Electronic Supplementary Material (ESI) for RSC Advances.

Design, synthesis, and biological evaluation of novel bioactive thalidomide analogs as anticancer immunomodulatory agents

Anas Ramadan Kotb[a], Dina A. Bakhotmah[b]*, Abdallah E. Abdallah[a], Hazem Elkady[a], Mohammed S. Taghour[a], Ibrahim. H. Eissa[a], Mohamed Ayman El-Zahabi[a]*

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

[b] Department of Chemistry, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia.

* **Corresponding author:**

Mohamed Ayman El-Zahabi

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

Email: malzahaby@azhar.edu.eg

Ibrahim. H. Eissa

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

Email: Ibrahimeissa@azhar.edu.eg

Dina A. Bakhotmah

Department of Chemistry, Faculty of Science, King Abdulaziz University, Jeddah, kingdom of Saudi Arabia.

Email: dbakhotmah@kau.edu.sa

Content

1- Biological testing

a- *In vitro* anti-proliferative activities

Anti-proliferative activity screening of the synthesized compounds was performed against three human cancer cell lines: hepatocellular carcinoma (HepG-2), prostate cancer (PC3), and mammary gland cancer (MCF-7), using MTT assay as follows: The cells were cultured in RPMI-1640 medium with 10% fetal bovine serum. Penicillin (100 units/ml) and streptomycin (100µg/ml) were added at 37 C in a 5% CO_2 incubator. The cells were seeded in 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 the synthesized compounds and incubated for 24 h. Then, 20 µl of MTT solution of a concentration of 5mg/ml was added and incubated for 4 h. DMSO (100 µl) was added into each well to dissolve the formed purple formazan. The color intensity was measured and recorded at absorbance of 570 nm using a plate reader (EXL 800, USA). The percentage of cell viability in was calculated as (A570 of treated samples/A570 of untreated sample) X 100.

b- Supernatant preparation

Cell culture supernatants were prepared from HepG-2 cell line. Cells were cultured in RPMI 1640 medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) having 100 U/ml of penicillin and 50 µg/ml of streptomycin at 37°C in a humidified 5% CO_2/air mixture. The candidate compounds and thalidomide were dissolved in DMSO as a stock solution at 100 mmol/L and diluted with an FBS-free medium to achieve the designated concentrations (10 µM). The same concentration of DMSO without any compounds was used as a control. The HepG-2 cells were cultured onto 6-well plates (1x106 cells/ well), allowed to adhere for 24 h followed by treatment with the tested compounds. After 72 h, the media was collected and centrifuged for 15 min at 5000 rpm and equal volume of cell culture supernatants were collected and utilized for immunoassay using different kits.

c- Estimation of TNF-α, CASP8, and VEGF in HepG-2 supernatant

The levels of TNF-α, CASP8, and VEGF in cell culture supernatants were estimated by ELISA technique using commercially available matched paired antibodies (R&D Systems Inc., Minneapolis, MN) according to reported procedure.

d- Cell lysate preparation

The tested compounds and thalidomide were incubated for 72 h with HepG-2 cells. Then, the cells were treated with trypsin/EDTA solution (0.25 mM trypsin and 1 mM EDTA dissolved in a phosphate buffer). Cell lysate were washed three times with phosphate buffer saline (PBS, Sigma Chemical Company, St. Louis, MO, USA) and lysed by three repetitive freezing/thawing cycles (thawing at 37 °C for 2 min and freezing at –80°C for 15 min), followed by homogenization of the cells by passing through a 20G needle.

e- Estimation of nuclear factor kappa-B P65 (NF-κB P65) in HepG-2 cell lysate

The cell lysate samples were applied to the microtiter plates in a concentration of 50 µL/well. The plates were incubated at 37°C for 1 h, then kept at 4°C for 12 h in a humidified chamber. Next, the cell lysate was extracted from the wells and the plates were washed three times with a buffer (PBS/0.05%Tween-20). Blocking buffer (PBS/0.05%Tween-20/5% FBS) (200 µL) was added to each well and incubated at 37 °C for 1.5 h then washed. Anti-rabbit NF-κB P65 polyclonal antibody was dispensed as 50 µL/well and incubated for 2 h at 37 °C. The plates were washed and incubated for 1 h with 50 µL/well of diluted polyclonal goat anti-rabbit-peroxidase conjugate (1:1000). After that, the plates were washed four times followed by addition of TMB and H_2O_2 in equal volume (50 µL/well). After the development of color, 50 µL of stopping buffer (1 M HCl) were added per well. The absorbance was measured at 450 nm using the ELISA plate reader (FLUOstar OPTIMA)

2- *In silico* studies

1- Pharmacokinetic profiling study

To identify potential drug candidates, drug-likeness profiles were developed for the preliminary estimation of the physicochemical, drug-like parameters, and pharmacokinetics using Discovery studio 4.0 utilizing Lipinski's and Veber's rules. Lipinski suggested that the absorption of an orally administered compound is more likely to be better if the molecule satisfies at least three out of four of the following rules: (1) H bond donors (OH, NH, and SH) ≤ 5; (2) H bond acceptors (N, O, and S atoms) ≤ 10; (3) molecular weight < 500; (4) logP < 5. Compounds violating more than one of these rules could not have good bioavailability. Moreover, reduced molecular flexibility, as

measured by the number of rotatable bonds, and low polar surface area is found to be an important predictors of good oral bioavailability. Compounds that meet the criteria of 10 or fewer rotatable bonds and polar surface area equal to or less than 140 A will have a high probability of good oral bioavailability.

2- ADMET studies

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

3- Toxicity studies

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

4- Chemistry and materials

All solvents and reagents were commercially available and used without further purification. Progress of reactions was monitored by TLC using TLC sheets coated with UV fluorescent silica gel (Kieselgel 0.25mm, 60 F254, Merck Germany) with a developing solvent system of DCM/methanol (95:5) and was visualized using UV lamp. The melting points were determined using a Gallen lamp melting point apparatus. Elemental analyses were accomplished using a CHN analyzer. The infrared spectra were recorded on FT/IR-6600typeA spectrophotometer. ^{1}H NMR spectra were recorded at 400 and 700 MHz, while ^{13}C NMR spectra were run at 100 and 176 MHz, on a Bruker Avance NEO-600 equipped with a 1.7 mm TCI CryoProbe. Chemical shifts were expressed in δ (ppm) with reference to TMS and coupling constant (J) in Hertz using DMSO-d_6 and $CDCl_3$-d_6 as solvents. The mass spectra were recorded on an Agilent 6410 triple-quadrupole mass spectrometer equipped with an ESI source.

5- Spectral and elemental analysis

Lab code	Compound No.
AR-7	6
AR-8	7a
AR-9	7b
AR-10	24b
AR-11	24a
AR-12	24c
AR-13	13a
AR-14	15a
AR-15	11a
AR-17	19b
AR-18	19a
H2-1	13b
H2-2	15b
H2-3	11b

Viability assay against normal cell line

MTT protocol

Determination of sample cytotoxicity on cells (MTT protocol)

1-the 96 well tissue culture plate was inoculated with 1 X 10^5 cells / ml (100 ul / well) and incubated at 37°C for 24 hours to develop a complete monolayer sheet.

2- Growth medium was decanted from 96 well micro titer plates after confluent sheet of cells were formed, cell monolayer was washed twice with wash media.

3- two-fold dilutions of tested sample was made in RPMI medium with 2% serum (maintenance medium).

4- 0.1 ml of each dilution was tested in different wells leaving 3 wells as control, receiving only maintenance medium.

5- Plate was incubated at 37°C and examined. Cells were checked for any physical signs of toxicity, e.g. partial or complete loss of the monolayer, rounding, shrinkage, or cell granulation.

6- MTT solution was prepared (5mg/ml in PBS) (BIO BASIC CANADA INC).

8- 20ul MTT solution were added to each well. Place on a shaking table, 150rpm for 5 minutes, to thoroughly mix the MTT into the media.

9) Incubate (37C, 5% CO2) for 4 hours to allow the MTT to be metabolized.

10) Dump off the media. (dry plate on paper towels to remove residue if necessary.

11) Resuspend formazan (MTT metabolic product) in 200ul DMSO. Place on a shaking table, 150rpm for 5 minutes, to thoroughly mix the formazan into the solvent.

12) Read optical density at 560nm and subtract background at 620nm. Optical density should be directly correlated with cell quantity.

Morphological assay

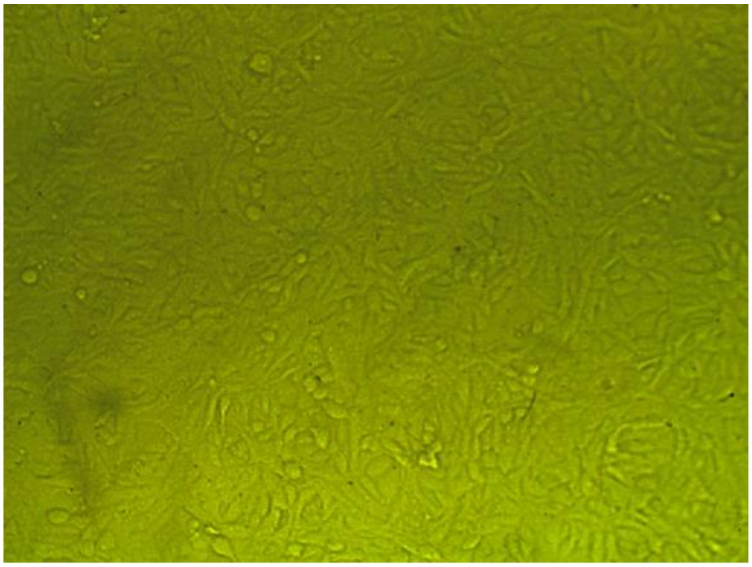
❖ Large-scale, morphological changes that occur at the cell surface, or in the cytoskeleton, can be followed and related to cell viability.

❖ Damage can be identified by large decreases in volume secondary to losses in protein and intracellular ions of due to altered permeability to sodium or potassium.

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

❖ Apoptotic cells: cell shrinkage, nuclear condansation, nuclear fragmentation

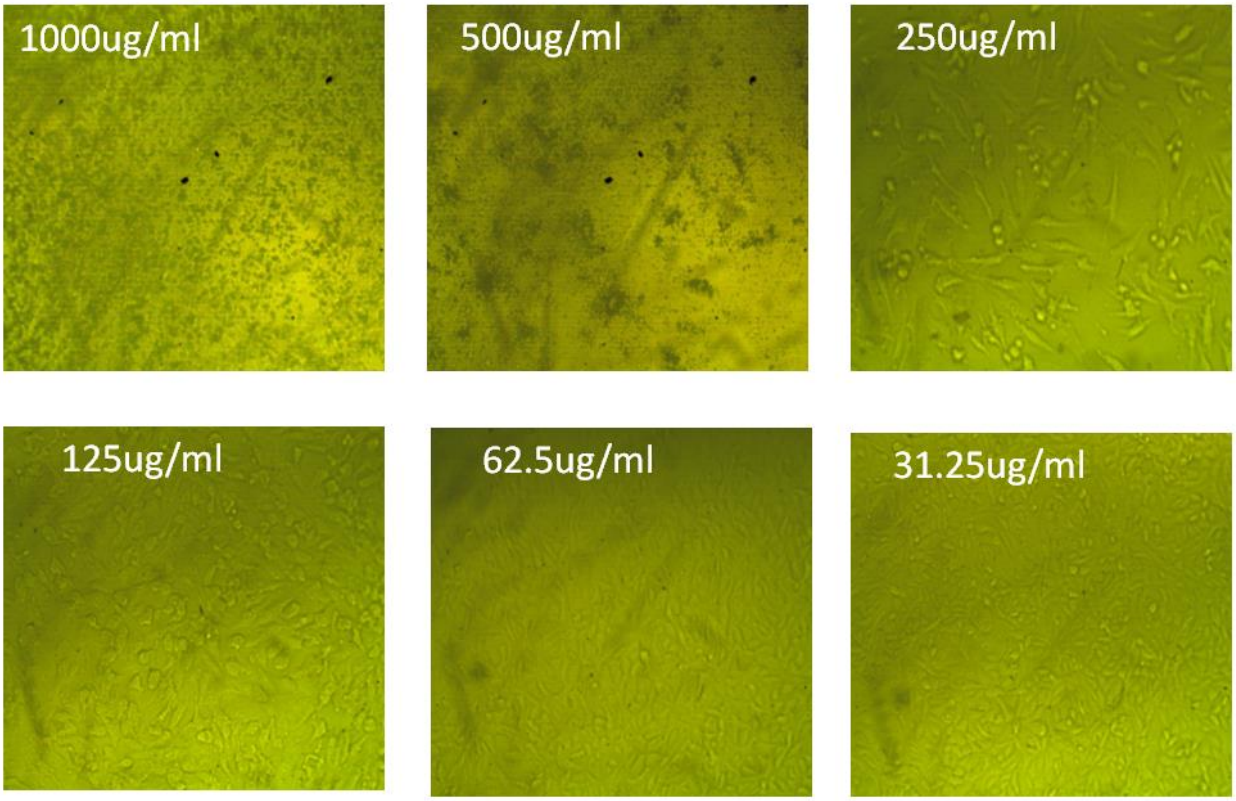
ID	ug/ml	O.D			Mean O.D	±SE	Viability %	Toxicity %	IC50 ± SD
vero	--------	0.842	0.833	0.83	0.835	0.003606	100	0	ug
AR 9	1000	0.022	0.026	0.018	0.022	0.002309	2.634730539	97.36526946	163.61 ± 1.12
	500	0.053	0.058	0.062	0.057667	0.002603	6.906187625	93.09381238	
	250	0.174	0.168	0.174	0.172	0.002	20.5988024	79.4011976	
	125	0.435	0.451	0.439	0.441667	0.004807	52.89421158	47.10578842	
	62.5	0.832	0.84	0.827	0.833	0.003786	99.76047904	0.239520958	
	31.25	0.833	0.841	0.831	0.835	0.003055	100	0	
AR 10	1000	0.019	0.021	0.019	0.019667	0.000667	2.355289421	97.64471058	161.12 ± 1.42
	500	0.032	0.035	0.052	0.039667	0.006227	4.750499002	95.249501	
	250	0.214	0.22	0.19	0.208	0.009165	24.91017964	75.08982036	
	125	0.436	0.436	0.447	0.439667	0.003667	52.65469062	47.34530938	
	62.5	0.738	0.742	0.73	0.736667	0.003528	88.22355289	11.77644711	
	31.25	0.828	0.841	0.832	0.833667	0.003844	99.84031936	0.159680639	

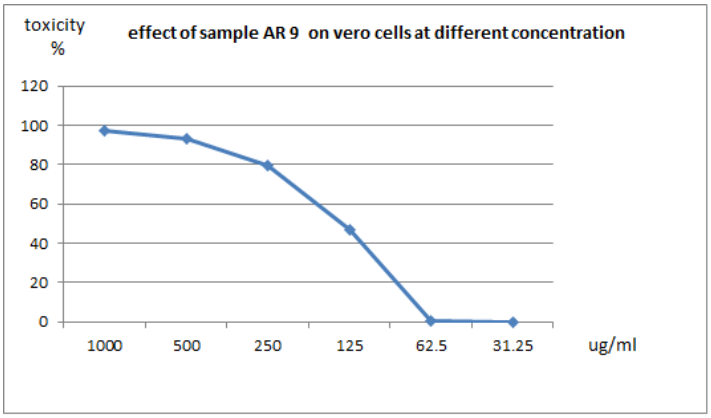


control
vero cells

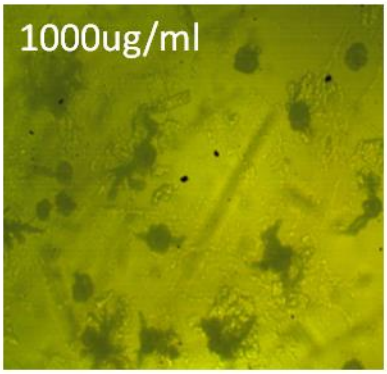
Organism :	*Cercopithecus aethiops*
Tissue :	kidney
Cell Type :	epithelial
Culture Properties :	adherent
Disease :	normal
ATCC :	CCL-81

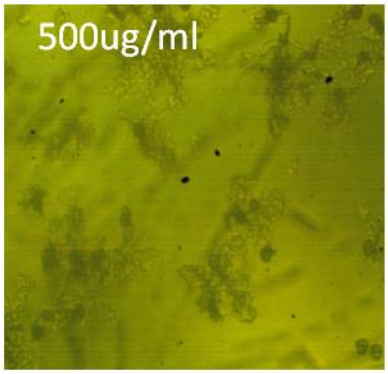
Effect of sample AR 9 on vero cells at different concentration

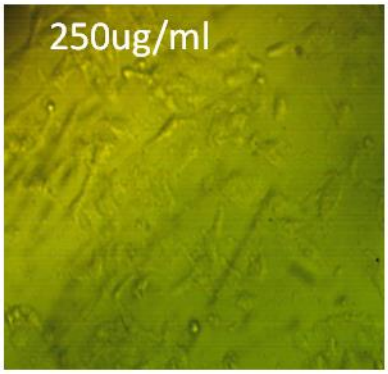


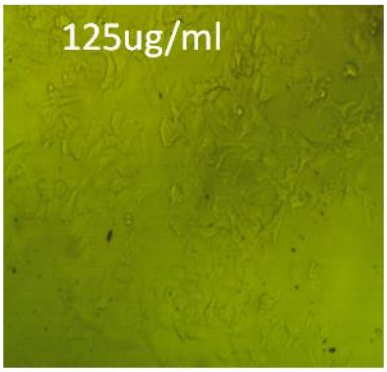


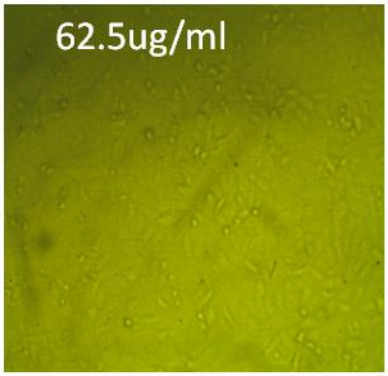
Effect of sample AR 10 on vero cells at different concentration

