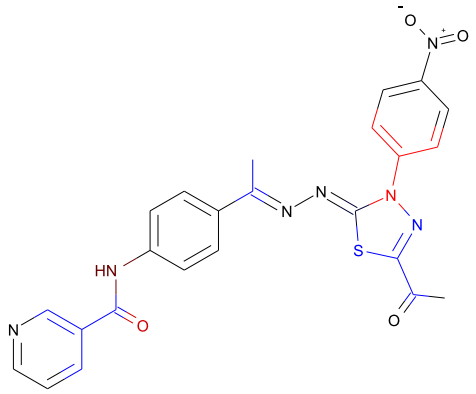
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]	0.672
FCFP_2	32	 [*]Cl	0.526
FCFP_2	1	 [*]=O	0.511

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489

FCFP_2	203677720	 [*]C(=[*])[c](:[cH]:[*]):[cH]:[*]	-0.406
FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S1	-0.348



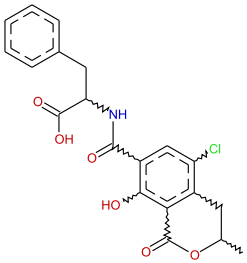
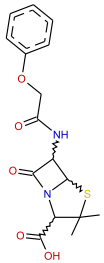
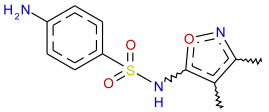
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 0.000203
Unit: g/kg_body_weight
Mahalanobis Distance: 13.1
Mahalanobis Distance p-value: 2.65e-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	1.158	1.261	1.381
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

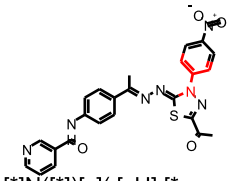
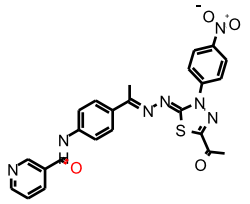
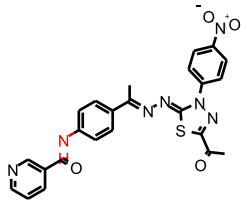
Model Applicability

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

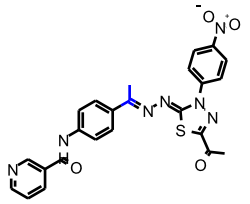
1. Molecular_Weight out of range. Value: 501.52. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 10. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. Molecular_PolarSurfaceArea out of range. Value: 170.49. Training min, max, mean, SD: 0, 138.03, 28.978, 32.1.
5. OPS PC1 out of range. Value: 8.2949. Training min, max, SD, explained variance: -4.0008, 7.9165, 2.861, 0.2531.
6. OPS PC6 out of range. Value: -3.1571. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
7. Unknown FCFP_2 feature: 8: [*][N+](=[*])[*]
8. Unknown FCFP_2 feature: 5: [*][O-]
9. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]
10. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
11. Unknown FCFP_2 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
12. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]
13. Unknown FCFP_2 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
14. Unknown FCFP_2 feature: 1872392852: [*][N+](=O)[*]
15. Unknown FCFP_2 feature: 260476081: [*][N+](=[*])[O-]

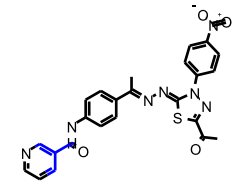
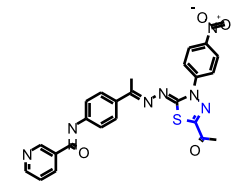
Feature Contribution

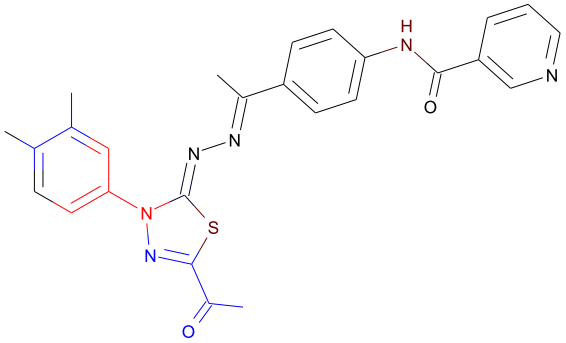
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]	0.672
FCFP_2	1	 [*]=O	0.511
FCFP_2	3	 [*]N[*]	0.104

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_2	136597326	 [*]C(=[*])C	-0.489

FCFP_2	203677720		-0.406
FCFP_2	565998553		-0.348



$C_{26}H_{24}N_6O_2S$

Molecular Weight: 484.57275

ALogP: 4.566

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: 0.000391

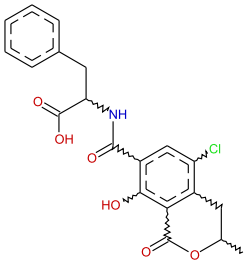
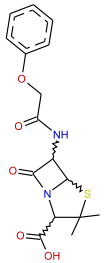
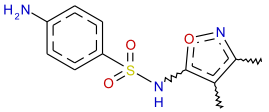
Unit: g/kg_body_weight

Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 1.53e-008

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

Structural Similar Compounds

Name	OCHRATOXIN	PENICILLIN VK	SULFISOOXAZOLE
Structure			
Actual Endpoint (-log C)	6.28396	2.54455	2.82494
Predicted Endpoint (-log C)	5.12358	3.9702	3.0705
Distance	0.967	1.133	1.151
Reference	NCI/NTP TR-358	NCI/NTP TR-336	NCI/NTP TR-138

Model Applicability

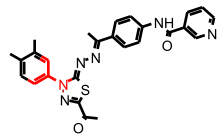
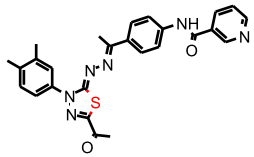
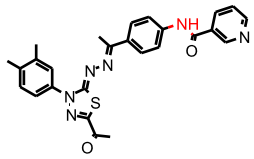
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. Molecular_Weight out of range. Value: 484.57. Training min, max, mean, SD: 68.074, 434.63, 171.13, 85.06.
2. Num_H_Acceptors out of range. Value: 8. Training min, max, mean, SD: 0, 6, 1.6146, 1.644.
3. Num_AromaticRings out of range. Value: 3. Training min, max, mean, SD: 0, 2, 0.5625, 0.693.
4. OPS PC6 out of range. Value: -3.2057. Training min, max, SD, explained variance: -2.4321, 2.9885, 1.256, 0.0488.
5. Unknown FCFP_2 feature: 580960234: [*]C(=NN=[*])[*]
6. Unknown FCFP_2 feature: -1986158408: [*]\N=C\1/S[*]=[*]N1[*]
7. Unknown FCFP_2 feature: -1549192822: [*]\N=C(/C)\[c](:[*]):[*]

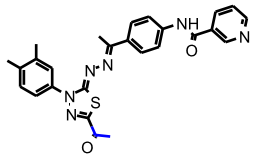
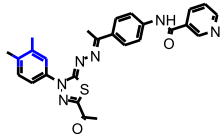
Feature Contribution

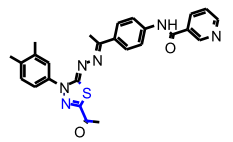
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

FCFP_2	332760439	 [*]N([*])[c](:[cH]:[*]):[c]([*]):[*]	0.672
FCFP_2	1	 [*]=O	0.511
FCFP_2	3	 [*]N[*]	0.104

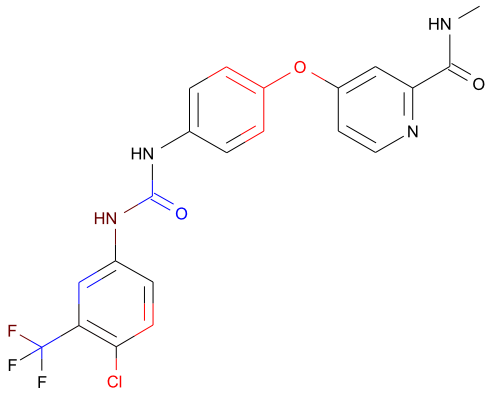
Top Features for negative contribution

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

FCFP_2	565998553	 [*]C(=[*])C1=N[*][*]S 1	-0.348

Sorafenib

TOPKAT_Rat_Maximum_Tolerated_Dose_Gavage



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 0.000918

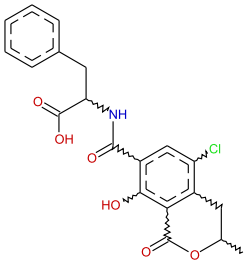
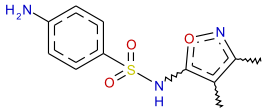
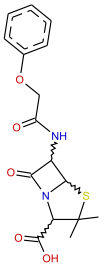
Unit: g/kg_body_weight

Mahalanobis Distance: 12.2

Mahalanobis Distance p-value: 4.69e-009

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

Structural Similar Compounds

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

Model Applicability

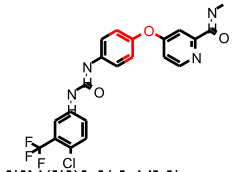
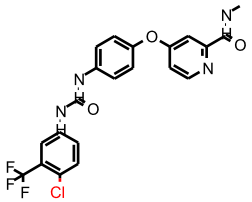
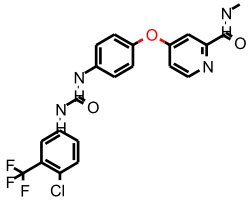
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

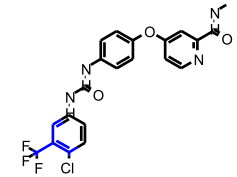
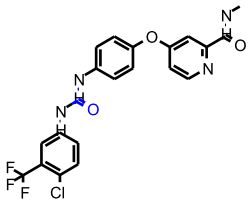
Feature Contribution