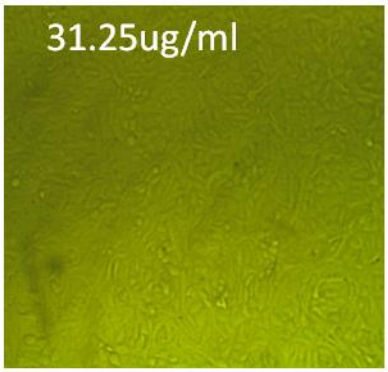


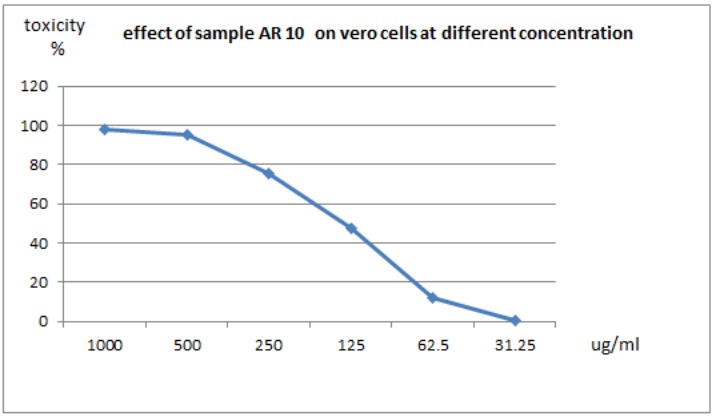




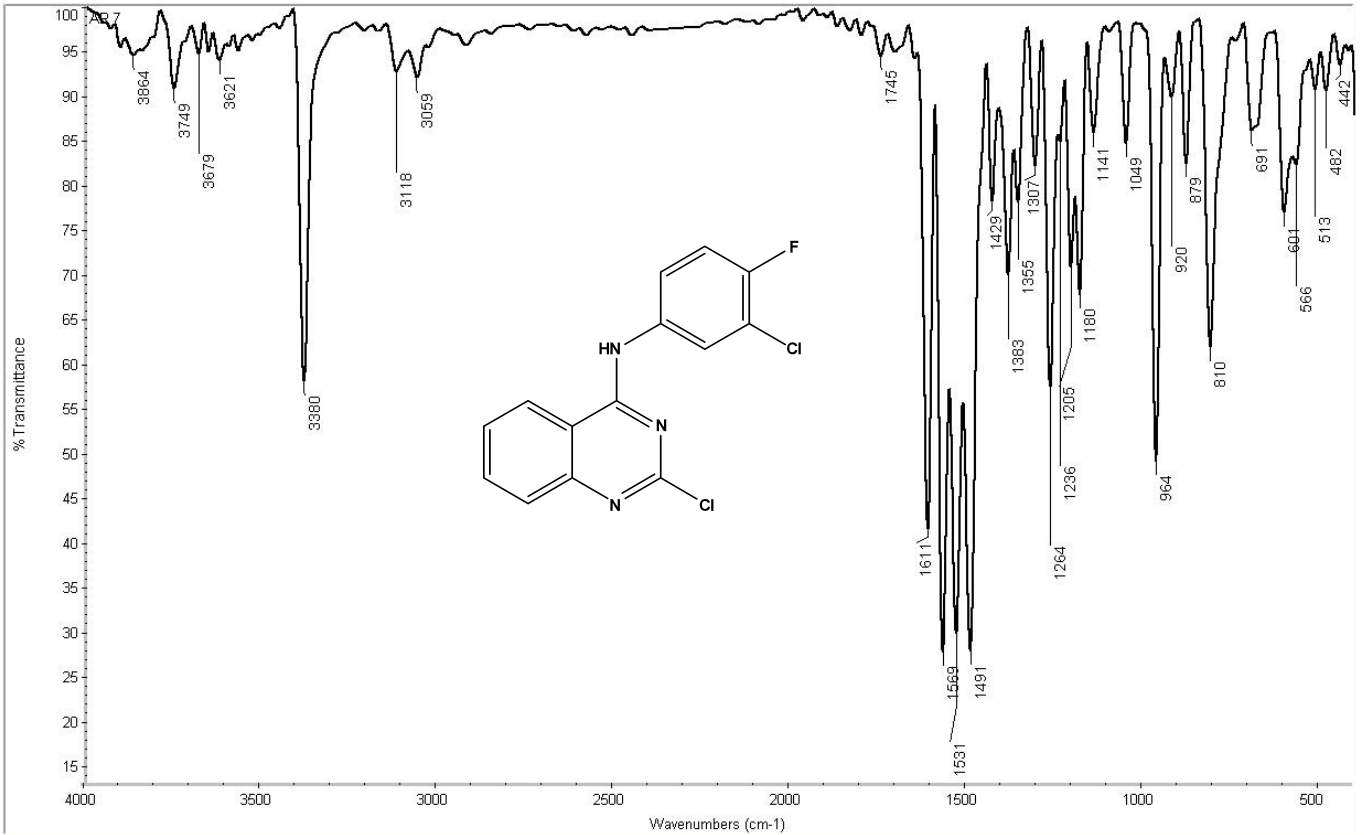




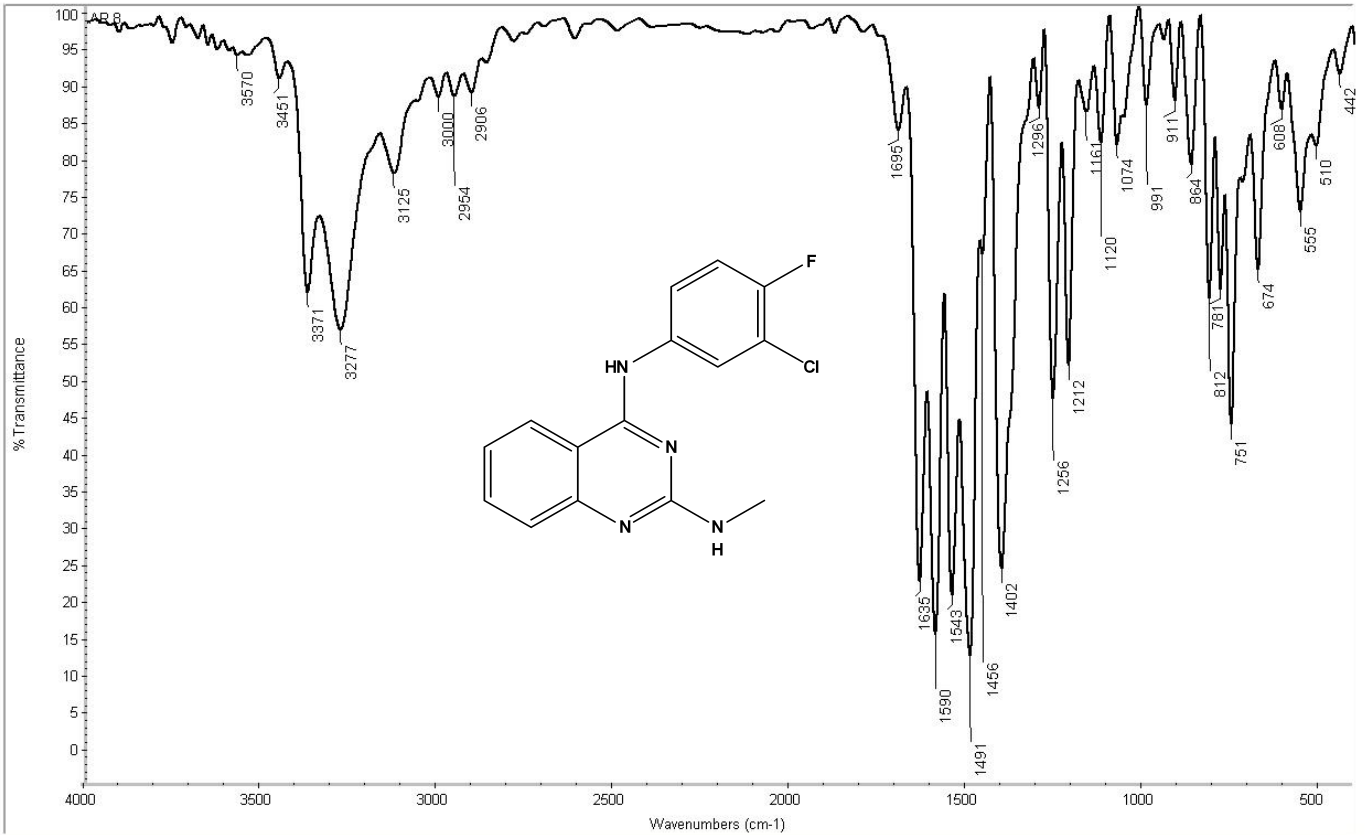




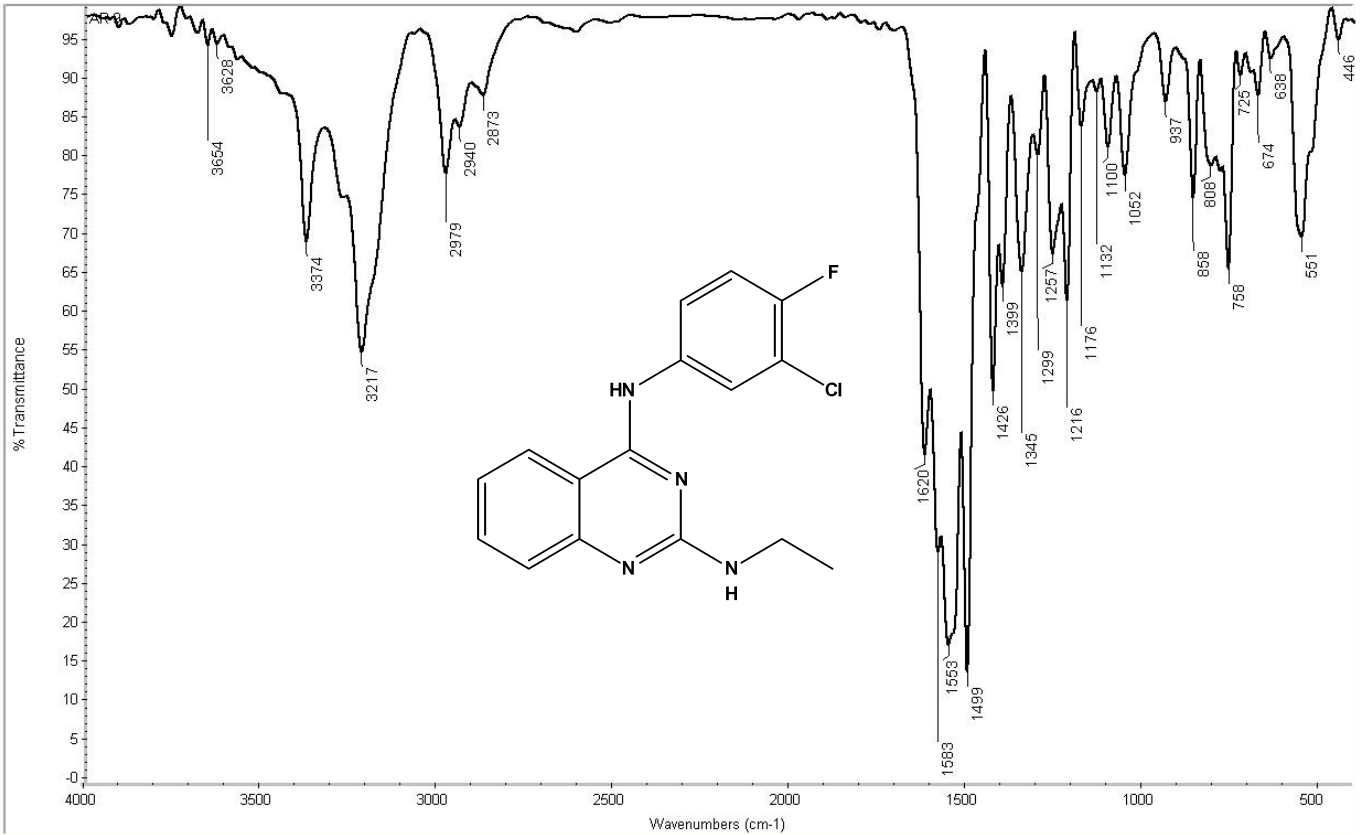
AR 7


AR 7
% Transmittance
Wavenumbers (cm-1)
3864
3749
3679
3621
3380
3118
3059
1745
1611
1569
1531
1491
1429
1383
1355
1307
1264
1236
1205
1180
1141
1049
964
920
879
810
691
601
566
513
482
442
F
Cl
HN
N
N
Cl

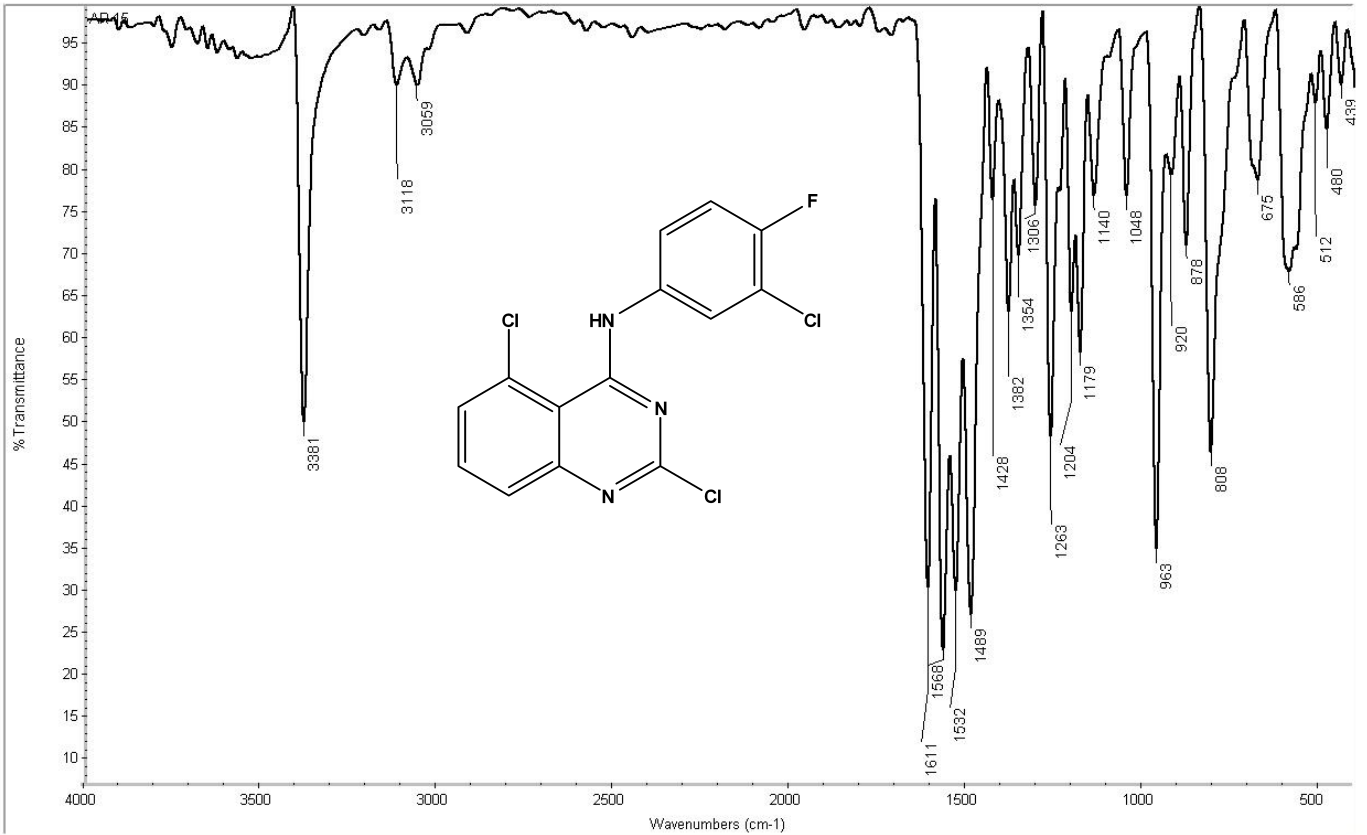
AR 8


AR 8
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
3570
3451
3371
3277
3125
3000
2954
2906
1695
1635
1590
1543
1491
1456
1402
1296
1256
1212
1161
1120
1074
991
911
864
812
781
751
674
608
555
510
442
F
Cl
HN
N
N
N
H

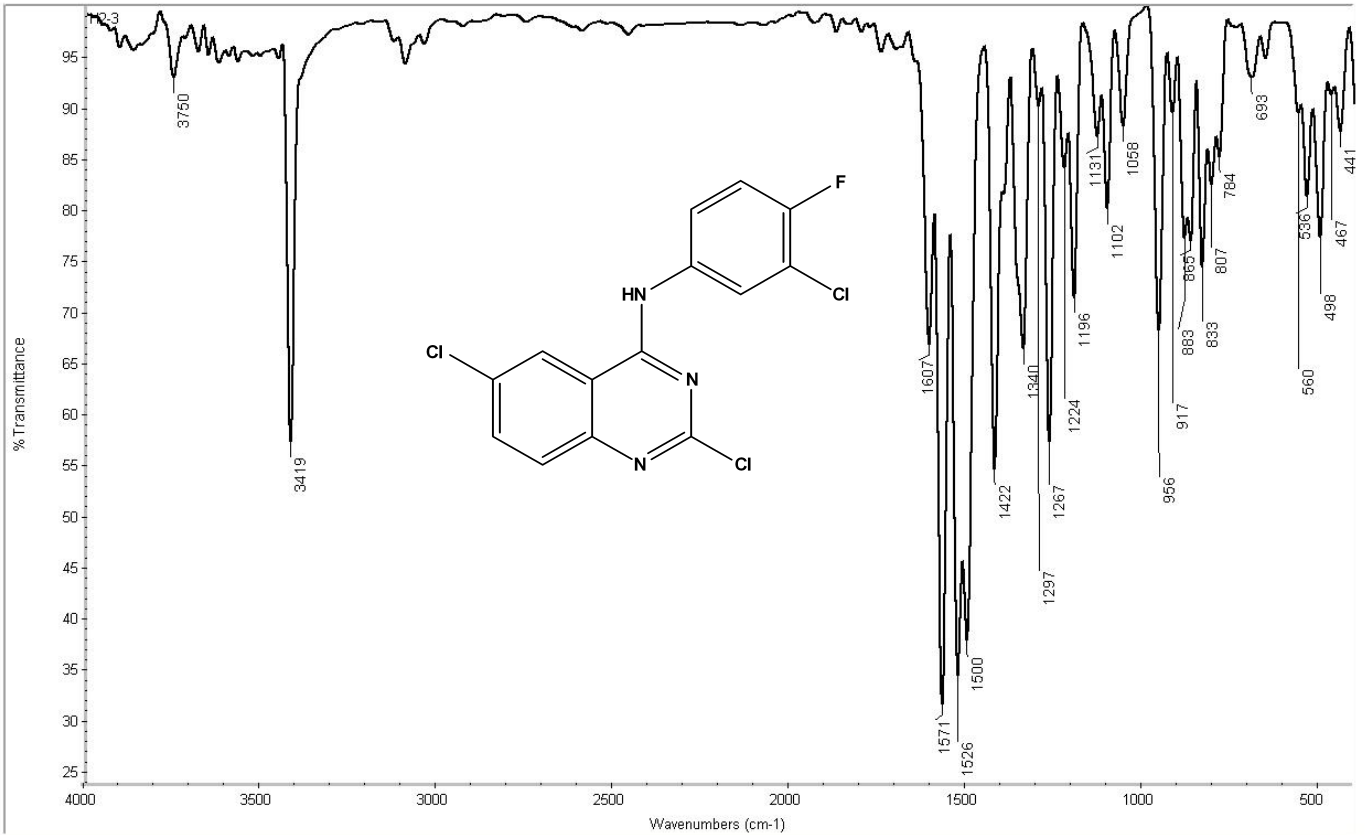
AR 9


%Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
3654
3628
3374
3217
2979
2940
2873
1620
1583
1553
1499
1426
1399
1345
1299
1257
1216
1176
1132
1100
1052
937
858
808
758
725
674
638
551
446
F
Cl
HN
N
N
N
H

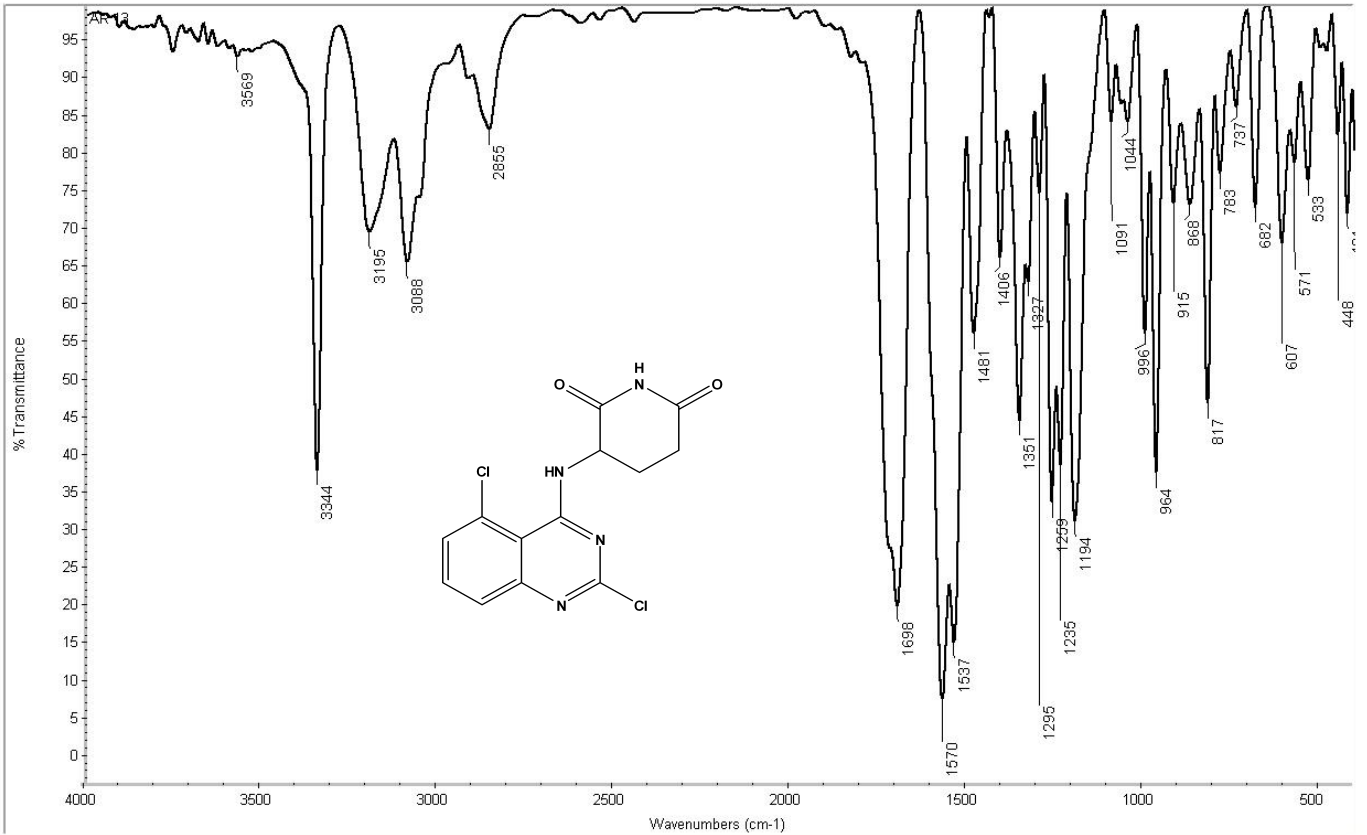
AR 15


AR 15
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
3381
3118
3059
1611
1568
1532
1489
1428
1382
1354
1306
1263
1204
1179
1140
1048
963
920
878
808
675
586
512
480
439
F
Cl
HN
Cl
N
N
Cl

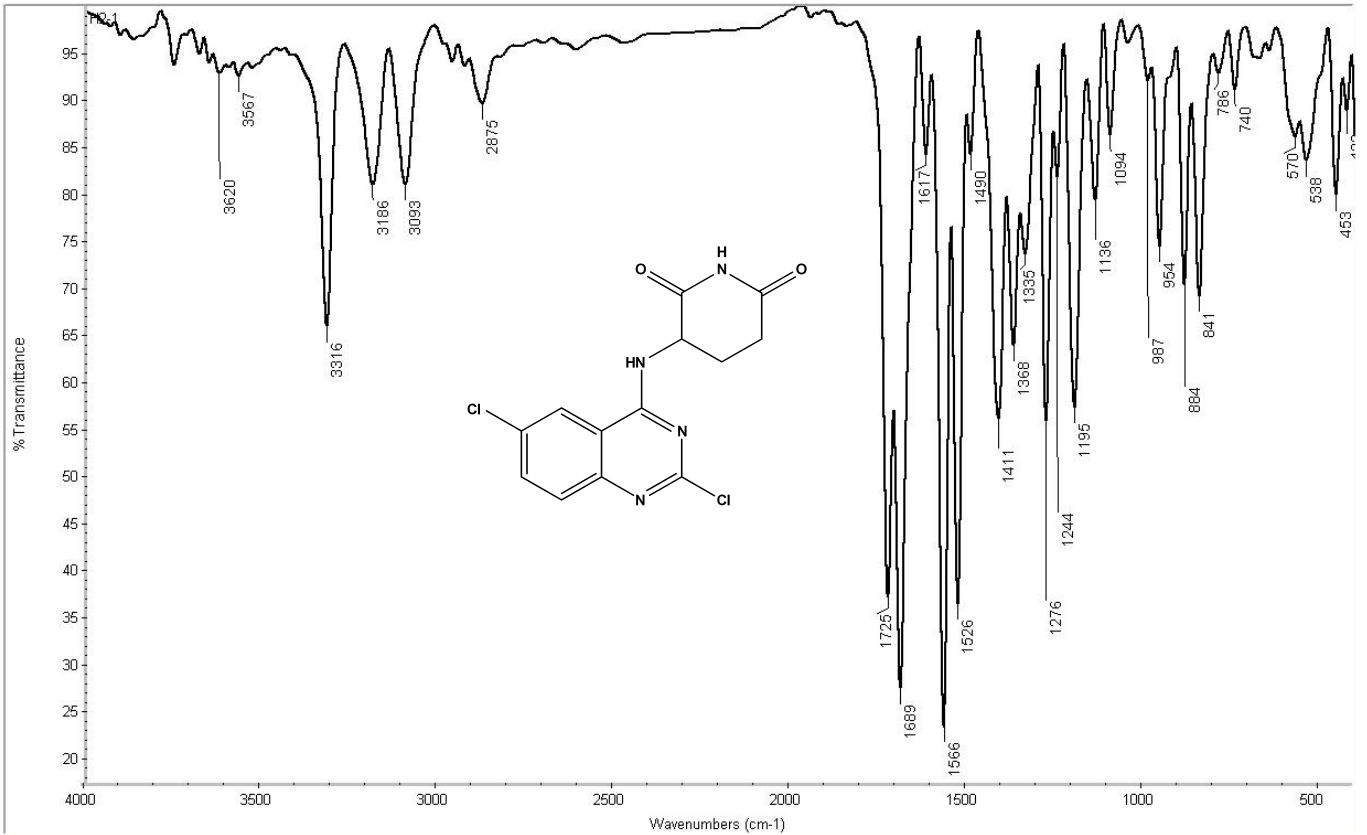
H2-3


H2-3
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
3750
3419
1607
1571
1526
1500
1422
1340
1297
1267
1224
1196
1131
1102
1058
956
917
883
865
833
807
784
693
560
536
498
467
441
F
Cl
HN
Cl
N
N
Cl

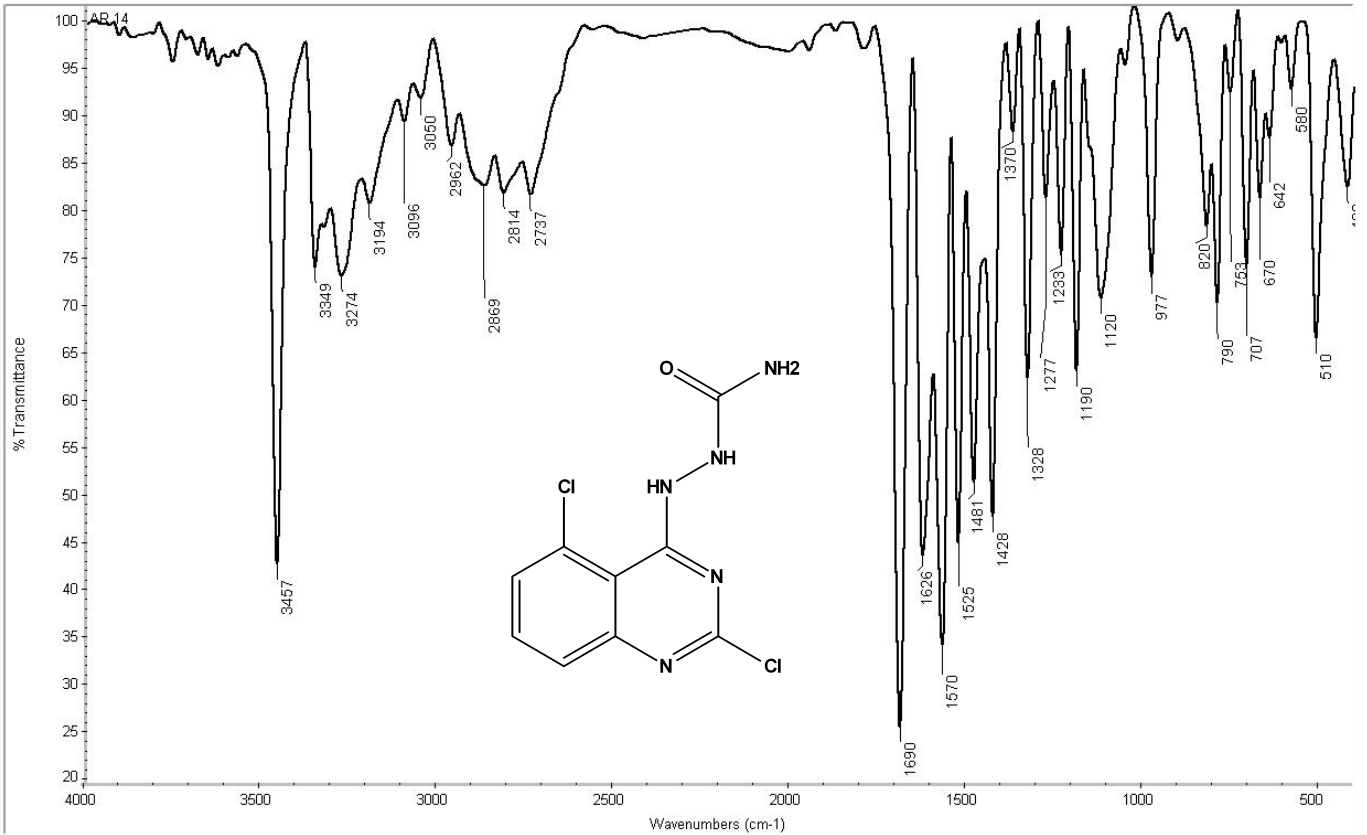
AR 13


AR-13
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
3569
3344
3195
3088
2855
1698
1570
1537
1481
1406
1351
1327
1295
1259
1235
1194
1091
1044
996
964
915
868
817
783
737
682
607
571
533
448
O
H
N
O
HN
Cl
N
N
Cl

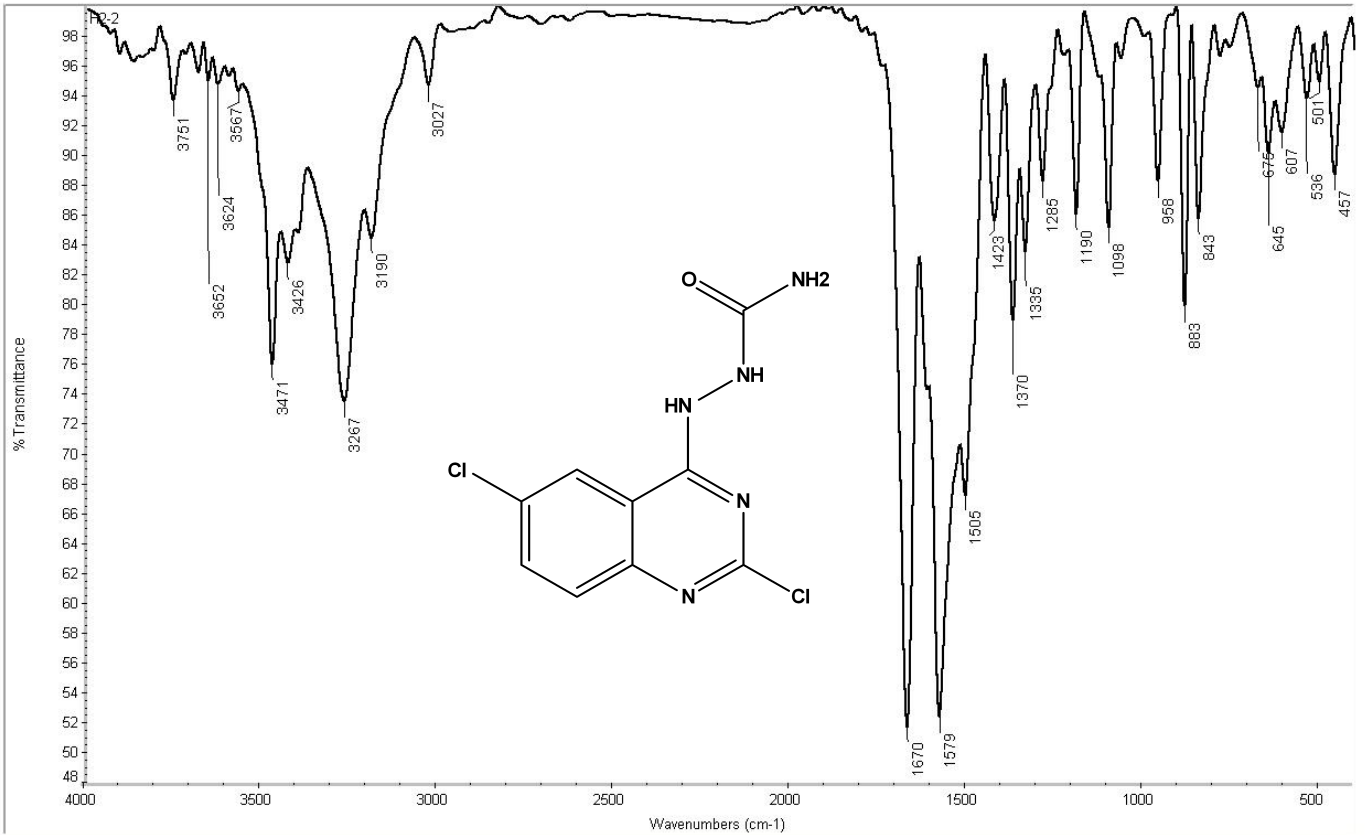
H2-1


% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
3620
3567
3316
3186
3093
2875
1725
1689
1617
1566
1526
1490
1411
1368
1335
1276
1244
1195
1136
1094
987
954
884
841
786
740
570
538
453
O
H
N
O
HN
Cl
N
N
Cl

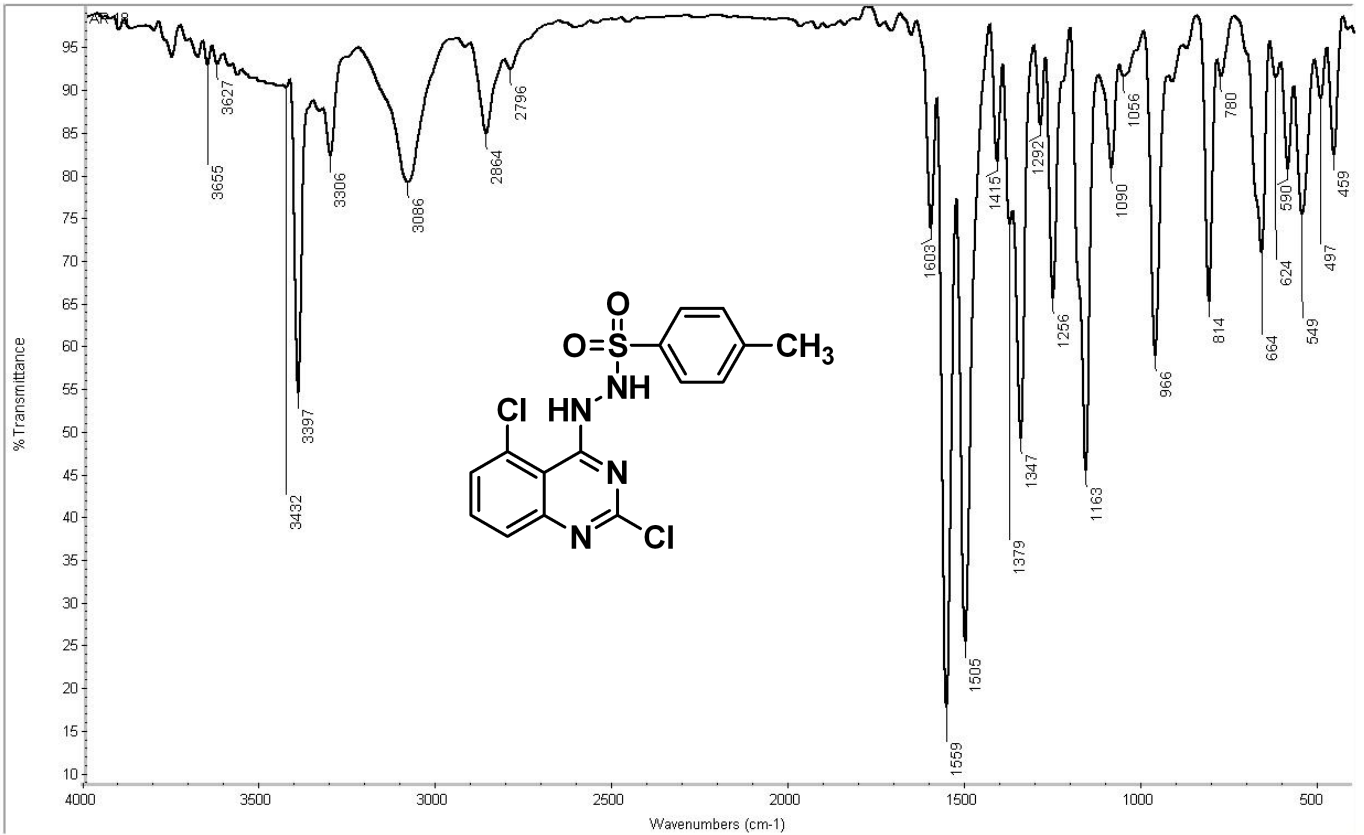
AR 14


AR 14
% Transmittance
Wavenumbers (cm-1)
3457
3349
3274
3194
3096
3050
2962
2869
2814
2737
1690
1626
1570
1525
1481
1428
1370
1328
1277
1233
1190
1120
977
820
790
753
707
670
642
580
510
O
NH2
NH
HN
Cl
N
N
Cl

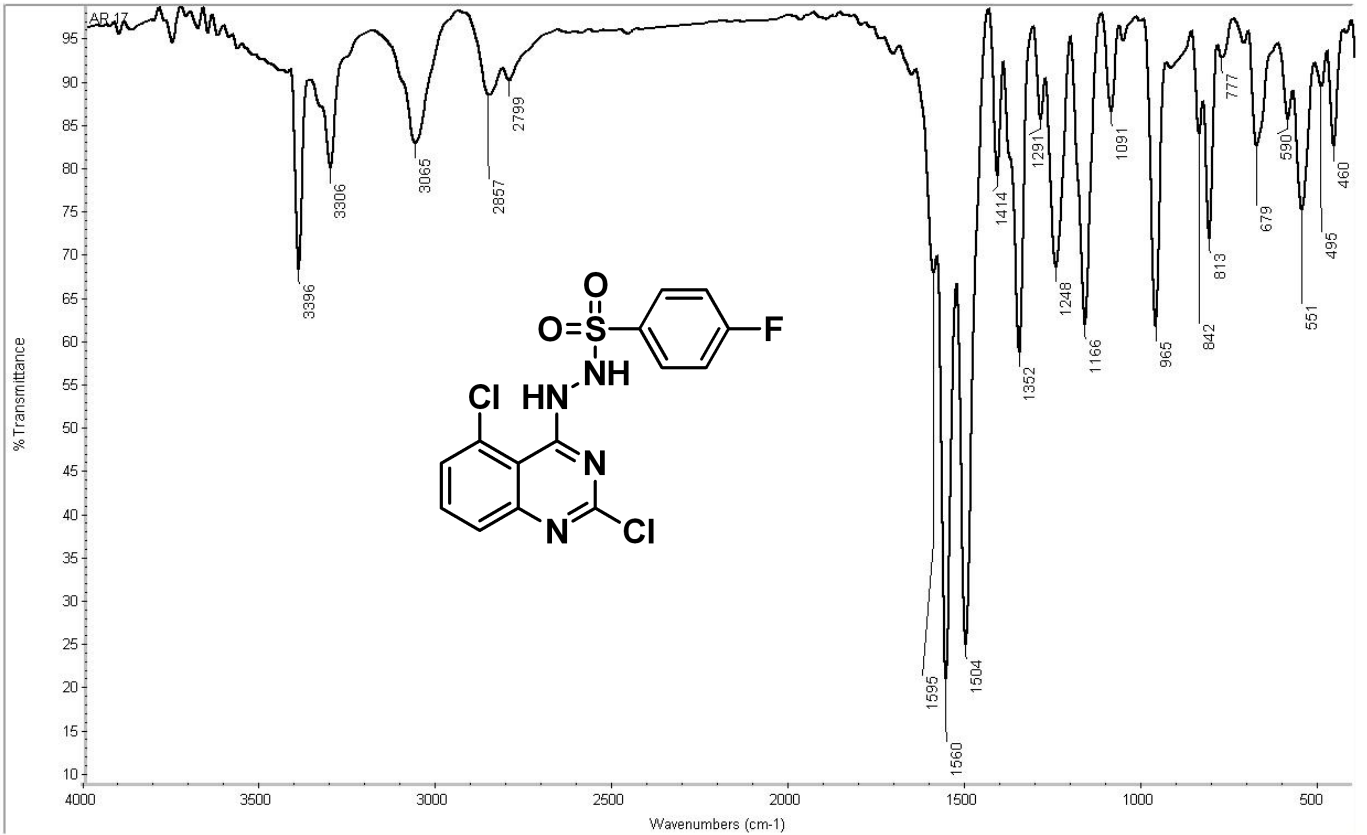
H2-2


H2-2
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
3751
3652
3624
3567
3471
3426
3267
3190
3027
1670
1579
1505
1423
1370
1335
1285
1190
1098
958
883
843
676
645
607
536
501
457
O
NH2
NH
HN
Cl
N
N
Cl

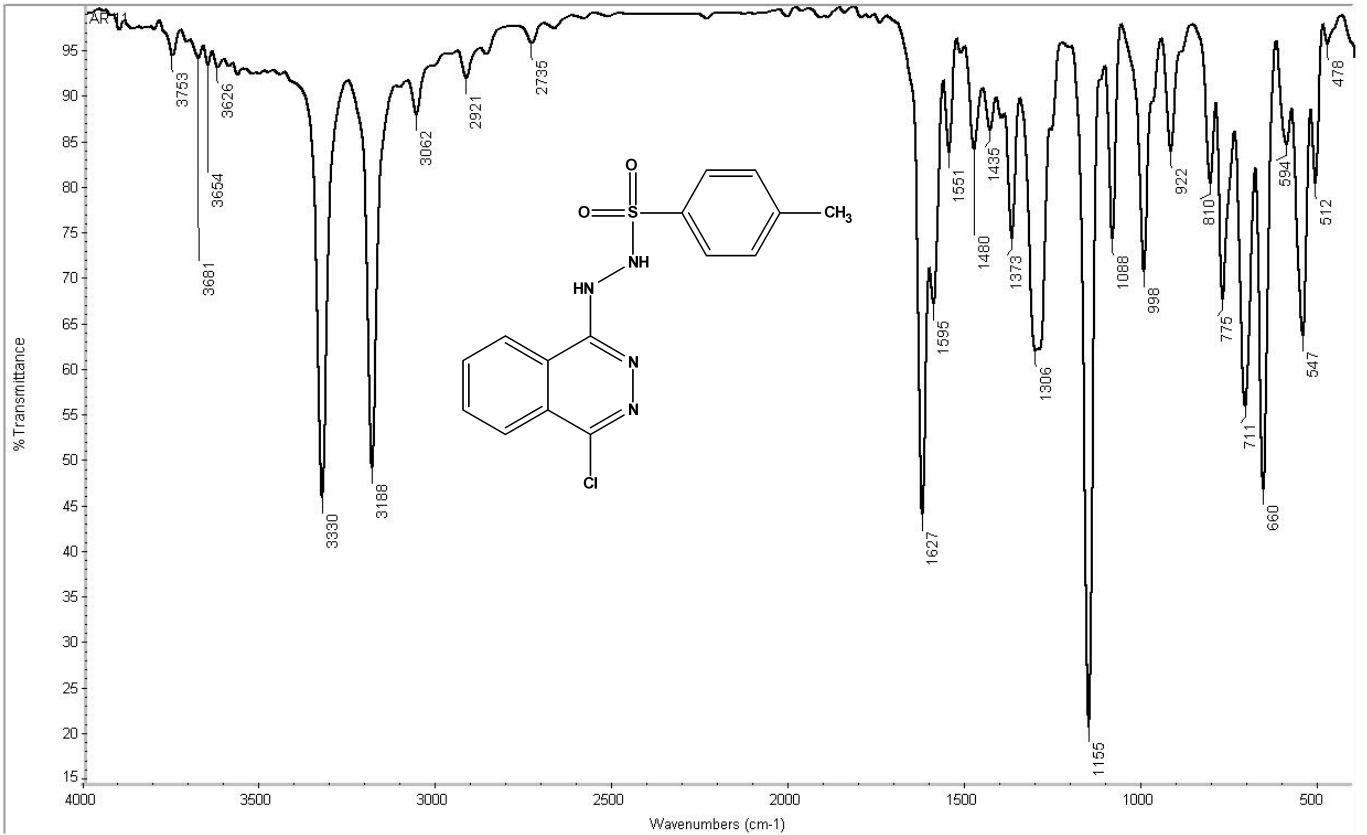
AR 18


AR-18
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
3655
3627
3432
3397
3306
3086
2864
2796
1603
1559
1505
1415
1379
1347
1292
1256
1163
1090
1056
966
814
780
664
624
590
549
497
459
O
O=S
NH
CH3
Cl
HN
N
N
Cl