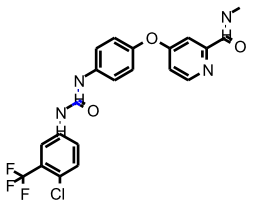
Top features for positive contribution

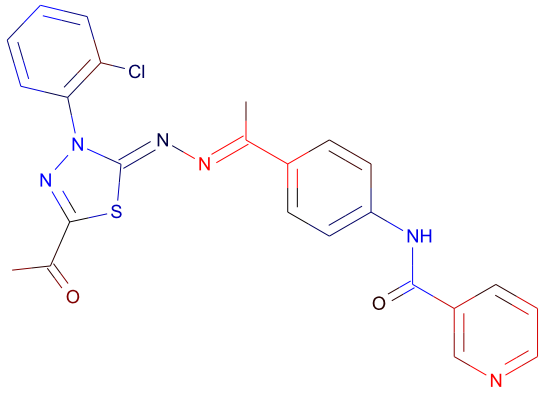
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

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

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



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96466

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: 0.531

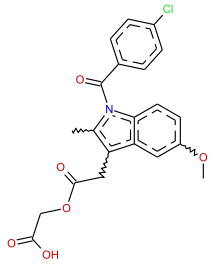
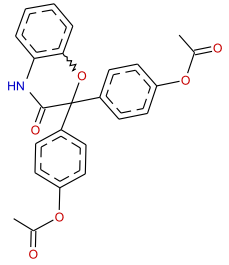
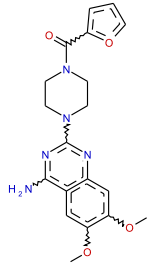
Unit: g/kg_body_weight

Mahalanobis Distance: 24.4

Mahalanobis Distance p-value: 5.41e-028

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

Structural Similar Compounds

Name	ACEMETACIN	bis-OXATIN ACETATE	PRASOZIN .HCl (HCl STRIPPED)
Structure			
Actual Endpoint (-log C)	4.235	1.717	2.294
Predicted Endpoint (-log C)	3.39415	2.40947	3.00765
Distance	0.662	0.681	0.710
Reference	ARZNAD 30;1398;80	NIIRDN 6;609;82	NIIRDN 6;688;82

Model Applicability

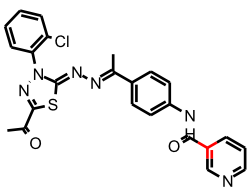
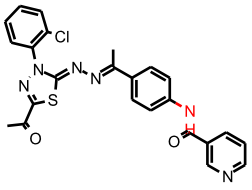
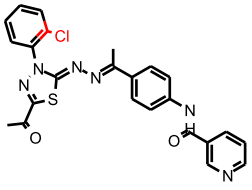
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
11. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]
12. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl

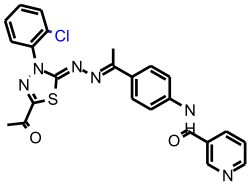
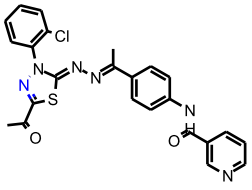
Feature Contribution

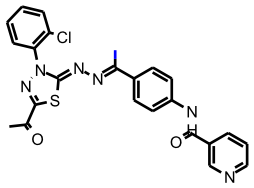
Top features for positive contribution

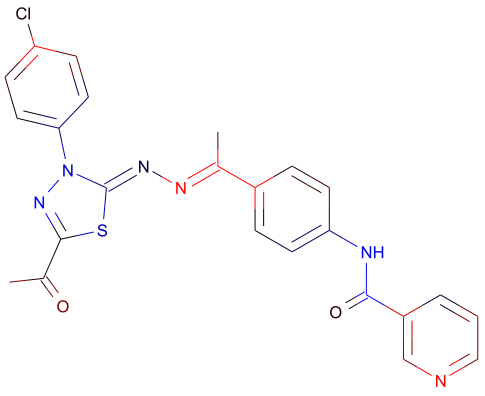
Fingerprint	Bit/Smiles	Feature Structure	Score

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	-817402818	 [*]Cl	-0.263
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201



$C_{24}H_{19}ClN_6O_2S$

Molecular Weight: 490.96465

ALogP: 4.258

Rotatable Bonds: 6

Acceptors: 8

Donors: 1

Model Prediction

Prediction: 0.247

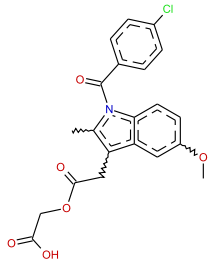
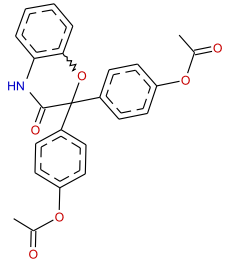
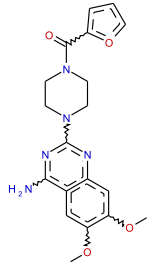
Unit: g/kg_body_weight