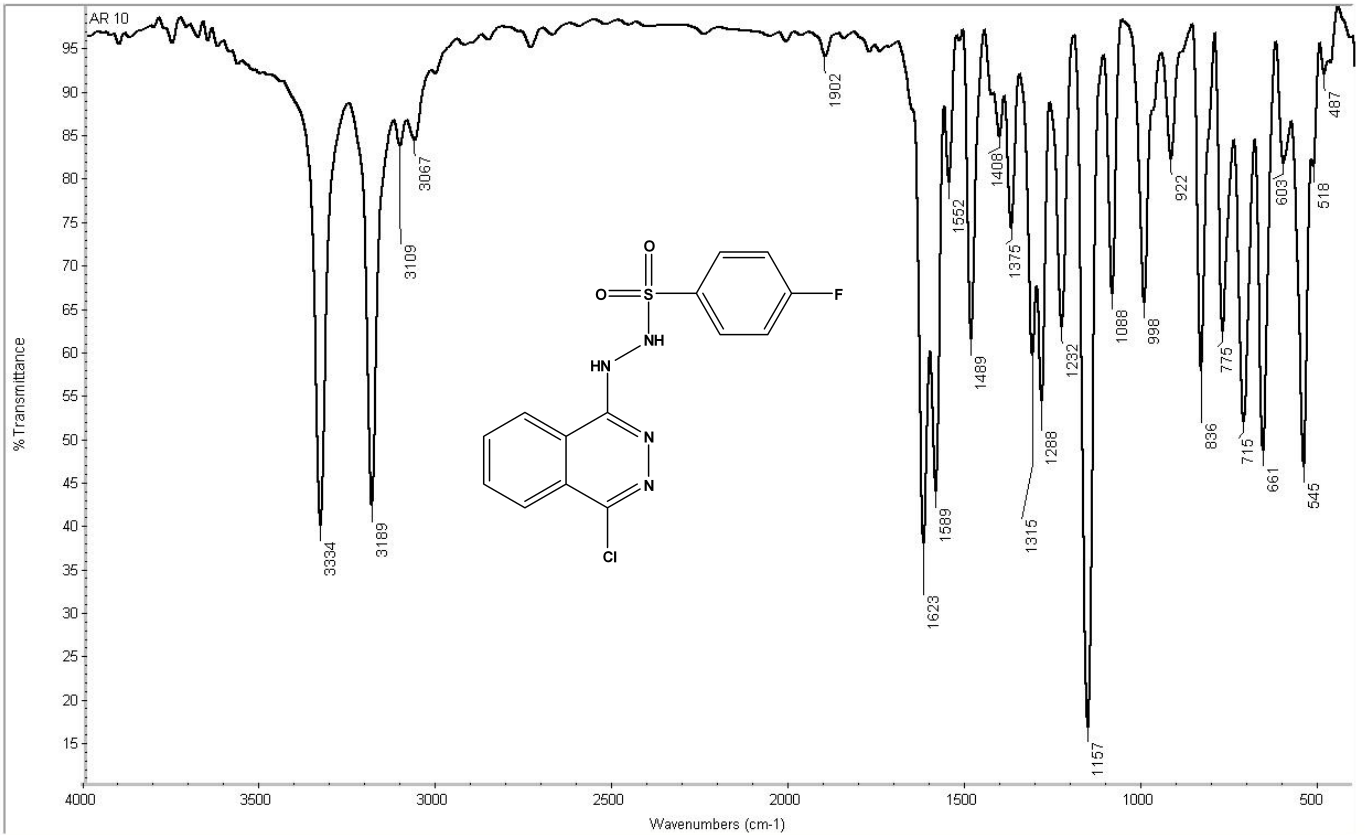
AR 17


AR 17
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
3396
3306
3065
2857
2799
1595
1560
1504
1414
1352
1291
1248
1166
1091
965
842
813
777
679
590
551
495
460
O
O=S
NH
F
Cl
HN
N
N
Cl

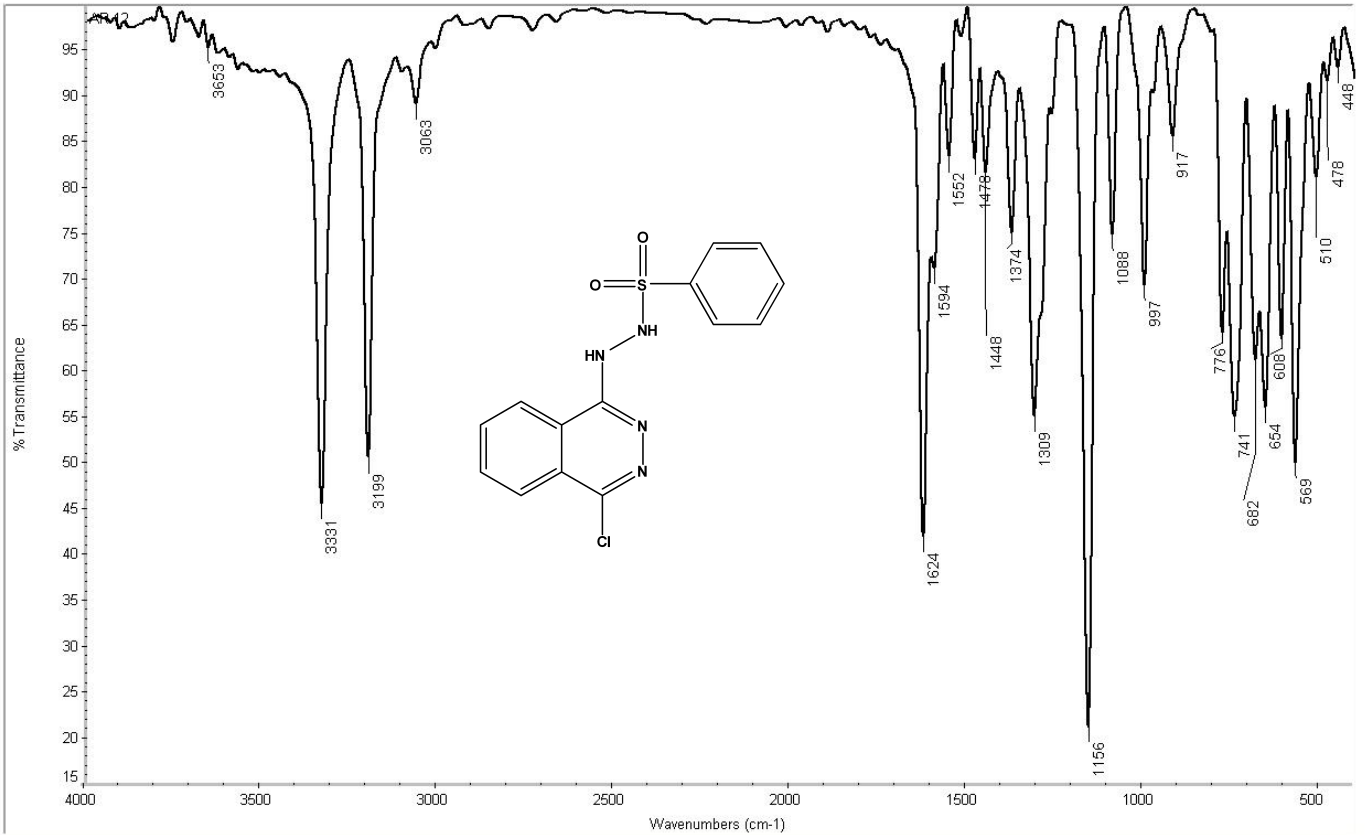
AR 11

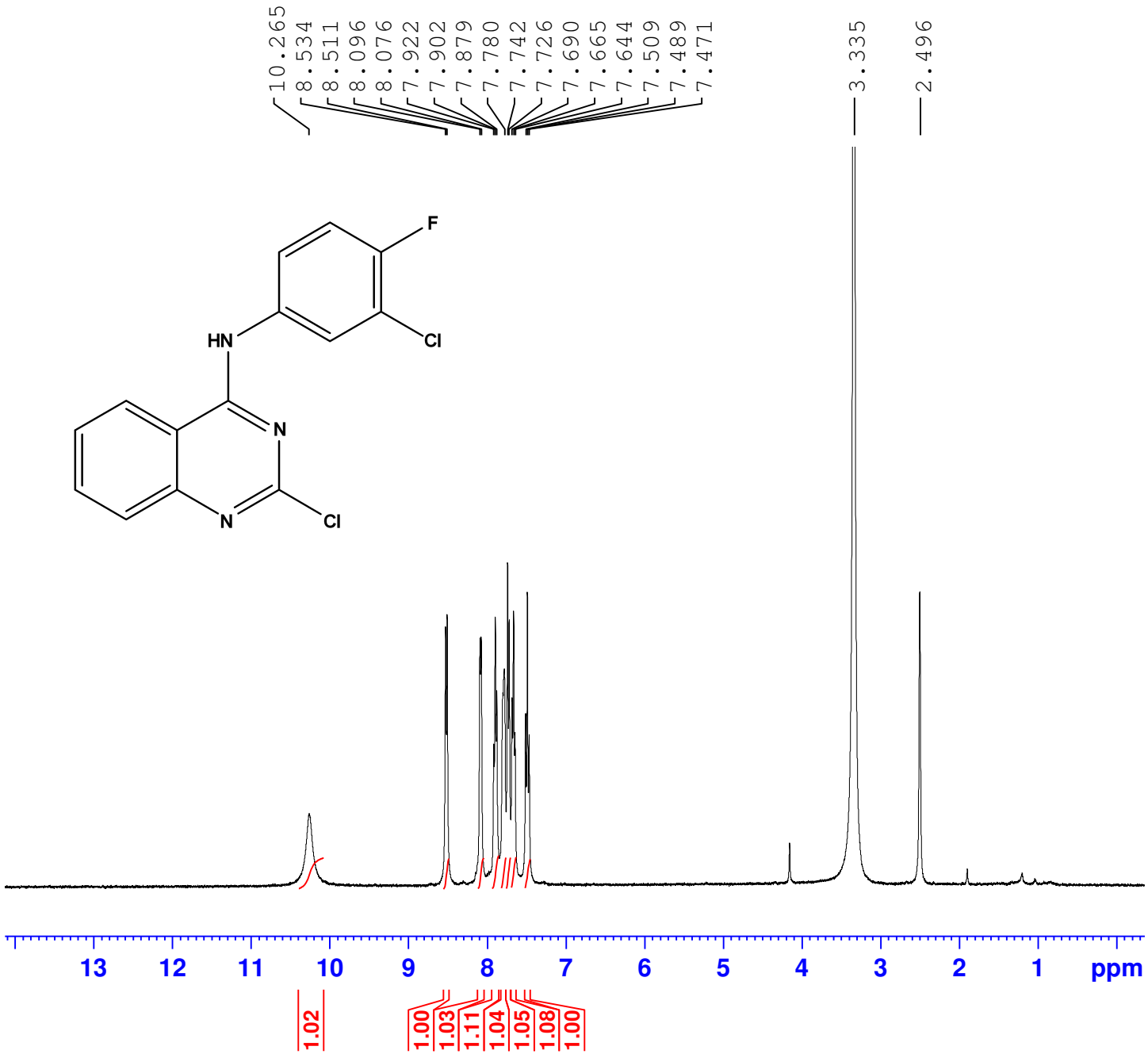

AR 11
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
3753
3681
3654
3626
3330
3188
3062
2921
2735
1627
1595
1551
1480
1435
1373
1306
1155
1088
998
922
810
775
711
660
594
547
512
478
O
S
O
CH3
NH
HN
N
N
Cl

AR 10


AR 10
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
3334
3189
3109
3067
1902
1623
1589
1552
1489
1408
1376
1315
1288
1232
1157
1088
998
922
836
775
715
661
603
545
518
487
O
S
F
NH
HN
N
N
Cl

AR 12


AR12
% Transmittance
Wavenumbers (cm-1)
4000
3500
3000
2500
2000
1500
1000
500
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
3653
3331
3199
3063
1624
1594
1552
1478
1448
1374
1309
1156
1088
997
917
776
741
682
654
608
569
510
478
448
O
O
S
NH
HN
N
N
Cl

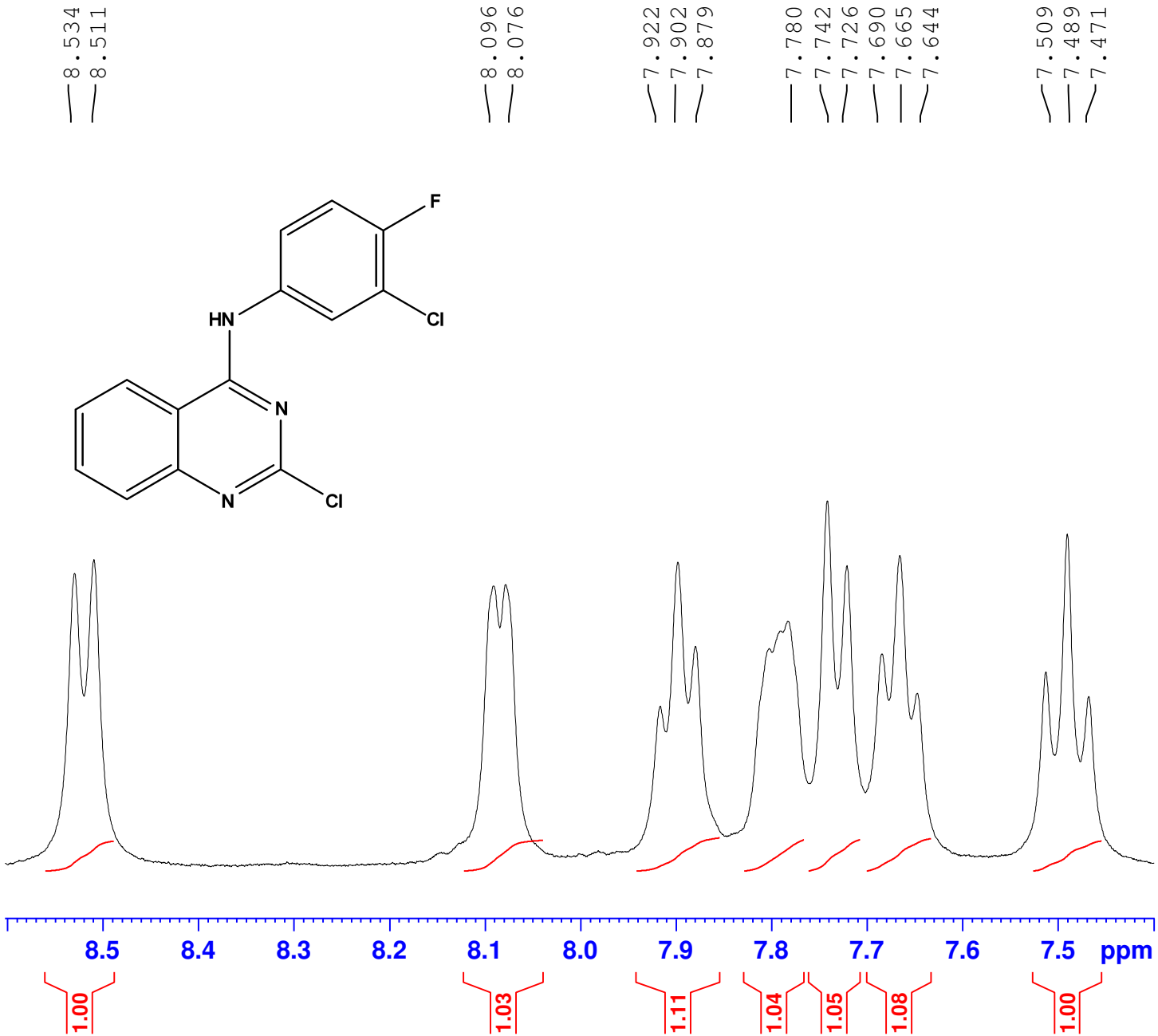


Current Data Parameters
NAME albraa-AR7
EXPNO 1
PROCNO 1

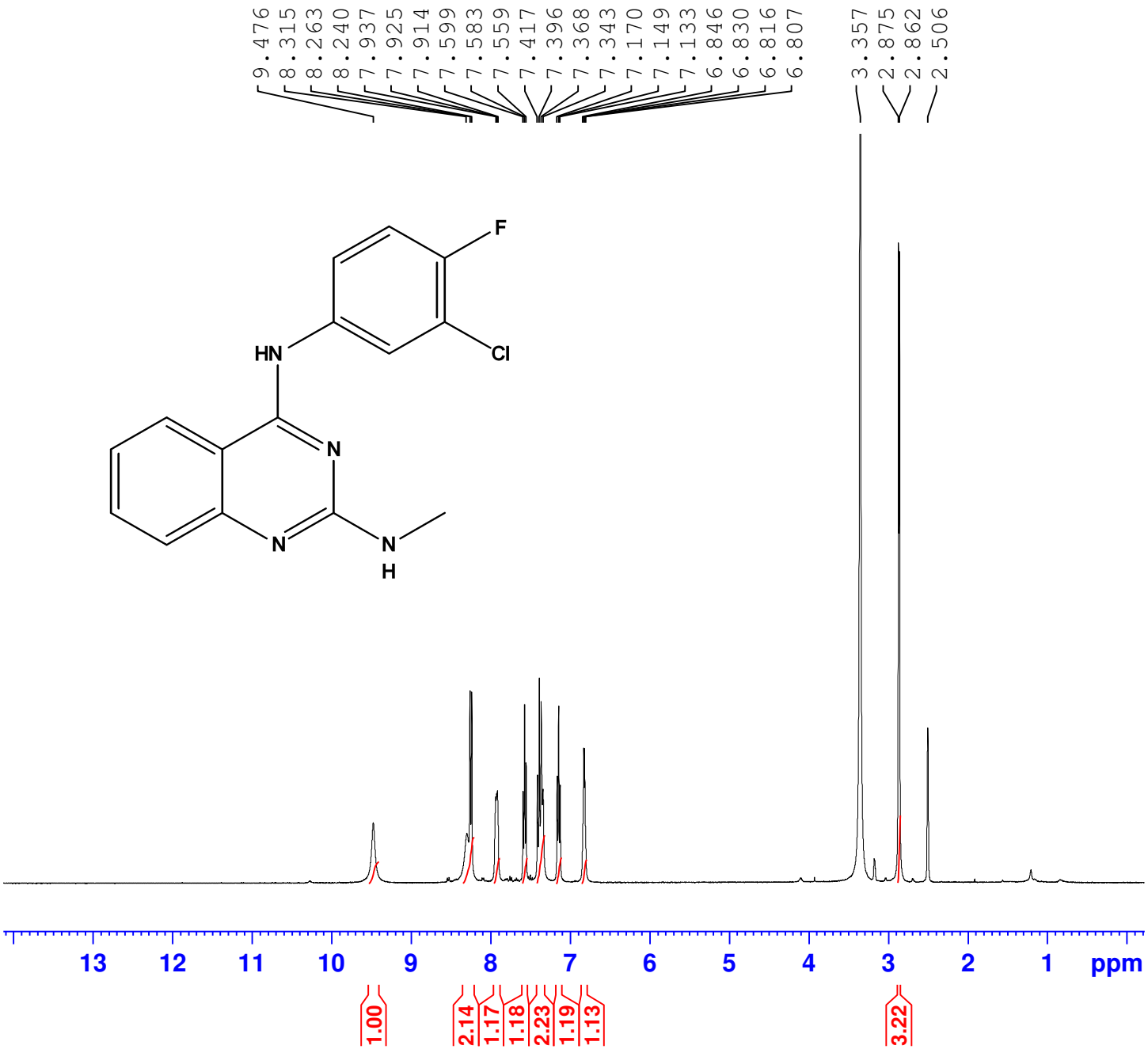
F2 - Acquisition Parameters
Date_ 20210225
Time 11.20
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 43
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00




```
Current Data Parameters
NAME          albraa-AR7
EXPNO                  1
PROCNO                 1

F2 - Acquisition Parameters
Date_           20210225
Time               11.20
INSTRUM            spect
PROBHD   5 mm PABBO BB/
PULPROG             zg30
TD                 65536
SOLVENT             DMSO
NS                    43
DS                     2
SWH             8012.820 Hz
FIDRES          0.122266 Hz
AQ             4.0894465 sec
RG                205.37
DW                62.400 usec
DE                 6.50 usec
TE                300.0 K
D1             1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
SFO1         400.1524711 MHz
NUC1                   1H
P1                 12.00 usec
PLW1         18.00000000 W

F2 - Processing parameters
SI                 65536
SF           400.1500000 MHz
WDW                   EM
SSB       0
LB                 0.30 Hz
GB        0
PC                 1.00
```



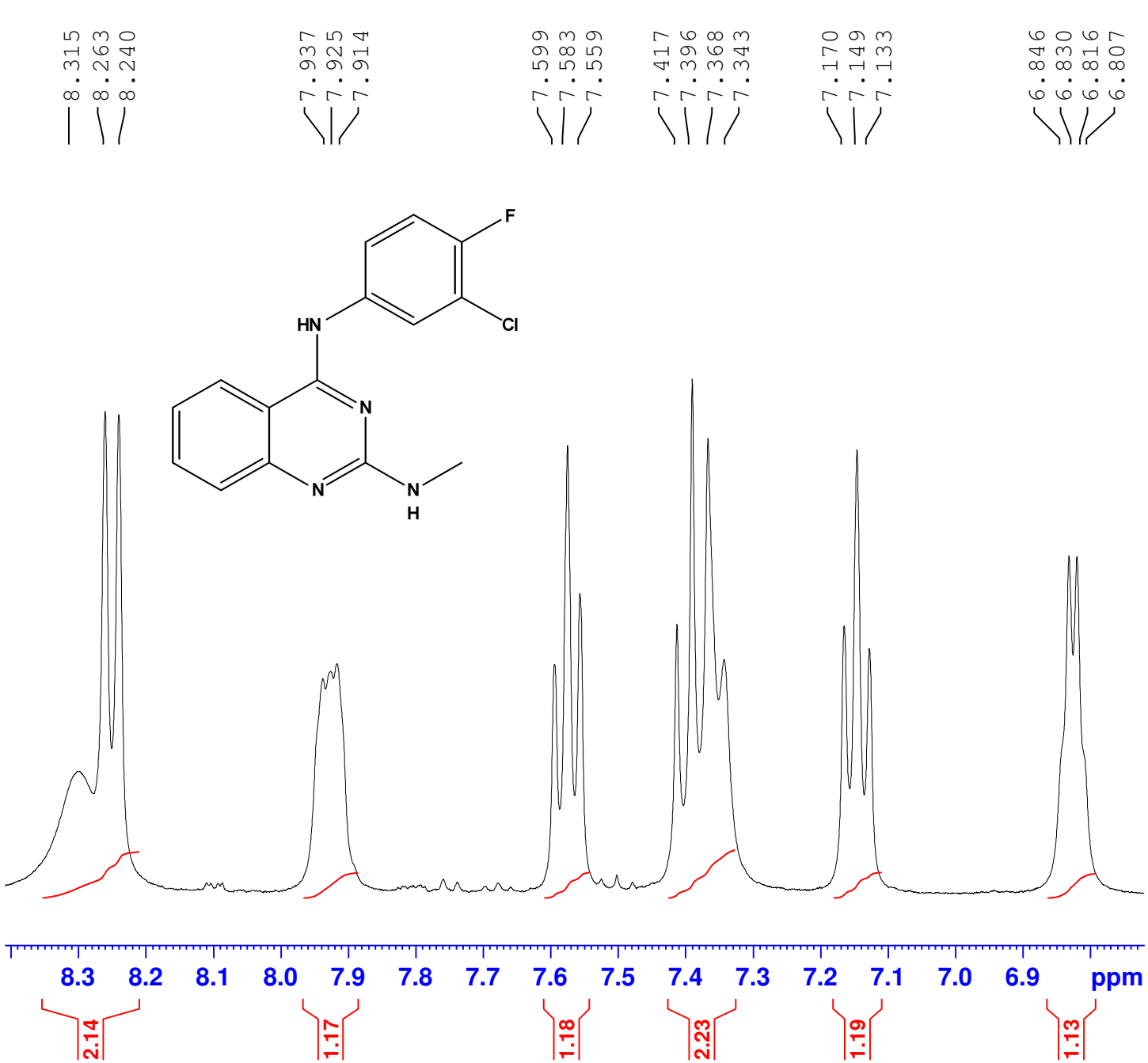
BRUKER

Current Data Parameters
NAME albraa-AR8
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210225
Time 11.27
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 55
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

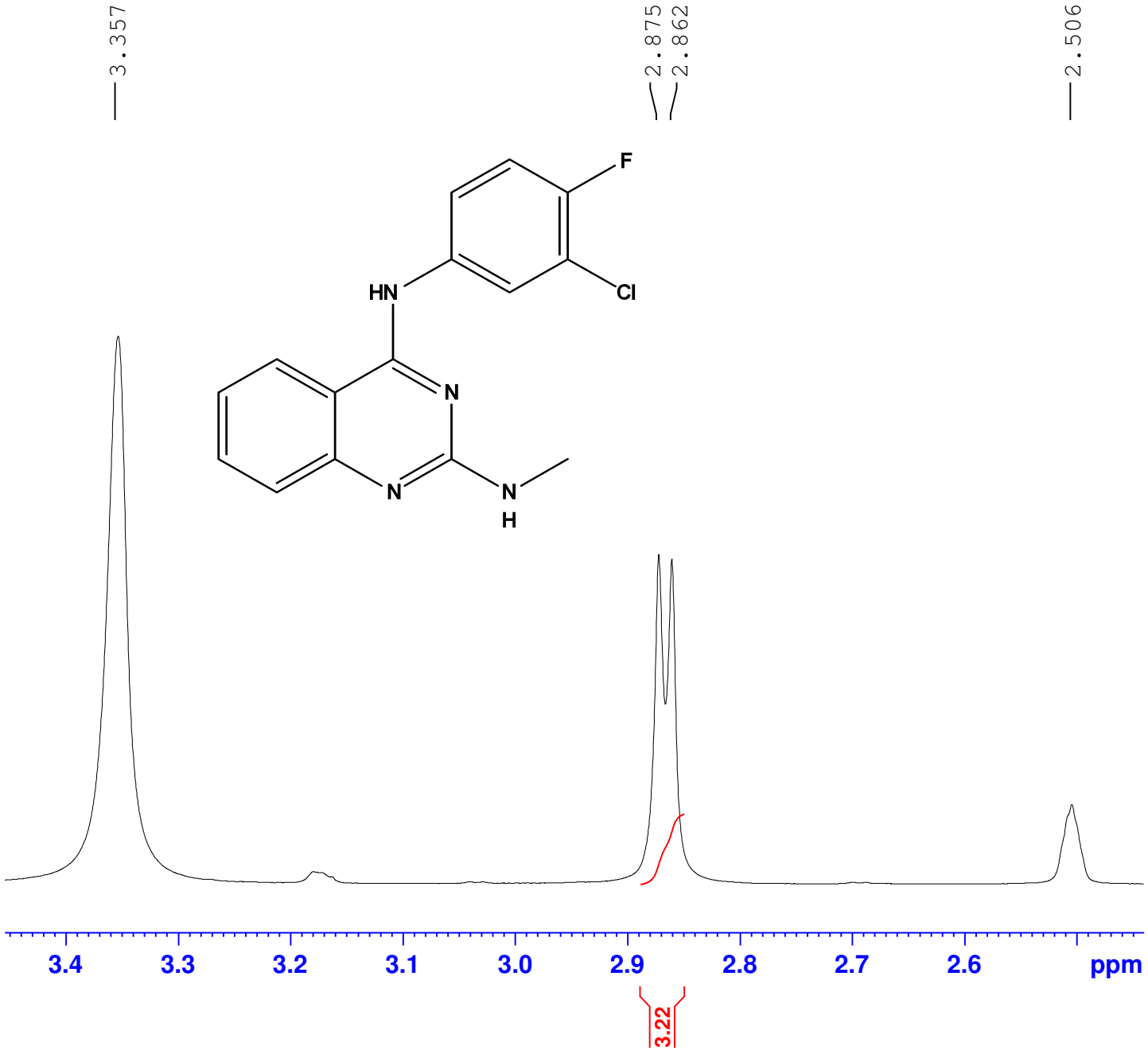


Current Data Parameters
NAME albraa-AR8
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210225
Time 11.27
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 55
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

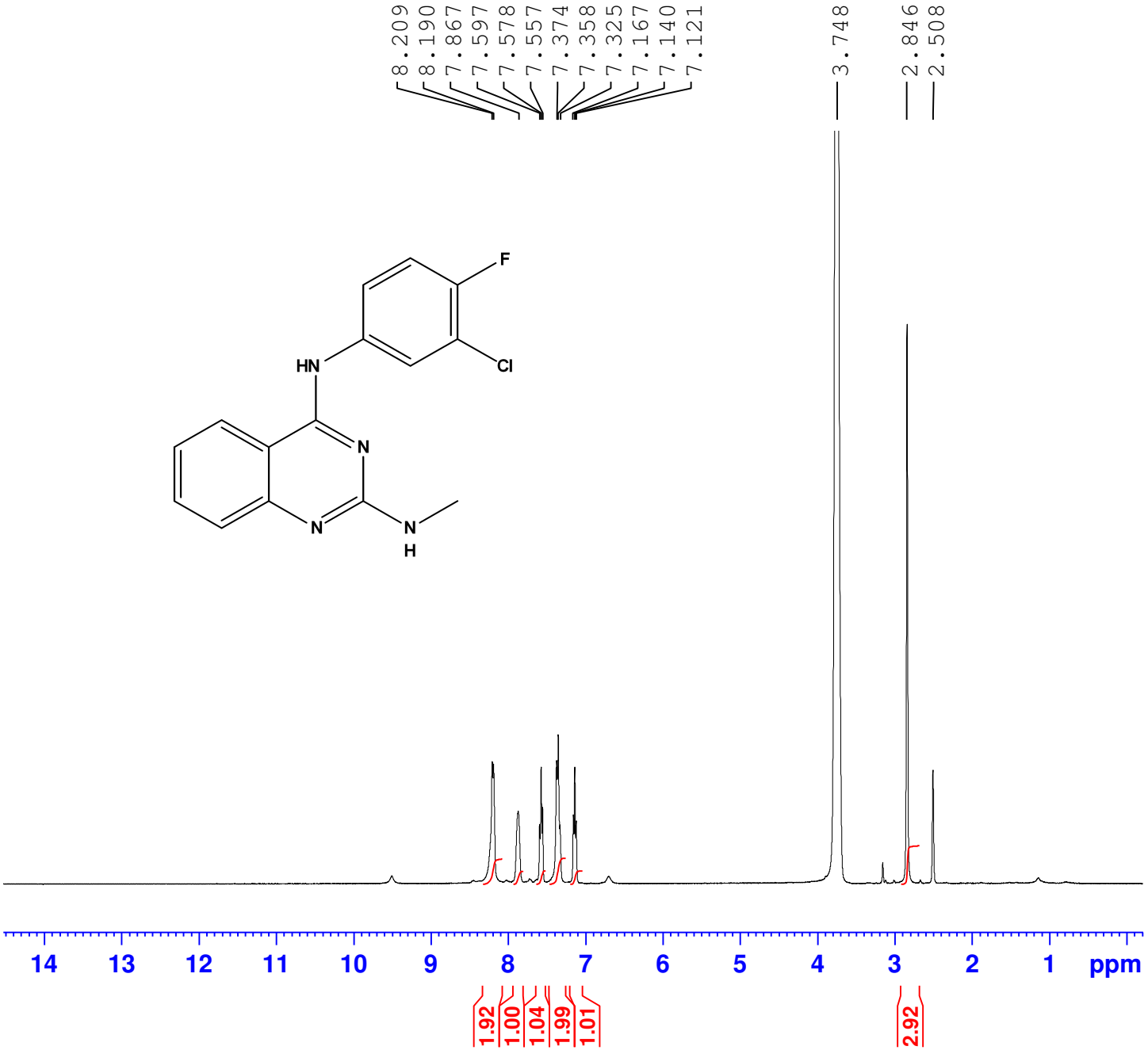
======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00



BRUKER

```
Current Data Parameters
NAME          albraa-AR8
EXPNO                  1
PROCNO                 1

F2 - Acquisition Parameters
Date_           20210225
Time               11.27
INSTRUM            spect
PROBHD   5 mm PABBO BB/
PULPROG             zg30
TD                 65536
SOLVENT             DMSO
NS                    55
DS                     2
SWH             8012.820 Hz
FIDRES          0.122266 Hz
AQ             4.0894465 sec
RG                205.37
DW                62.400 usec
DE                 6.50 usec
TE                300.0 K
D1           1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
SFO1         400.1524711 MHz
NUC1                   1H
P1                12.00 usec
PLW1        18.00000000 W

F2 - Processing parameters
SI                 65536
SF          400.1500000 MHz
WDW                   EM
SSB       0
LB                0.30 Hz
GB        0
PC                1.00
```



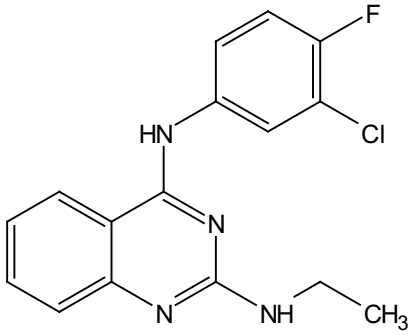
Current Data Parameters
NAME albraa-AR8-d2o
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210228
Time 11.32
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 59
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

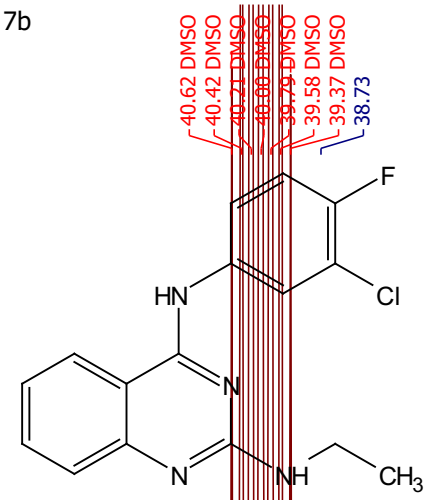
13C NMR of 7b



158.59
152.19
137.34
133.68
123.43
122.69
122.38
119.35
119.16
117.02
116.81
40.62 DMSO
40.42 DMSO
40.21 DMSO
40.00 DMSO
39.79 DMSO
39.58 DMSO
39.37 DMSO
38.73
15.08

f1 (ppm)

13C NMR of 7b



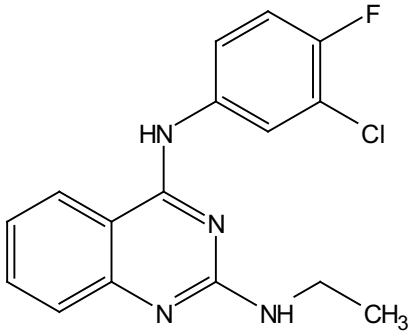
40.62 DMSO
40.42 DMSO
40.21 DMSO
40.00 DMSO
39.79 DMSO
39.58 DMSO
39.37 DMSO
38.73
15.08

3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
-200

48 46 44 42 40 38 36 34 32 30 28 26 24 22 20 18 16 14 12 10 8 6

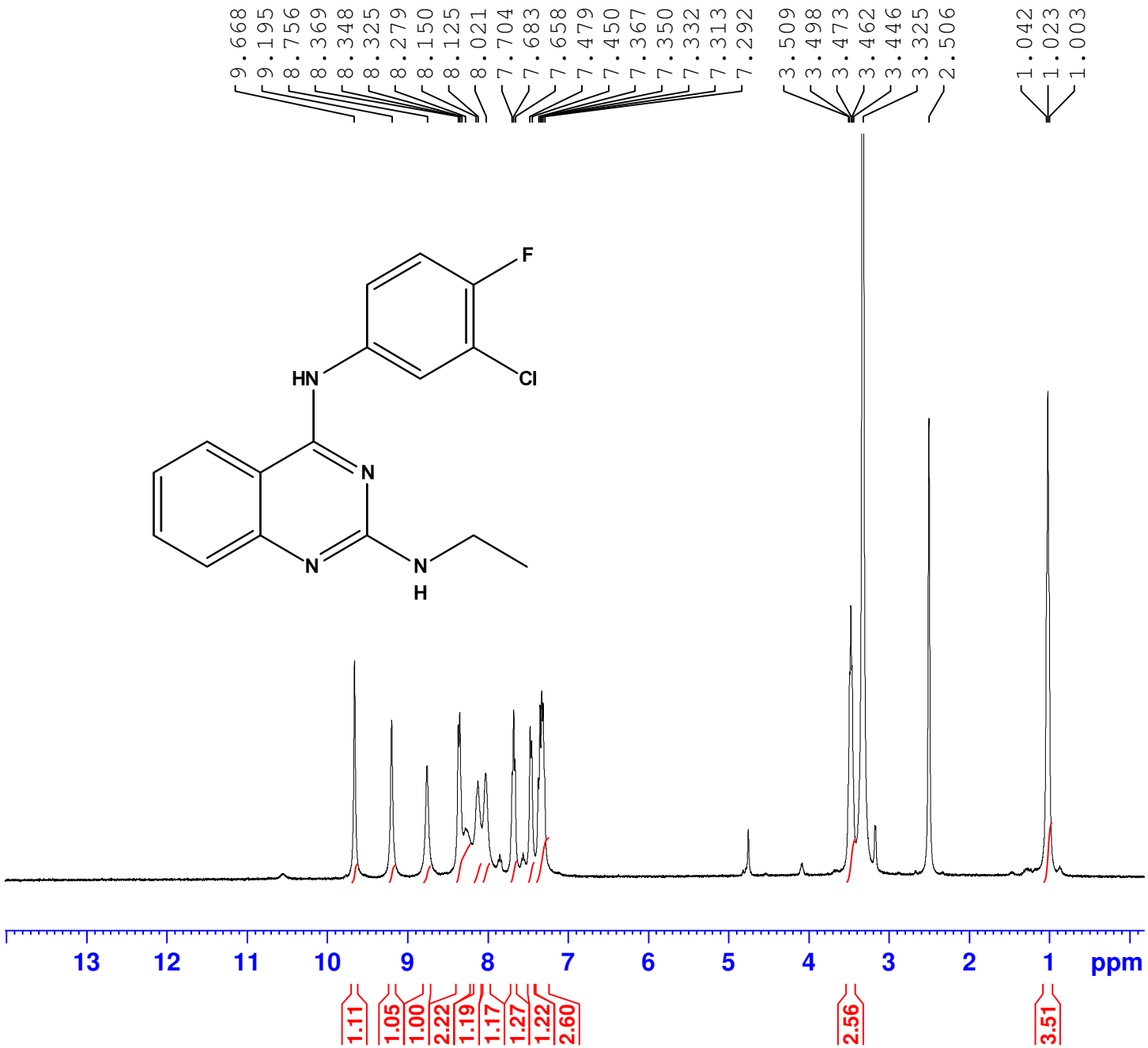
f1 (ppm)

13C NMR of 7b



158.59
152.19
137.34
133.68
123.43
122.69
122.38
119.35
119.16
117.02
116.81

f1 (ppm)



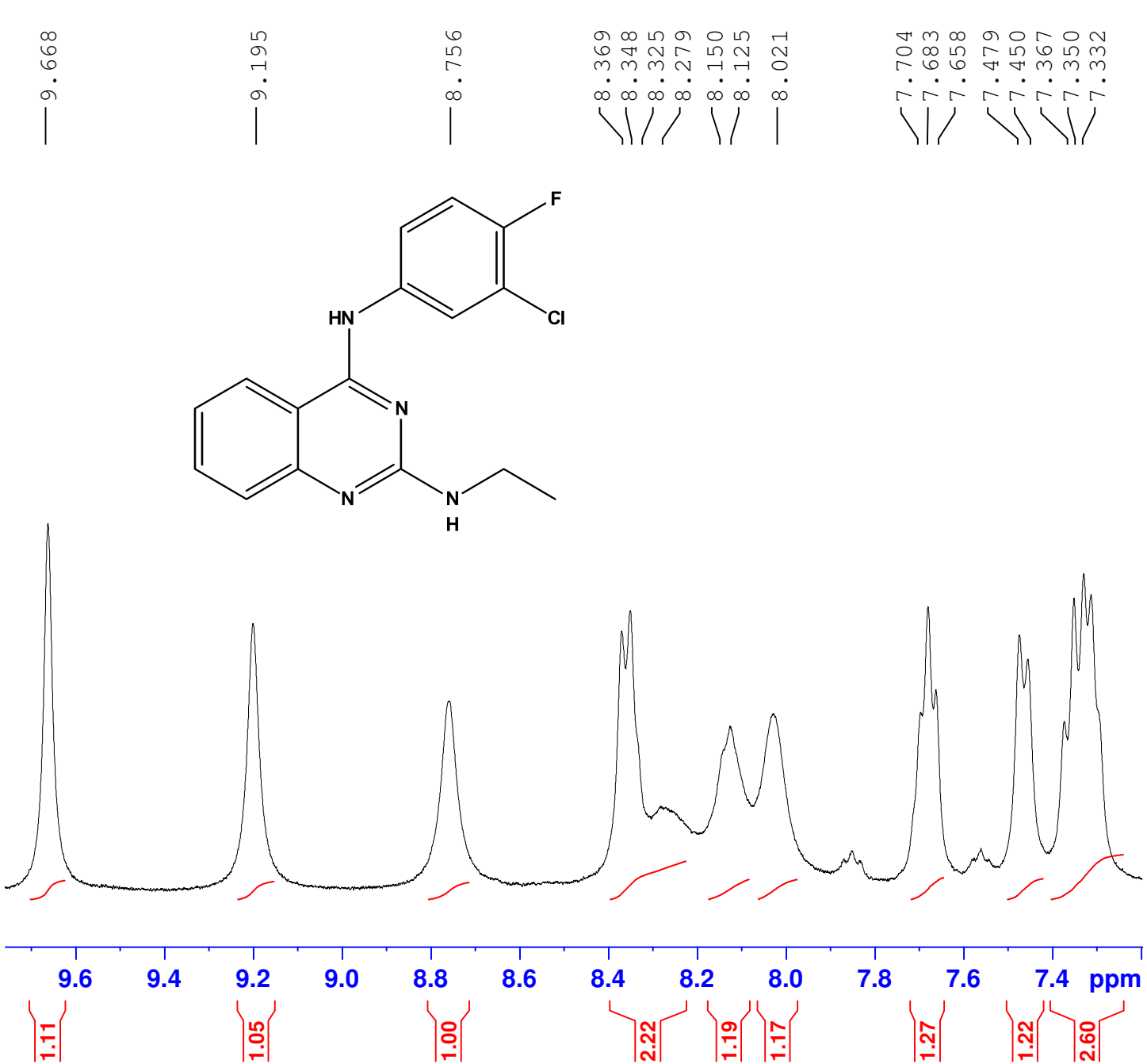
BRUKER

Current Data Parameters
NAME albraa-AR9
EXPNO 1
PROCNO 1

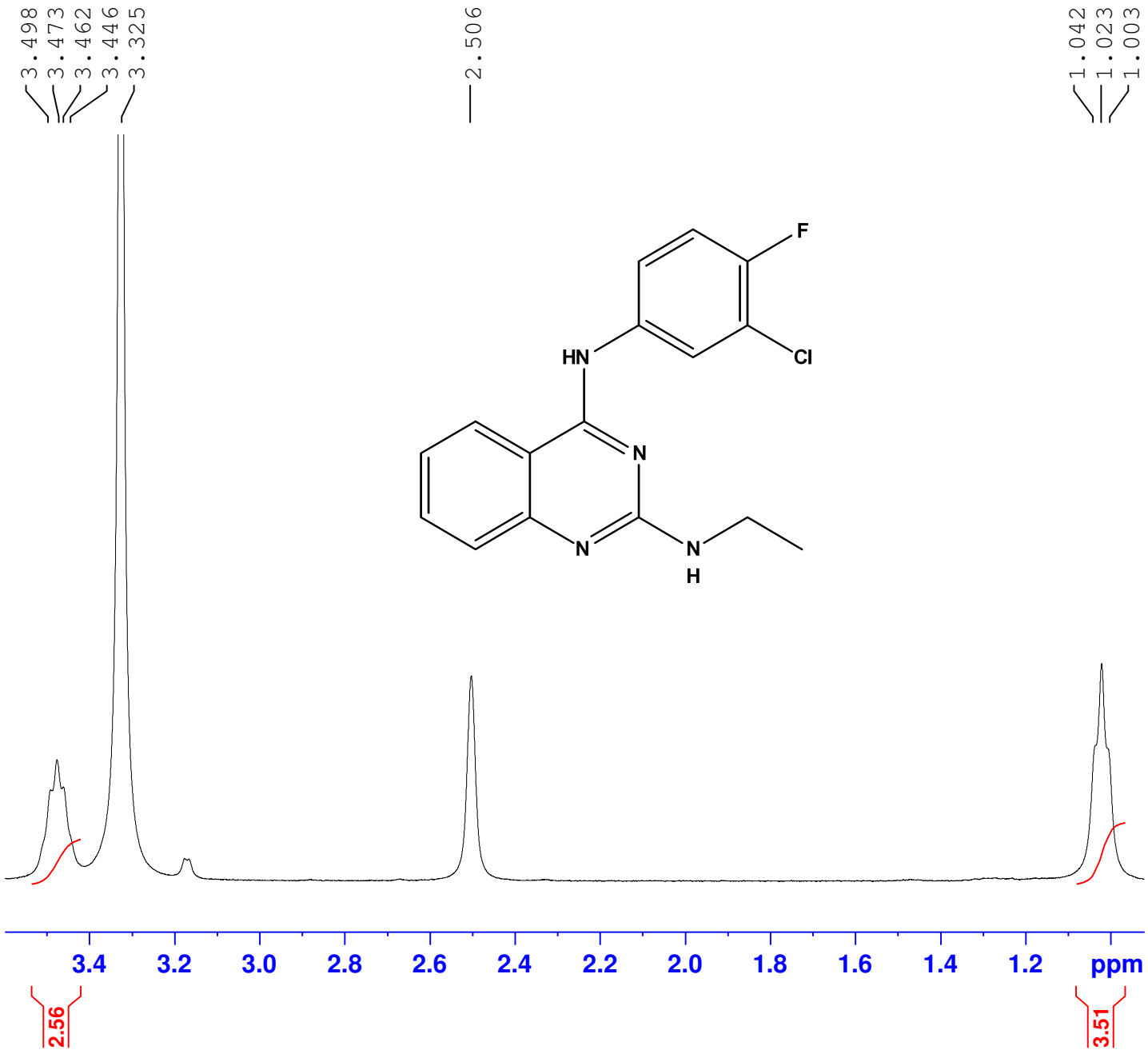
F2 - Acquisition Parameters
Date_ 20210225
Time 11.40
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 119
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00




```
Current Data Parameters
NAME          albraa-AR9
EXPNO                  1
PROCNO                 1

F2 - Acquisition Parameters
Date_           20210225
Time               11.40
INSTRUM            spect
PROBHD   5 mm PABBO BB/
PULPROG             zg30
TD                 65536
SOLVENT             DMSO
NS                   119
DS                     2
SWH             8012.820 Hz
FIDRES          0.122266 Hz
AQ             4.0894465 sec
RG                205.37
DW                62.400 usec
DE                 6.50 usec
TE                300.0 K
D1            1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
SFO1         400.1524711 MHz
NUC1                   1H
P1                 12.00 usec
PLW1         18.00000000 W

F2 - Processing parameters
SI                 65536
SF           400.1500000 MHz
WDW                   EM
SSB       0
LB                 0.30 Hz
GB        0
PC                 1.00
```