Mahalanobis Distance: 24.3

Mahalanobis Distance p-value: 3.27e-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	ACEMETACIN	bis-OXATIN ACETATE	PRASOZIN .HCl (HCl STRIPPED)
Structure			
Actual Endpoint (-log C)	4.235	1.717	2.294
Predicted Endpoint (-log C)	3.39415	2.40947	3.00765
Distance	0.655	0.687	0.712
Reference	ARZNAD 30;1398;80	NIIRDN 6;609;82	NIIRDN 6;688;82

Model Applicability

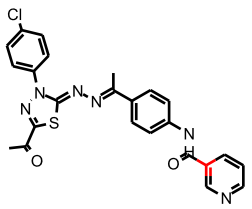
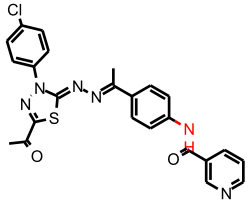
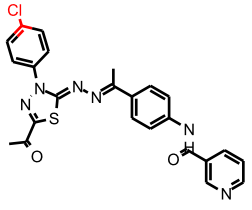
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
11. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl
12. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]

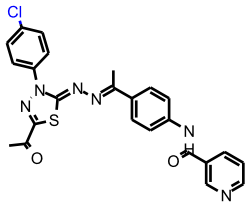
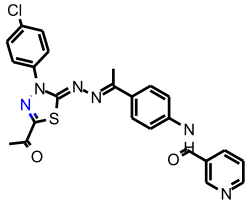
Feature Contribution

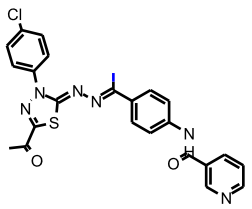
Top features for positive contribution

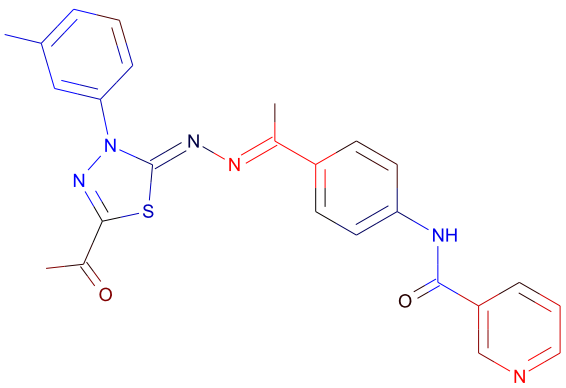
Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-817402818	 [*]Cl	-0.263
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201



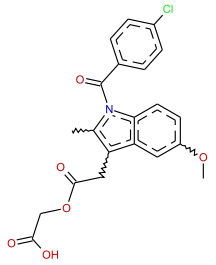
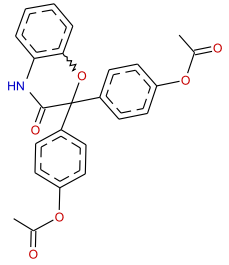
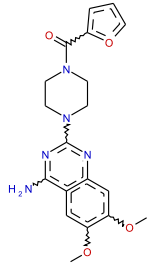
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.418
Unit: g/kg_body_weight
Mahalanobis Distance: 24.2
Mahalanobis Distance p-value: 6.26e-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	ACEMETACIN	bis-OXATIN ACETATE	PRASOZIN .HCl (HCl STRIPPED)
Structure			
Actual Endpoint (-log C)	4.235	1.717	2.294
Predicted Endpoint (-log C)	3.39415	2.40947	3.00765
Distance	0.659	0.662	0.674
Reference	ARZNAD 30;1398;80	NIIRDN 6;609;82	NIIRDN 6;688;82

Model Applicability

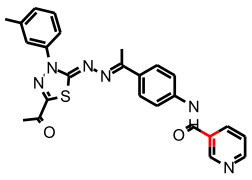
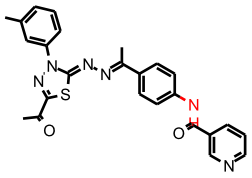
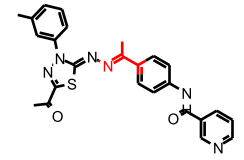
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
11. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]

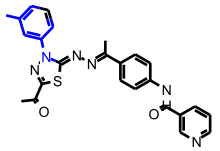
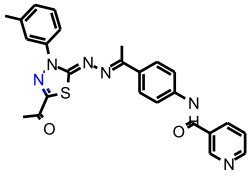
Feature Contribution