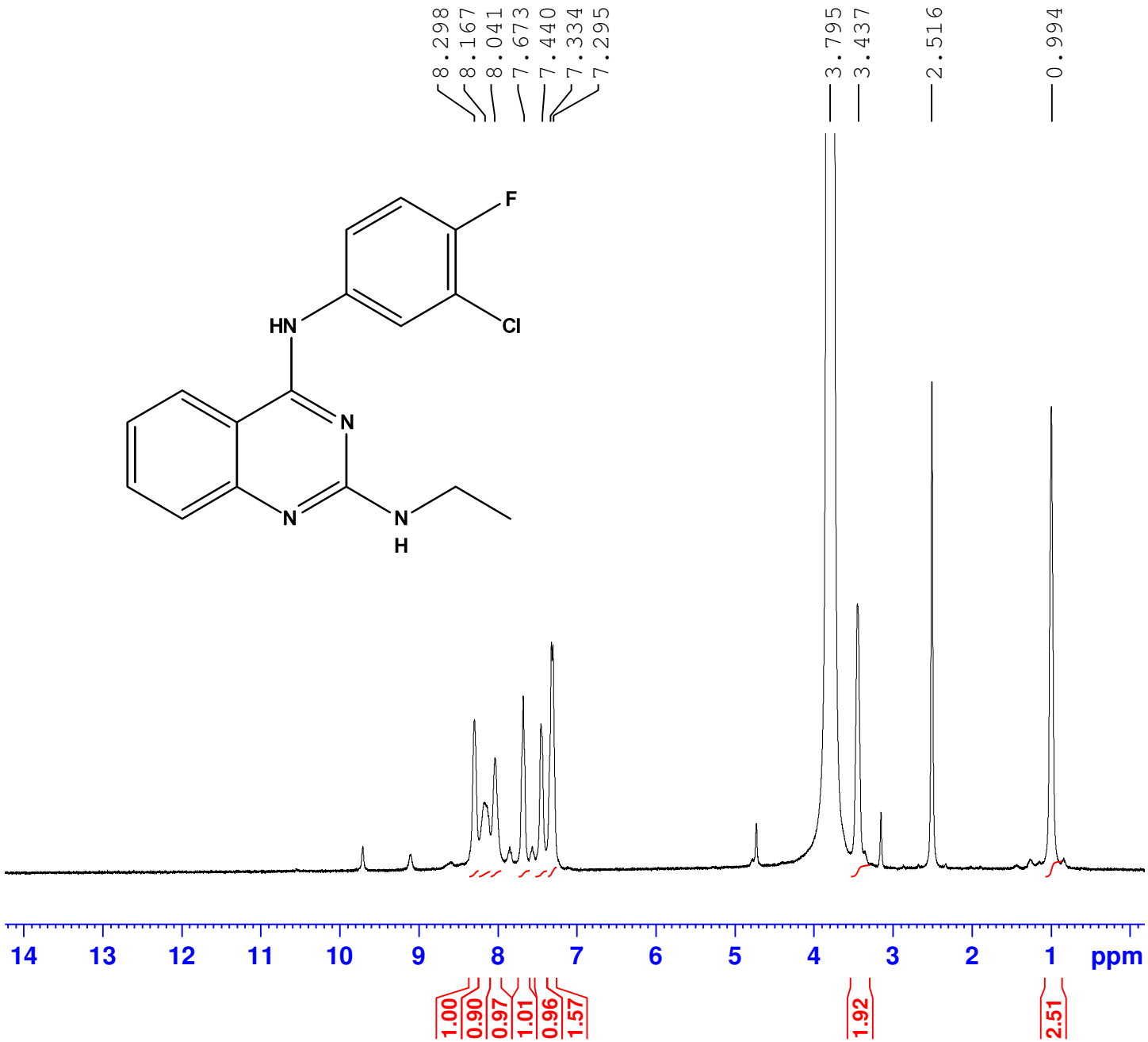


Current Data Parameters
NAME albraa-AR9
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210225
Time 11.40
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 119
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

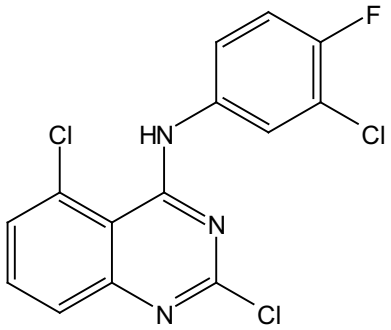
======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00




```
Current Data Parameters
NAME      albraa-AR9-d2o
EXPNO                   1
PROCNO                  1

F2 - Acquisition Parameters
Date_            20210228
Time                11.46
INSTRUM             spect
PROBHD    5 mm PABBO BB/
PULPROG              zg30
TD                  65536
SOLVENT              DMSO
NS                     128
DS                       2
SWH              8012.820 Hz
FIDRES           0.122266 Hz
AQ               4.0894465 sec
RG                  205.37
DW                  62.400 usec
DE                   6.50 usec
TE                  300.0 K
D1             1.00000000 sec
TD0                     1

======== CHANNEL f1 ========
SFO1          400.1524711 MHz
NUC1                    1H
P1                  12.00 usec
PLW1          18.00000000 W

F2 - Processing parameters
SI                    65536
SF            400.1500000 MHz
WDW                     EM
SSB       0
LB                   0.30 Hz
GB        0
PC                   1.00
```

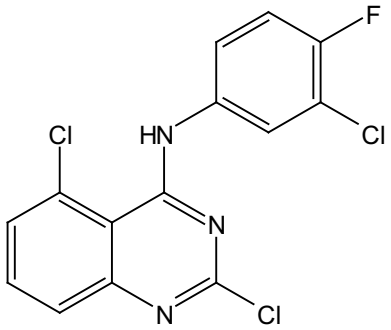
Hazem ElQady_C_AR-15.10.fid
Hazem ElQady_C_AR-15

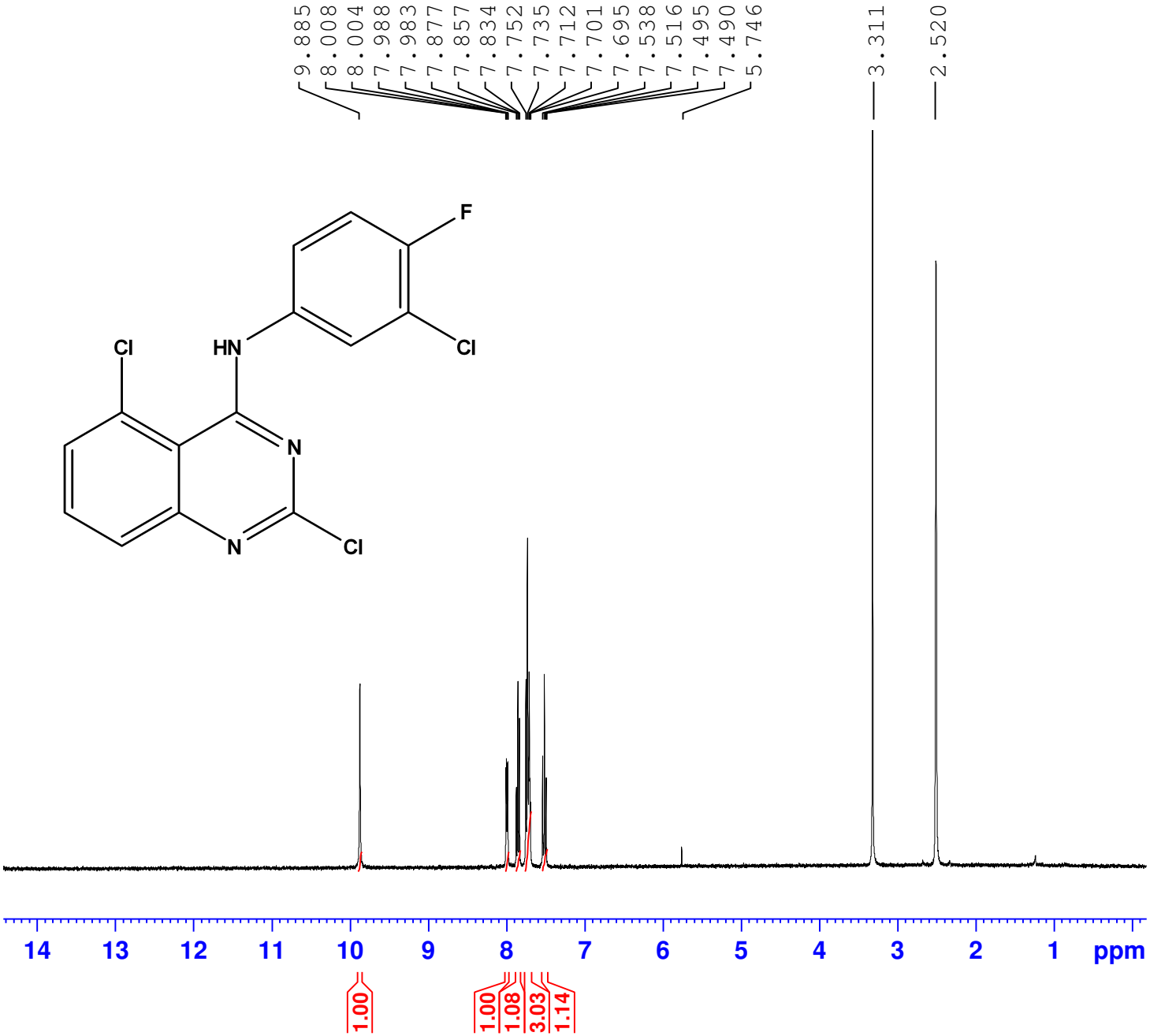


158.73
156.19
153.85
134.56
129.74
129.08
127.23
125.62
124.50
124.43
117.35
117.13
111.97
40.63 DMSO
40.42 DMSO
40.21 DMSO
40.00 DMSO
39.79 DMSO
39.59 DMSO
39.38 DMSO

f1 (ppm)

Hazem ElQady_C_AR-15.10.fid
Hazem ElQady_C_AR-15



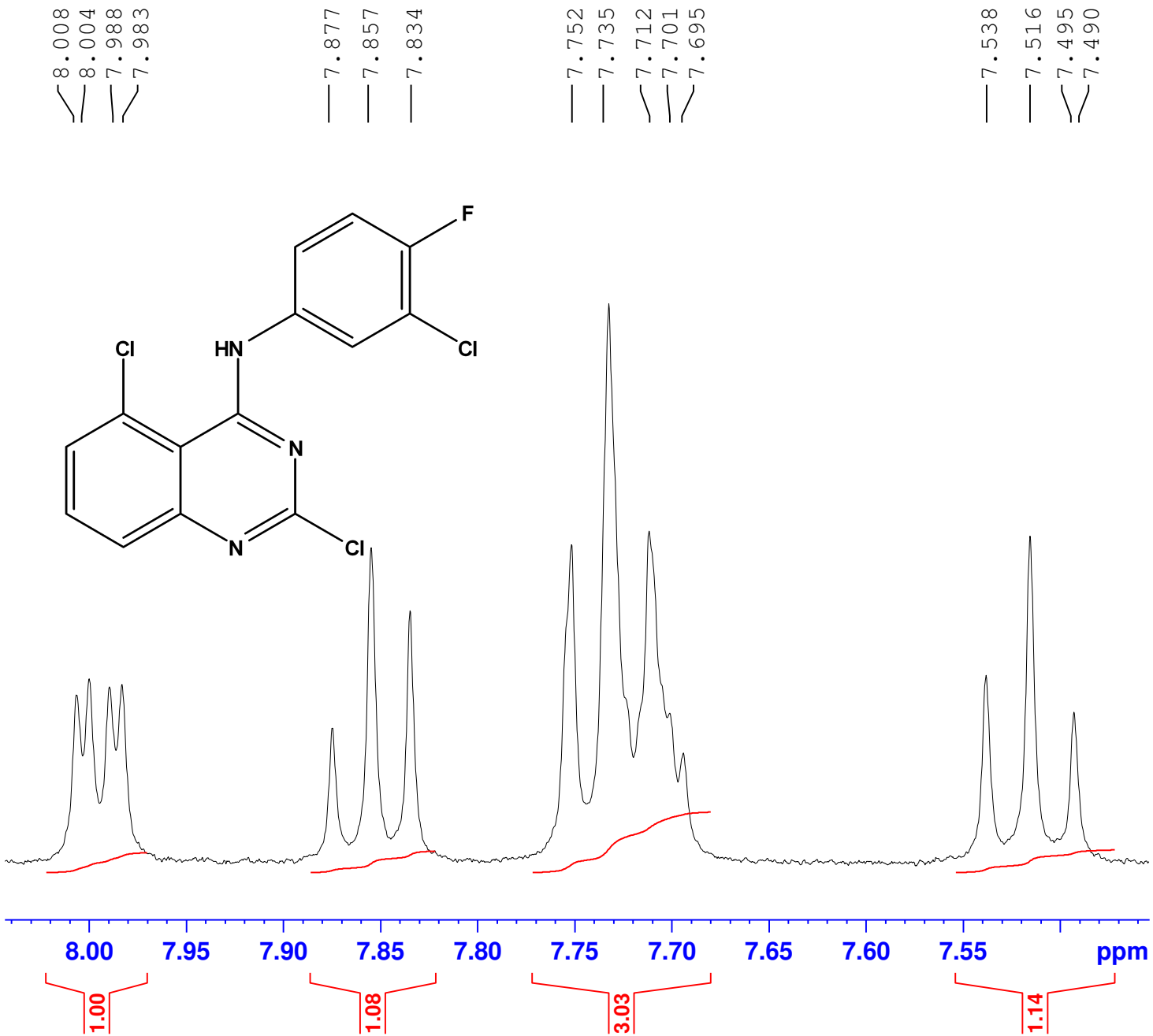


Current Data Parameters
NAME albraa-ibrahim-H4
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210401
Time 13.57
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 32
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00



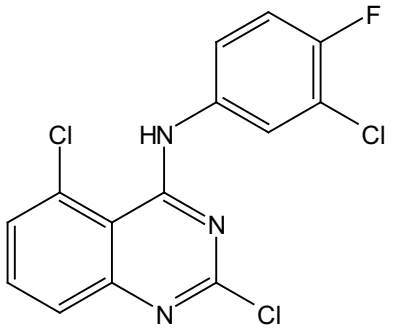
Current Data Parameters
NAME albraa-ibrahim-H4
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210401
Time 13.57
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 32
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

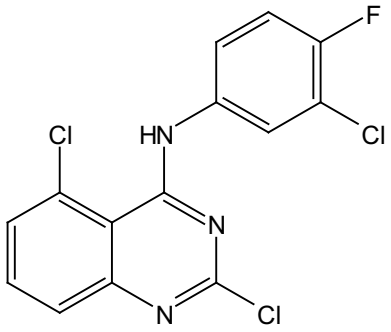
13C NMR of 11a



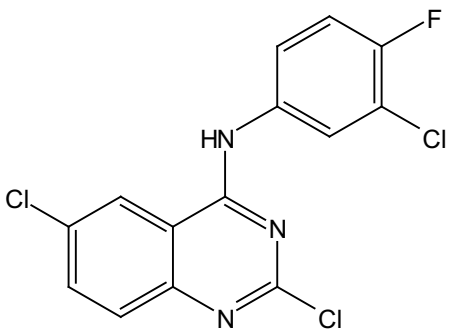
158.73
156.19
156.12
153.85
153.69
135.23
135.20
134.56
129.74
129.08
127.23
125.62
124.50
124.43
119.71
119.53
117.35
117.13
111.97
40.63 DMSO
40.42 DMSO
40.21 DMSO
40.00 DMSO
39.79 DMSO
39.59 DMSO
39.38 DMSO

f1 (ppm)

13C NMR of 11a



13C NMR of 11b

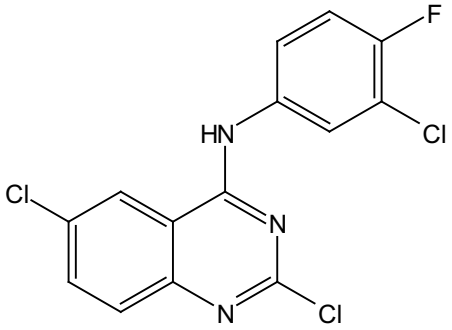


158.77
156.68
149.91
134.92
131.40
129.49
124.62
123.41
123.34
123.13
117.37
117.15
115.05
40.62 DMSO
40.41 DMSO
40.20 DMSO
39.99 DMSO
39.78 DMSO
39.57 DMSO
39.37 DMSO

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)

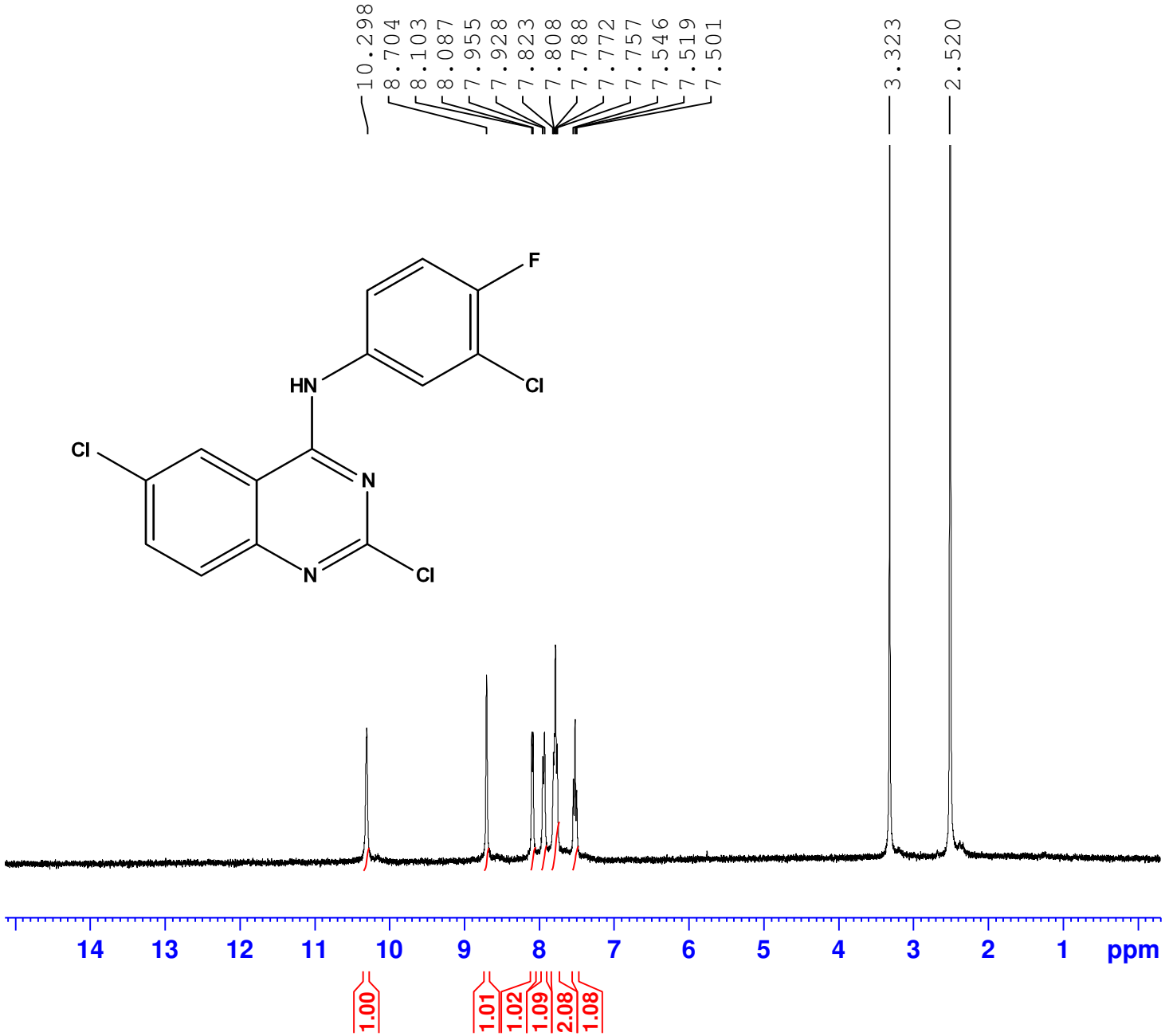
13C NMR of 11b



158.77
156.68
149.91
134.92
131.40
129.49
124.62
123.41
123.34
123.13
117.37
117.15
115.05

14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000

170
165
160
155
150
145
140
135
130
125
120
115
f1 (ppm)

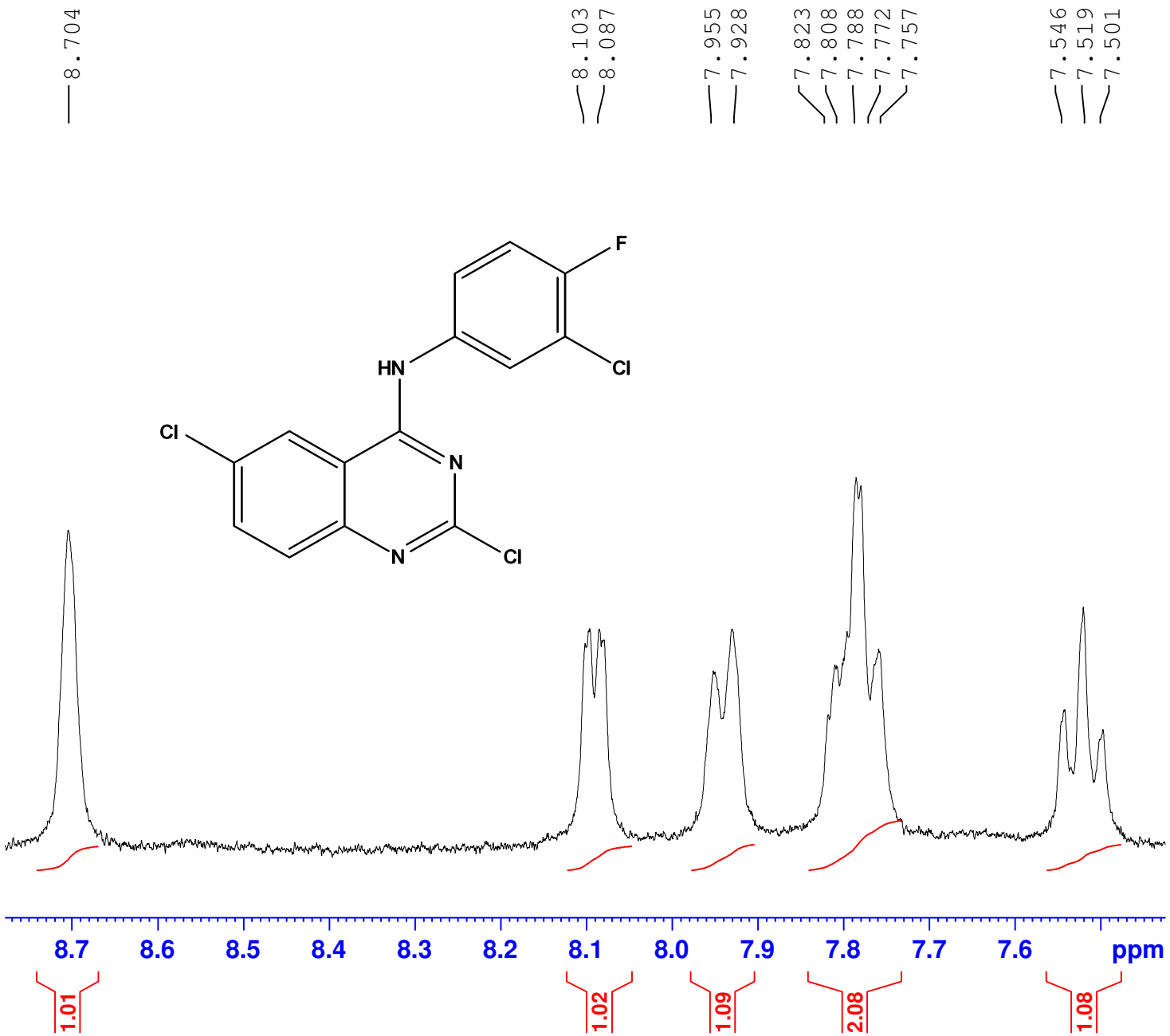


Current Data Parameters
NAME anas-ramadan-h2-3
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210401
Time 13.51
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 41
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

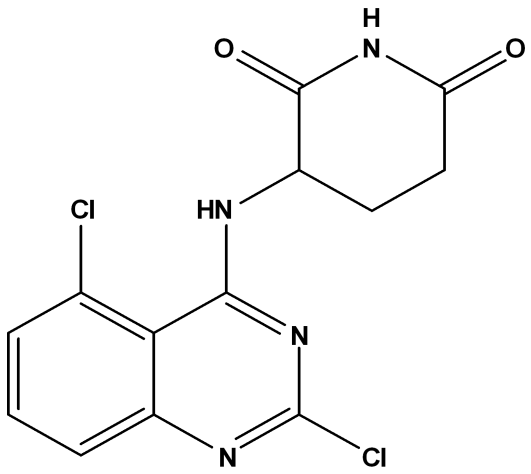
======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00



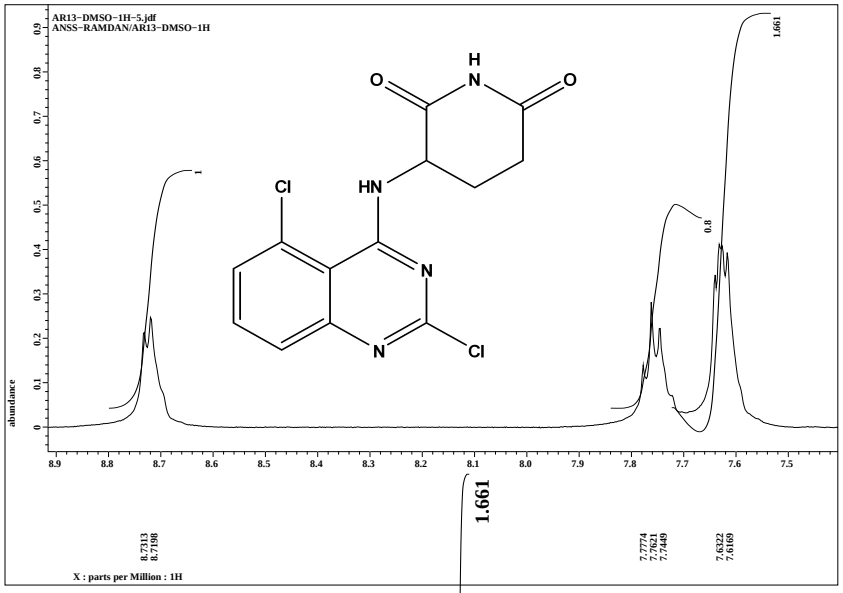

```
Current Data Parameters
NAME      anas-ramadan-h2-3
EXPNO                     1
PROCNO                    1

F2 - Acquisition Parameters
Date_              20210401
Time                 13.51
INSTRUM              spect
PROBHD    5 mm PABBO BB/
PULPROG               zg30
TD                   65536
SOLVENT               DMSO
NS                       41
DS                        2
SWH               8012.820 Hz
FIDRES           0.122266 Hz
AQ               4.0894465 sec
RG                  205.37
DW                  62.400 usec
DE                   6.50 usec
TE                  300.0 K
D1             1.00000000 sec
TD0                       1

======== CHANNEL f1 ========
SFO1          400.1524711 MHz
NUC1                    1H
P1                  12.00 usec
PLW1          18.00000000 W

F2 - Processing parameters
SI                    65536
SF            400.1500000 MHz
WDW                     EM
SSB       0
LB                   0.30 Hz
GB        0
PC                   1.00
```

AR13-DMSO-1H-5.jdf
ANSS-RAMDAN/AR13-DMSO-1H



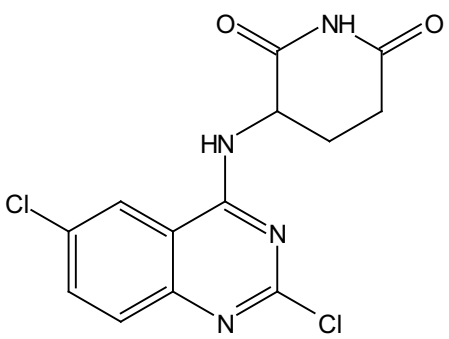
Author	= delta3
Content	= ANSS-RAMDAN/AR13
Creation_time	= 21-MAR-2021 19:3
Current_time	= 21-MAR-2021 12:3
Data_format	= 1D REAL
Dim_size	= 13107
Dim_title	= 1H
Dim_units	= [ppm]
Dimensions	= X
Filename	= AR13-DMSO-1H-5.j
Machine	= scc
Revision_time	= 21-MAR-2021 12:2
Sample_id	= ANSS-RAMDAN/AR13
Site	= ECA500 (Datum BL
Spectrometer	= DELTA2_NMR
Scans	= 32
Mod_return	= 1
Total_scans	= 32
X_points	= 16384
X_prescans	= 1
X_domain	= 1H
X_offset	= 5.0[ppm]
X_freq	= 500.15991521[MHz
X_sweep	= 15.6641604[kHz]
X_resolution	= 0.95606448[Hz]
Irr_domain	= 1H
Irr_offset	= 5.0[ppm]
Irr_freq	= 500.15991521[MHz
Tri_domain	= 1H
Tri_offset	= 5.0[ppm]
Tri_freq	= 500.15991521[MHz
X_acq_duration	= 1.04595456[s]
Digital_filter	= TRUE
Filter_factor	= 8
Af_version	= 1
Delay_of_start	= 1.99999974[s]
Actual_start_time	= 21-MAR-2021 19:3
Acq_delay	= 7.94[us]
Digital_filter_status	= 2P
Clipped	= FALSE
Dc_balanced	= FALSE
X90	= 12[us]
Irr90	= 12[us]
Tri90	= 10[us]
Qua90	= 10[us]
Qui90	= 10[us]
Sex90	= 10[us]
Sep90	= 10[us]
Oct90	= 10[us]
Non90	= 10[us]
Dec90	= 10[us]
X90_hi	= 92[us]
Irr90_hi	= 92[us]
Tri90_hi	= 10[us]
Qua90_hi	= 10[us]
Qui90_hi	= 10[us]
Sex90_hi	= 10[us]
Sep90_hi	= 10[us]
Oct90_hi	= 10[us]
Non90_hi	= 10[us]
Dec90_hi	= 10[us]
X90_lo	= 92[us]
Irr90_lo	= 92[us]
Tri90_lo	= 10[us]
Qua90_lo	= 10[us]
Qui90_lo	= 10[us]
Sex90_lo	= 10[us]
Sep90_lo	= 10[us]
Oct90_lo	= 10[us]
Non90_lo	= 10[us]

AR13−DMSO−1H−5.jdf
ANSS−RAMDAN/AR13−DMSO−1H

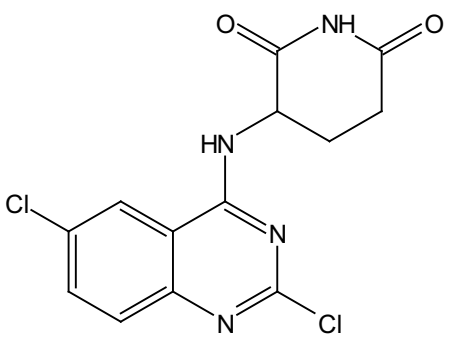


Author	= delta3
Content	= ANSS−RAMDAN/AR13
Creation_time	= 21−MAR−2021 19:3
Current_time	= 21−MAR−2021 12:3
Data_format	= 1D REAL
Dim_size	= 13107
Dim_title	= 1H
Dim_units	= [ppm]
Dimensions	= X
Filename	= AR13−DMSO−1H−5.j
Machine	= scc
Revision_time	= 21−MAR−2021 12:2
Sample_id	= ANSS−RAMDAN/AR13
Site	= ECA500 (Datum BL
Spectrometer	= DELTA2_NMR
Scans	= 32
Mod_return	= 1
Total_scans	= 32
X_points	= 16384
X_prescans	= 1
X_domain	= 1H
X_offset	= 5.0[ppm]
X_freq	= 500.15991521[MHz
X_sweep	= 15.6641604[kHz]
X_resolution	= 0.95606448[Hz]
Irr_domain	= 1H
Irr_offset	= 5.0[ppm]
Irr_freq	= 500.15991521[MHz
Tri_domain	= 1H
Tri_offset	= 5.0[ppm]
Tri_freq	= 500.15991521[MHz
X_acq_duration	= 1.04595456[s]
Digital_filter	= TRUE
Filter_factor	= 8
Af_version	= 1
Delay_of_start	= 1.99999974[s]
Actual_start_time	= 21−MAR−2021 19:3
Acq_delay	= 7.94[us]
Digital_filter_status	= 2P
Clipped	= FALSE
Dc_balanced	= FALSE
X90	= 12[us]
Irr90	= 12[us]
Tri90	= 10[us]
Qua90	= 10[us]
Qui90	= 10[us]
Sex90	= 10[us]
Sep90	= 10[us]
Oct90	= 10[us]
Non90	= 10[us]
Dec90	= 10[us]
X90_hi	= 92[us]
Irr90_hi	= 92[us]
Tri90_hi	= 10[us]
Qua90_hi	= 10[us]
Qui90_hi	= 10[us]
Sex90_hi	= 10[us]
Sep90_hi	= 10[us]
Oct90_hi	= 10[us]
Non90_hi	= 10[us]
Dec90_hi	= 10[us]
X90_lo	= 92[us]
Irr90_lo	= 92[us]
Tri90_lo	= 10[us]
Qua90_lo	= 10[us]
Qui90_lo	= 10[us]
Sex90_lo	= 10[us]
Sep90_lo	= 10[us]
Oct90_lo	= 10[us]
Non90_lo	= 10[us]

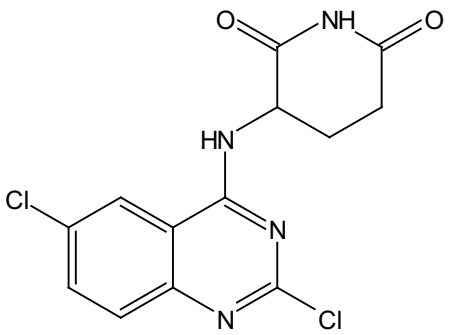
13C NMR of 13b

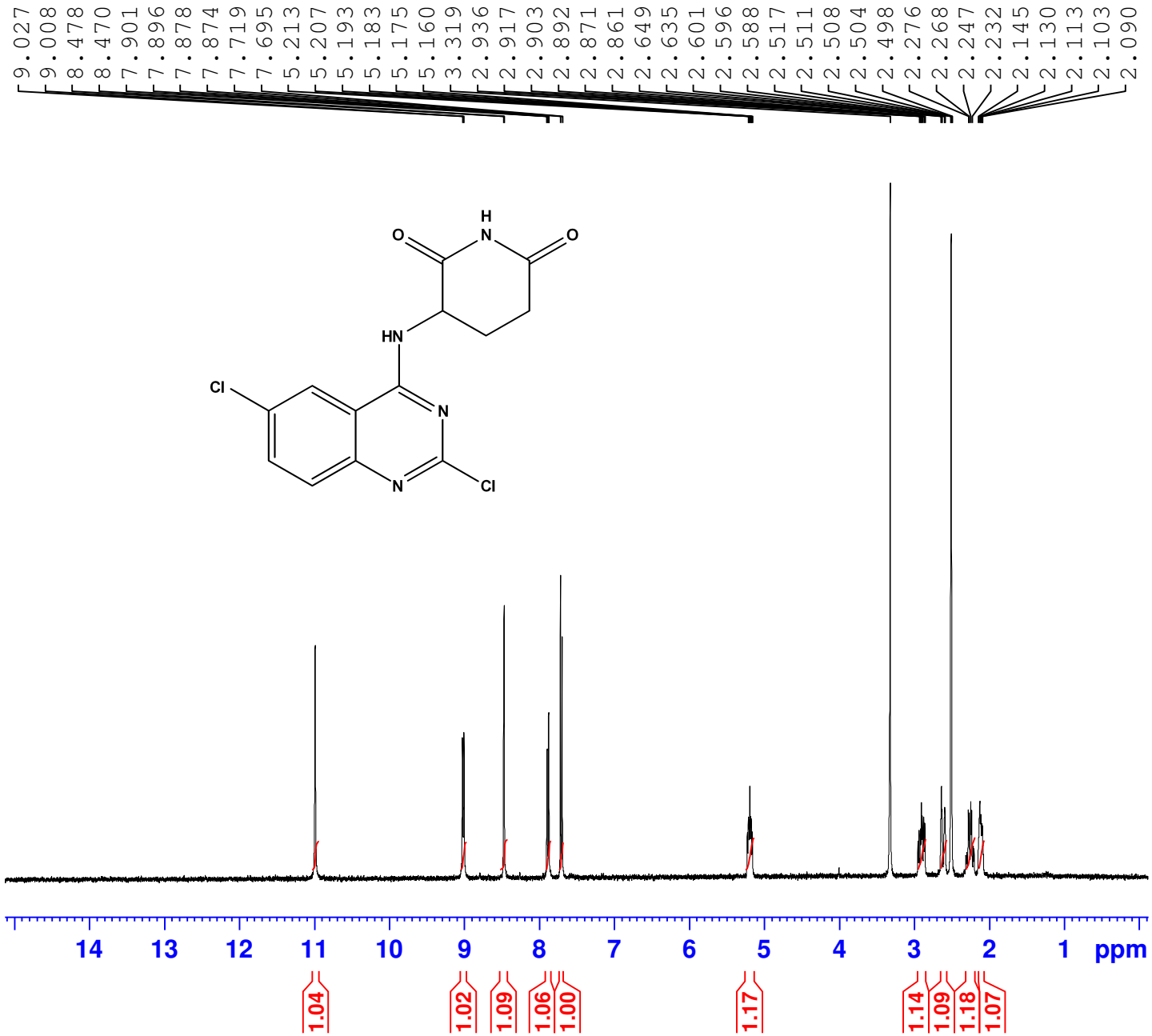


13C NMR of 13b



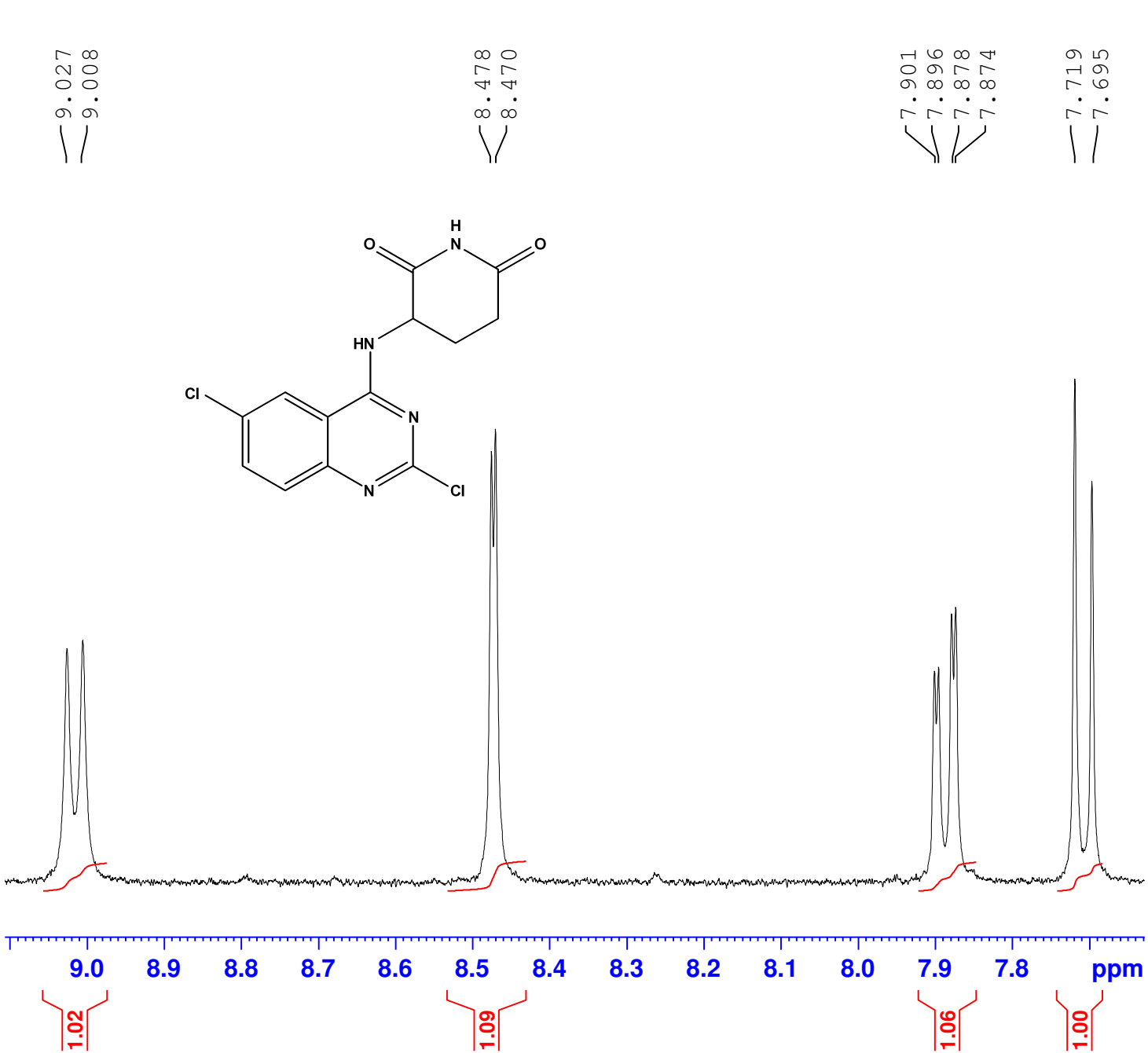
13C NMR of 13b





BRUKER

```
Current Data Parameters
NAME      anas-ramadan-h2-1
EXPNO                    1
PROCNO                   1

F2 - Acquisition Parameters
Date_           20210401
Time               13.31
INSTRUM            spect
PROBHD   5 mm PABBO BB/
PULPROG             zg30
TD                 65536
SOLVENT             DMSO
NS                    16
DS                     2
SWH             8012.820 Hz
FIDRES          0.122266 Hz
AQ             4.0894465 sec
RG                205.37
DW                62.400 usec
DE                 6.50 usec
TE                300.0 K
D1           1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
SFO1         400.1524711 MHz
NUC1                   1H
P1                 12.00 usec
PLW1         18.00000000 W

F2 - Processing parameters
SI                 65536
SF           400.1500000 MHz
WDW                   EM
SSB       0
LB                 0.30 Hz
GB        0
PC                 1.00
```