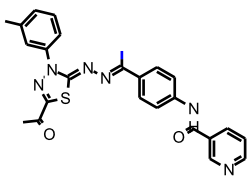
Top features for positive contribution

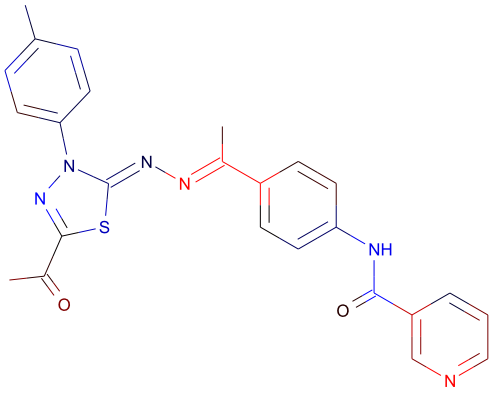
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]) :[*]	0.168

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1676877079	 [*]N([*])[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	-0.254
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201



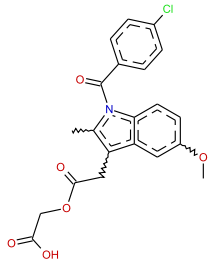
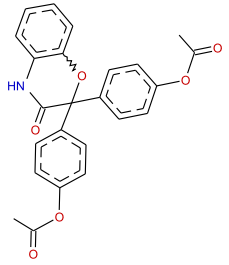
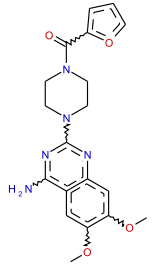
$C_{25}H_{22}N_6O_2S$
Molecular Weight: 470.54617
ALogP: 4.08
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.33
Unit: g/kg_body_weight
Mahalanobis Distance: 24.2
Mahalanobis Distance p-value: 6.26e-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	ACEMETACIN	bis-OXATIN ACETATE	PRASOZIN .HCl (HCl STRIPPED)
Structure			
Actual Endpoint (-log C)	4.235	1.717	2.294
Predicted Endpoint (-log C)	3.39415	2.40947	3.00765
Distance	0.660	0.663	0.675
Reference	ARZNAD 30;1398;80	NIIRDN 6;609;82	NIIRDN 6;688;82

Model Applicability

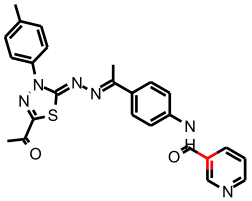
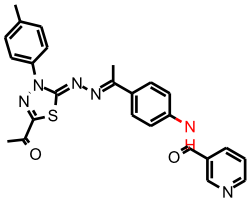
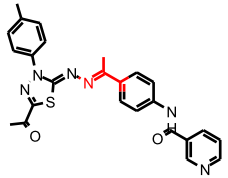
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
3. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
7. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
8. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
9. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
10. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
11. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]

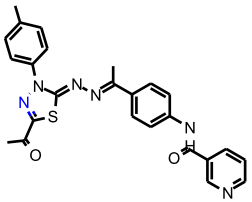
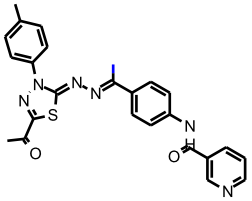
Feature Contribution

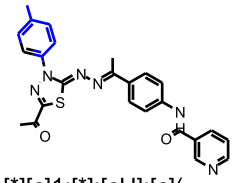
Top features for positive contribution

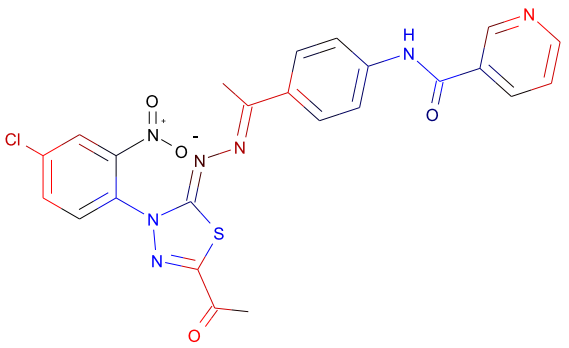
Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]) :[*]	0.168

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]N=[*]	-0.239
ECFP_6	734603939	 [*]C	-0.201

ECFP_6	-210573707	 [*][c]1:[*]:[cH]:[c](C):[cH]:[cH]:1	-0.122



$C_{24}H_{18}ClN_7O_4S$

Molecular Weight: 535.96222

ALogP: 4.153

Rotatable Bonds: 7

Acceptors: 10

Donors: 1

Model Prediction

Prediction: 0.353

Unit: g/kg_body_weight

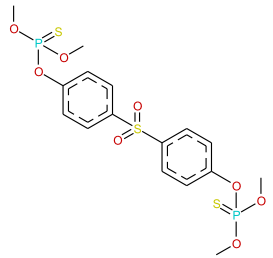
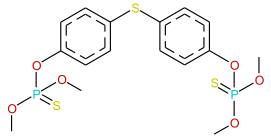
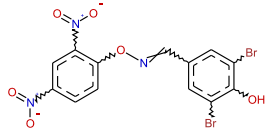
Mahalanobis Distance: 28.6

Mahalanobis Distance p-value: 1.02e-053

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	PHOSPHOROTHIOIC ACID; O;O'-(SULFONYLDI-p-PHENYLENE) O;O;O';O'-TETRAMETHYL ESTER	TEMEPHOS	BROMOFENOXIME
Structure			
Actual Endpoint (-log C)	2.397	2.669	2.622
Predicted Endpoint (-log C)	3.8515	3.65871	3.41798
Distance	0.768	0.803	0.813
Reference	TXAPA9 21;315;72	GUCHAZ 6;493;73	85ARAE 2;203;77

Model Applicability

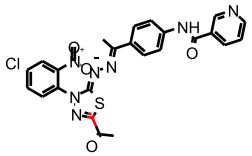
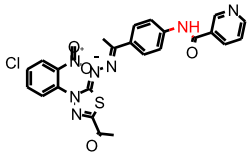
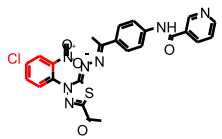
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1043790491: [*][N+](=[*])[*]
3. Unknown ECFP_2 feature: 781519895: [*][O-]
4. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
5. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
6. Unknown ECFP_2 feature: -1956535100: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
7. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
8. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
9. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
10. Unknown ECFP_2 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
11. Unknown ECFP_2 feature: 2104376220: [*][N+](=O)[*]
12. Unknown ECFP_2 feature: -659271057: [*][N+](=[*])[O-]
13. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
14. Unknown FCFP_6 feature: 8: [*][N+](=[*])[*]
15. Unknown FCFP_6 feature: 5: [*][O-]
16. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
17. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]

18. Unknown FCFP_6 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
19. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
20. Unknown FCFP_6 feature: 71476542: [*]:[c](:[*])Cl
21. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]
22. Unknown FCFP_6 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
23. Unknown FCFP_6 feature: 1872392852: [*][N+](=O)[*]
24. Unknown FCFP_6 feature: 260476081: [*][N+](=[*])[O-]

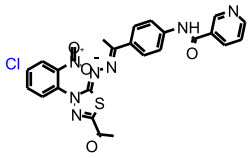
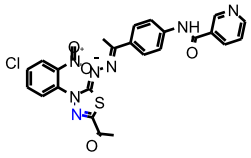
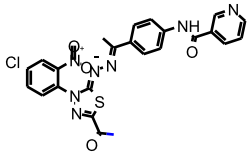
Feature Contribution

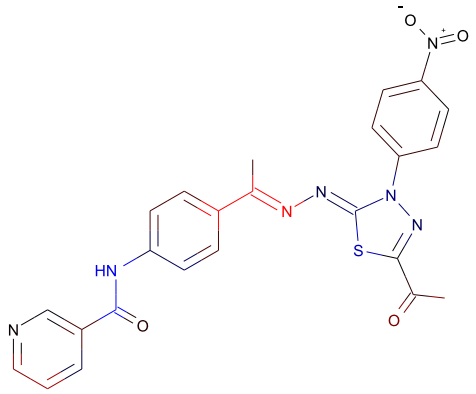
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
ECFP_6	577592657	 [*][c]1:[*]:[cH]:[cH] :[c](Cl):[cH]:1	0.194

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	-817402818	 [*]Cl	-0.263
ECFP_6	655739385	 [*]N=[*]	-0.239
ECFP_6	734603939	 [*]C	-0.201



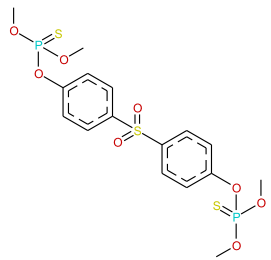
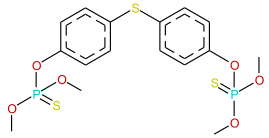
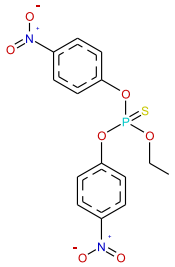
$C_{24}H_{19}N_7O_4S$
Molecular Weight: 501.51715
ALogP: 3.488
Rotatable Bonds: 7
Acceptors: 10
Donors: 1

Model Prediction

Prediction: 0.272
Unit: g/kg_body_weight
Mahalanobis Distance: 28
Mahalanobis Distance p-value: 1.32e-049

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	PHOSPHOROTHIOIC ACID; O;O'-(SULFONYLDI-p-PHENYLENE) O;O;O';O'-TETRAMETHYL ESTER	TEMEPHOS	PHOSPHOROTHIOIC ACID; O;O-bis-p-(NITROPHENYL); O-ETHYL ESTER
Structure			
Actual Endpoint (-log C)	2.397	2.669	3.759
Predicted Endpoint (-log C)	3.8515	3.65871	4.23595
Distance	0.758	0.803	0.809
Reference	TXAPA9 21;315;72	GUCHAZ 6;493;73	85JCAE -;1169;86

Model Applicability

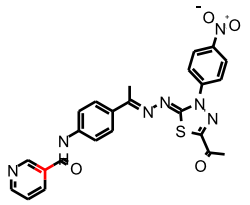
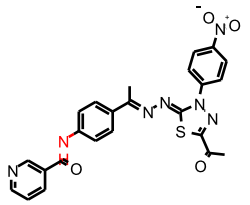
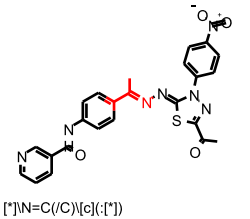
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1043790491: [*][N+](=[*])[*]
3. Unknown ECFP_2 feature: 781519895: [*][O-]
4. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
5. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
6. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
7. Unknown ECFP_2 feature: -179073144: [*][N+](=[*])[c](:[cH]:[*]):[cH]:[*]
8. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
9. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
10. Unknown ECFP_2 feature: -215026467: [*]:[c](:[*])[N+](=O)[O-]
11. Unknown ECFP_2 feature: 2104376220: [*][N+](=O)[*]
12. Unknown ECFP_2 feature: -659271057: [*][N+](=[*])[O-]
13. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
14. Unknown FCFP_6 feature: 8: [*][N+](=[*])[*]
15. Unknown FCFP_6 feature: 5: [*][O-]
16. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
17. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]

18. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
19. Unknown FCFP_6 feature: -828984032: [*][c](:[*]):[c](:[cH]:[*])[N+](=[*])[*]
20. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
21. Unknown FCFP_6 feature: -1338588315: [*]:[c](:[*])[N+](=O)[O-]
22. Unknown FCFP_6 feature: 1872392852: [*][N+](=O)[*]
23. Unknown FCFP_6 feature: 260476081: [*][N+](=[*])[O-]

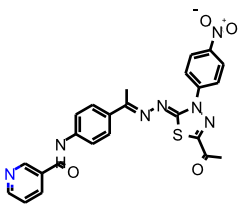
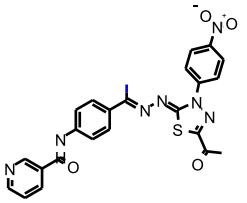
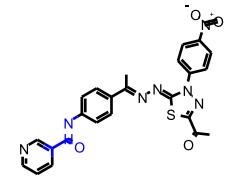
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	-1897341097	 [*]N[*]	0.216
FCFP_6	-1549192822	 [*]\N=C(/C)\[c](:[*]) :[*]	0.168

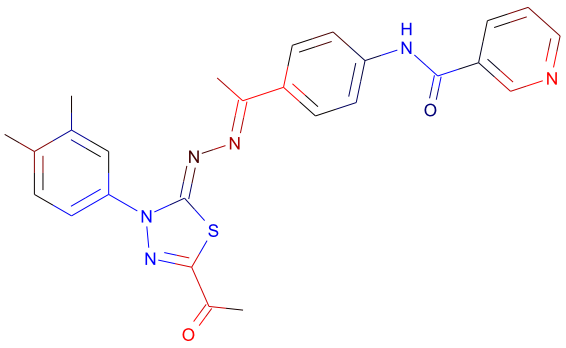
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	655739385	 [*]N=[*]	-0.239
ECFP_6	734603939	 [*]C	-0.201
FCFP_6	-1549103449	 [*]NC(=O)[c](:[*]):[*]	-0.117

7g

TOPKAT_Rat_Oral_LD50



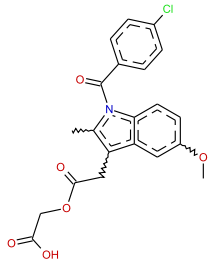
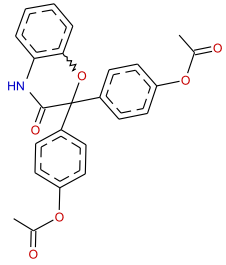
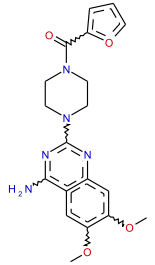
$C_{26}H_{24}N_6O_2S$
Molecular Weight: 484.57275
ALogP: 4.566
Rotatable Bonds: 6
Acceptors: 8
Donors: 1

Model Prediction

Prediction: 0.331
Unit: g/kg_body_weight
Mahalanobis Distance: 24.5
Mahalanobis Distance p-value: 2.36e-028

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

Structural Similar Compounds

Name	ACEMETACIN	bis-OXATIN ACETATE	PRASOZIN .HCl (HCl STRIPPED)
Structure			
Actual Endpoint (-log C)	4.235	1.717	2.294
Predicted Endpoint (-log C)	3.39415	2.40947	3.00765
Distance	0.675	0.686	0.718
Reference	ARZNAD 30;1398;80	NIIRDN 6;609;82	NIIRDN 6;688;82

Model Applicability

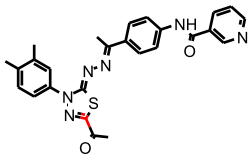
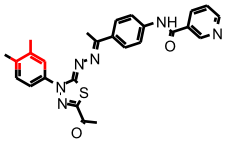
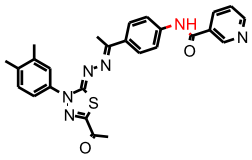
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -934225701: [*]C(=[*])C1=N[*][*]S1
3. Unknown ECFP_2 feature: 189949281: [*]\N=C\1/S[*]=[*]N1[*]
4. Unknown ECFP_2 feature: -819426257: [*]C(=NN=[*])[*]
5. Unknown ECFP_2 feature: 562081661: [*]C(=NN=[*])[*]
6. Unknown ECFP_2 feature: 128986386: [*]\N=C(/C)\[c](:[*]):[*]
7. Unknown FCFP_6 feature: 16: [*][c](:[*]):[*]
8. Unknown FCFP_6 feature: 580960234: [*]C(=NN=[*])[*]
9. Unknown FCFP_6 feature: 675799546: [*]=C1[*][*]=NN1[c](:[*]):[*]
10. Unknown FCFP_6 feature: 1618154665: [*][c](:[*]):[cH]:[cH]:[*]
11. Unknown FCFP_6 feature: 1747237384: [*]:[cH]:n:[cH]:[*]

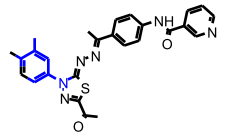
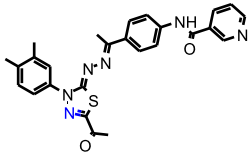
Feature Contribution

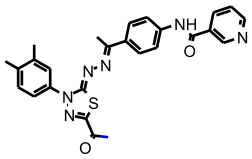
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

ECFP_6	642810091	 [*][c](:[*]):[*]	0.281
ECFP_6	2147419938	 [*][c](:[*]):[c](C):[cH]:[*]	0.263
ECFP_6	-1897341097	 [*]N[*]	0.216

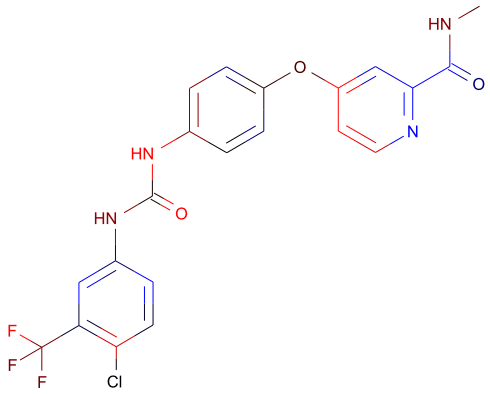
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1676877079	 [*]N([*])[c]1:[cH]:[*]:[cH]:[c](C):[cH]:1	-0.254
ECFP_6	655739385	 [*]N=[*]	-0.239

ECFP_6	734603939	 [*]C	-0.201

Sorafenib

TOPKAT_Rat_Oral_LD50



$C_{21}H_{16}ClF_3N_4O_3$

Molecular Weight: 464.82494

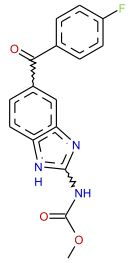
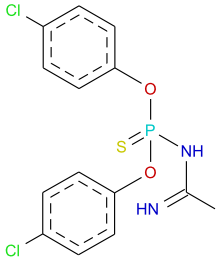
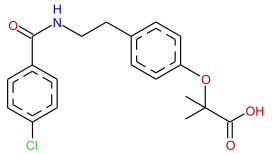
ALogP: 4.175

Rotatable Bonds: 6

Acceptors: 4

Donors: 3

Structural Similar Compounds

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

Model Prediction

Prediction: 0.823

Unit: g/kg_body_weight

Mahalanobis Distance: 21

Mahalanobis Distance p-value: 1.93e-012

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

Model Applicability

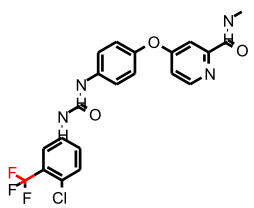
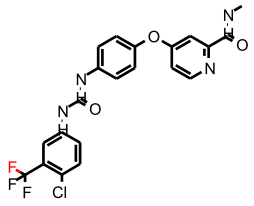
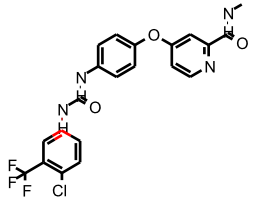
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

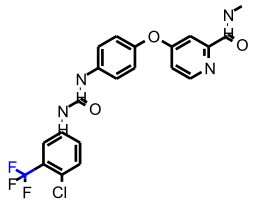
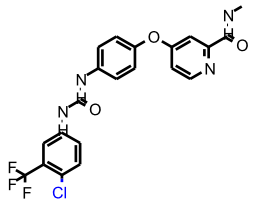
Feature Contribution

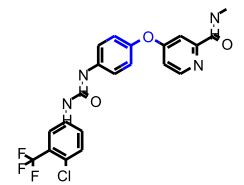
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score

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

Top Features for negative contribution

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

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