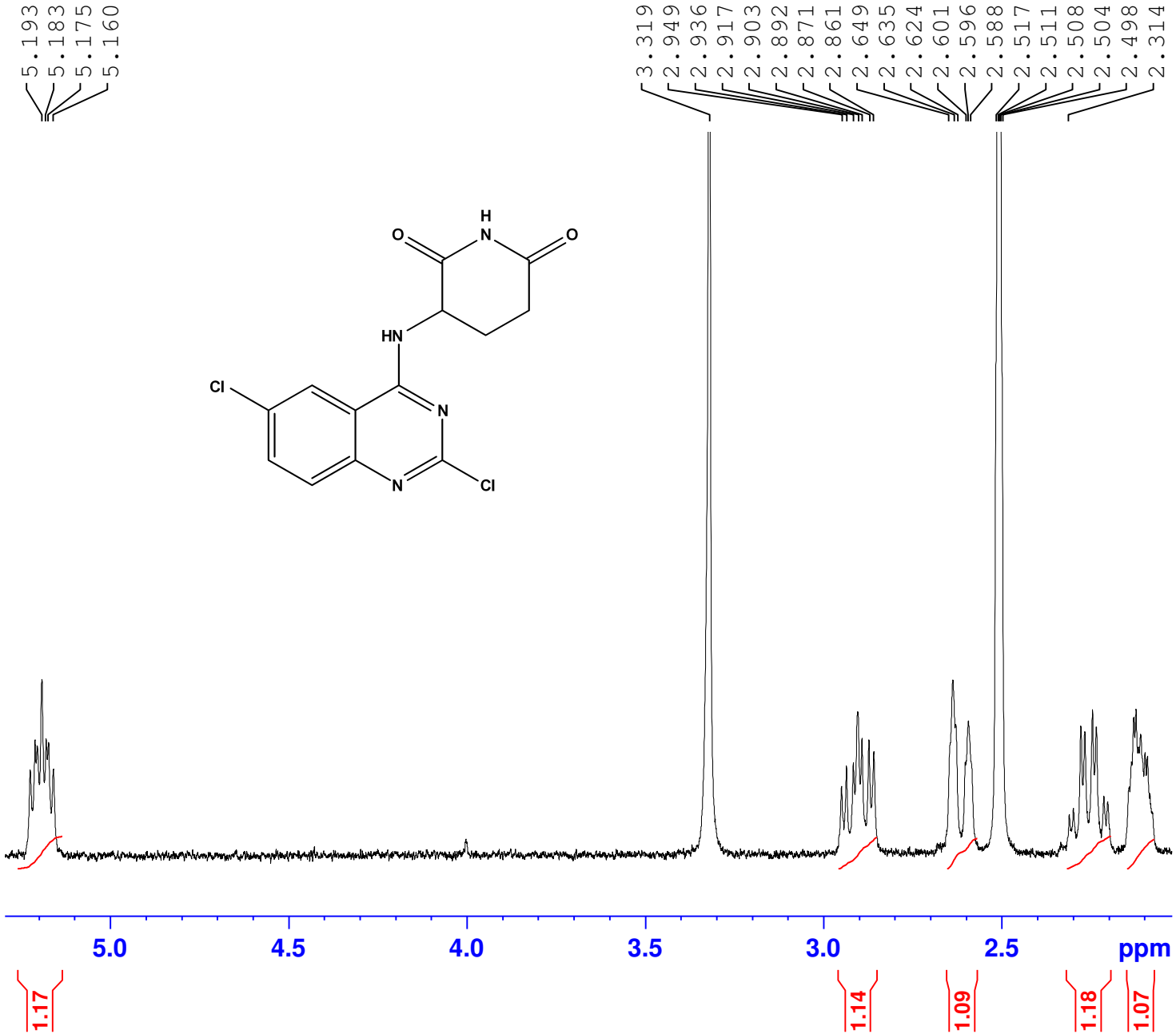


Current Data Parameters
NAME anas-ramadan-h2-1
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210401
Time 13.31
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 16
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00



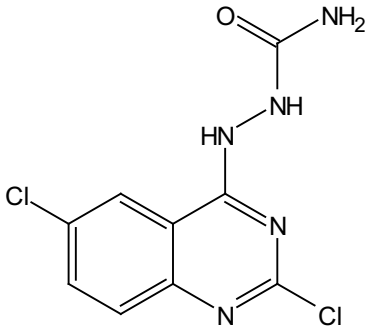
Current Data Parameters
NAME anas-ramadan-h2-1
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210401
Time 13.31
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 16
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

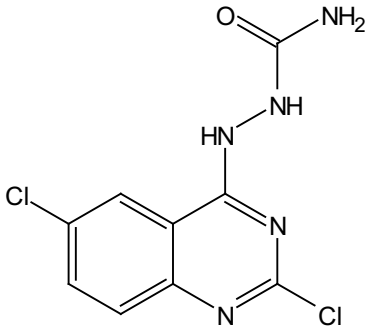
13C NMR of 15b

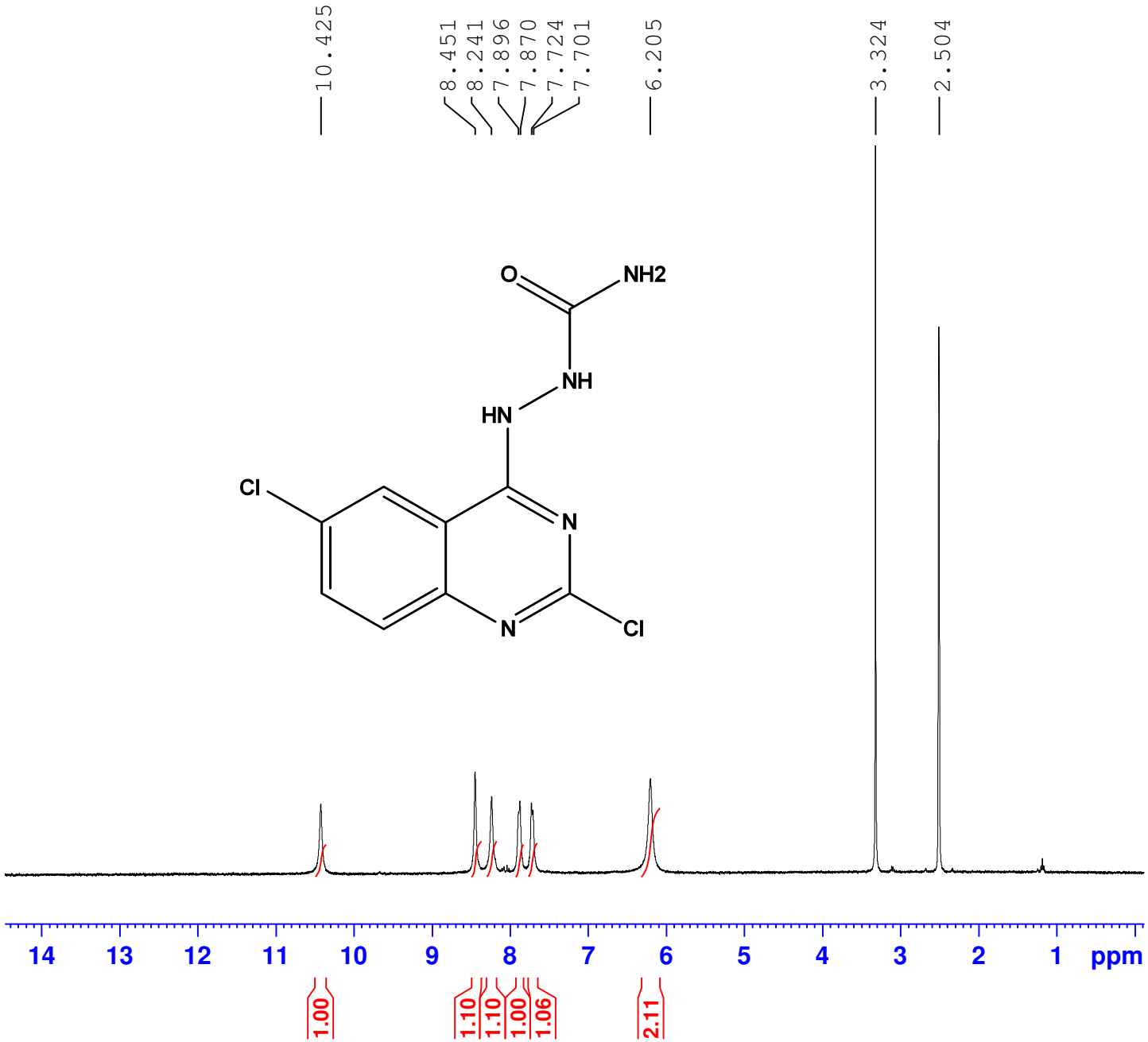


161.90
157.71
134.74
130.83
129.18
123.26
113.78
40.61 DMSO
40.40 DMSO
40.19 DMSO
39.78 DMSO
39.57 DMSO
39.36 DMSO

f1 (ppm)

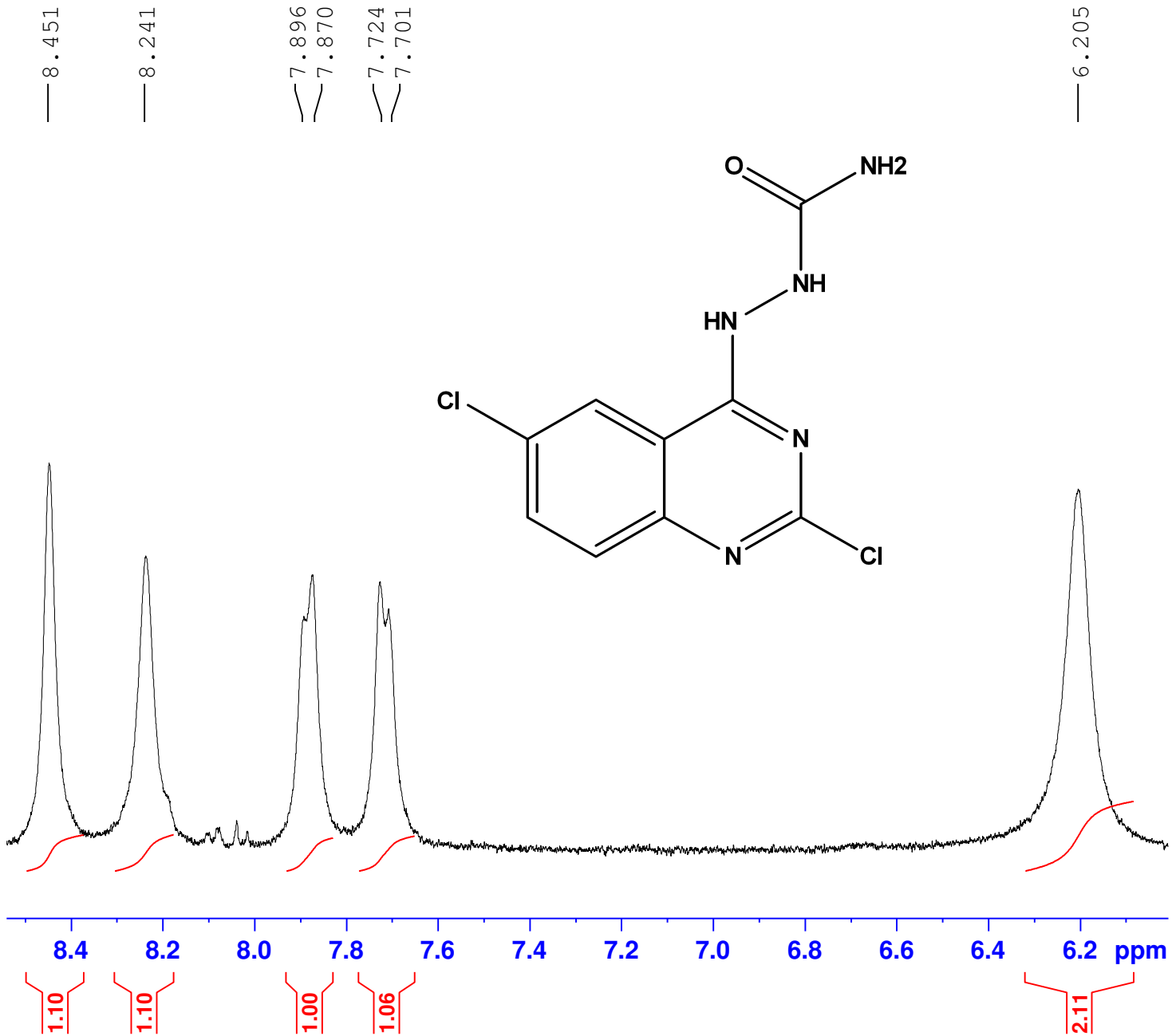
13C NMR of 15b





BRUKER

```
Current Data Parameters
NAME      anas-ramadan-h2-2
EXPNO                    1
PROCNO                   1

F2 - Acquisition Parameters
Date_           20210401
Time               13.44
INSTRUM            spect
PROBHD   5 mm PABBO BB/
PULPROG             zg30
TD                 65536
SOLVENT             DMSO
NS                    64
DS                     2
SWH             8012.820 Hz
FIDRES          0.122266 Hz
AQ              4.0894465 sec
RG                205.37
DW                62.400 usec
DE                 6.50 usec
TE                300.0 K
D1           1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
SFO1         400.1524711 MHz
NUC1                   1H
P1                12.00 usec
PLW1         18.00000000 W

F2 - Processing parameters
SI                 65536
SF           400.1500000 MHz
WDW                   EM
SSB       0
LB                 0.30 Hz
GB        0
PC                 1.00
```

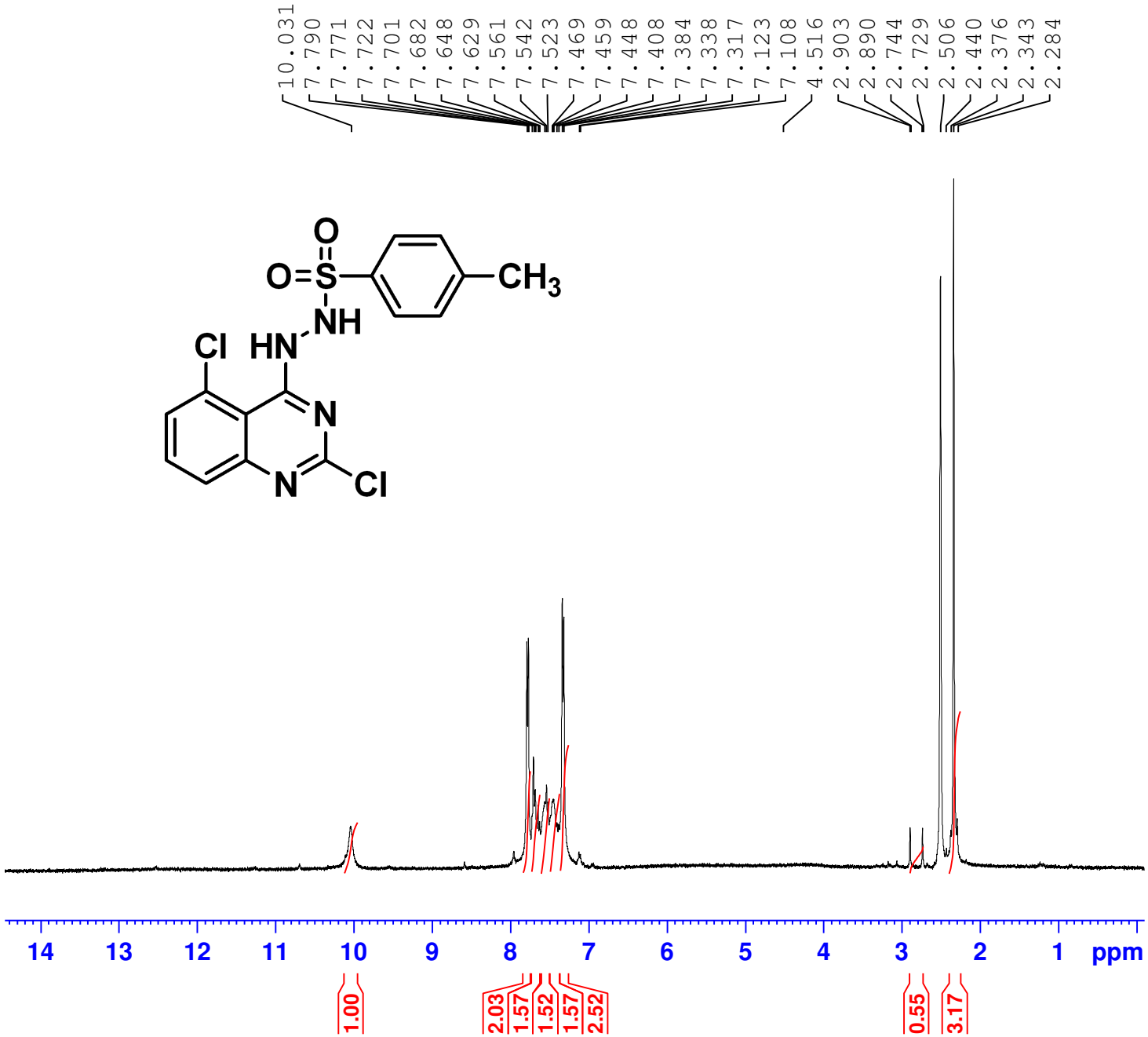


Current Data Parameters
NAME anas-ramadan-h2-2
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210401
Time 13.44
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 64
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

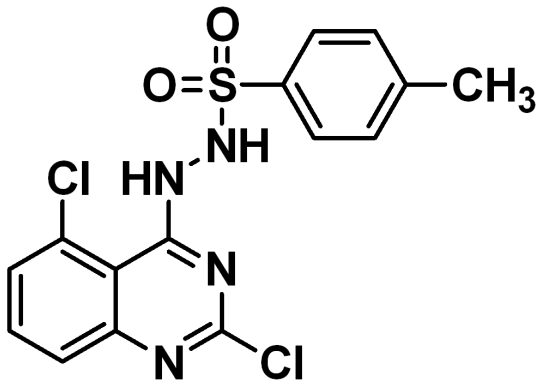
F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00



BRUKER

```
Current Data Parameters
NAME      albraa-ibrahim-AR18
EXPNO                       1
PROCNO                      1

F2 - Acquisition Parameters
Date_              20210401
Time                  14.27
INSTRUM               spect
PROBHD    5 mm PABBO BB/
PULPROG                zg30
TD                    65536
SOLVENT                DMSO
NS                       64
DS                        2
SWH                8012.820 Hz
FIDRES            0.122266 Hz
AQ                4.0894465 sec
RG                   205.37
DW                   62.400 usec
DE                    6.50 usec
TE                   300.0 K
D1              1.00000000 sec
TD0                       1

======== CHANNEL f1 ========
SFO1           400.1524711 MHz
NUC1                     1H
P1                   12.00 usec
PLW1           18.00000000 W

F2 - Processing parameters
SI                    65536
SF             400.1500000 MHz
WDW                      EM
SSB       0
LB                    0.30 Hz
GB        0
PC                    1.00
```

7.790
7.771
7.722
7.701
7.682
7.648
7.629
7.561
7.542
7.523
7.469
7.459
7.448
7.408
7.384
7.338
7.317



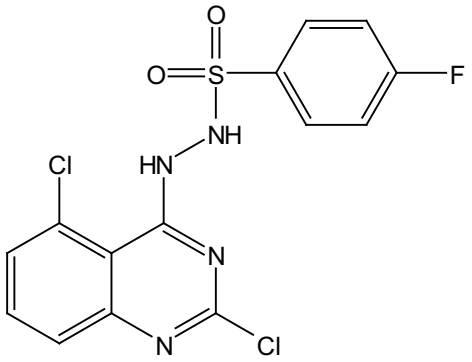
Current Data Parameters
NAME albraa-ibrahim-AR18
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210401
Time 14.27
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 64
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

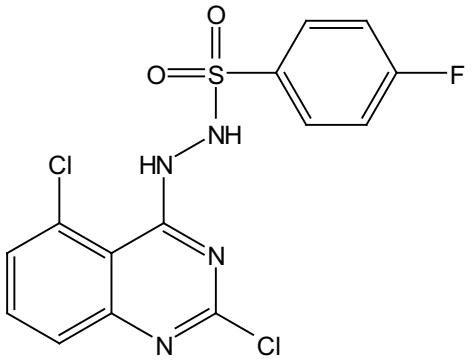
13C NMR of 19b

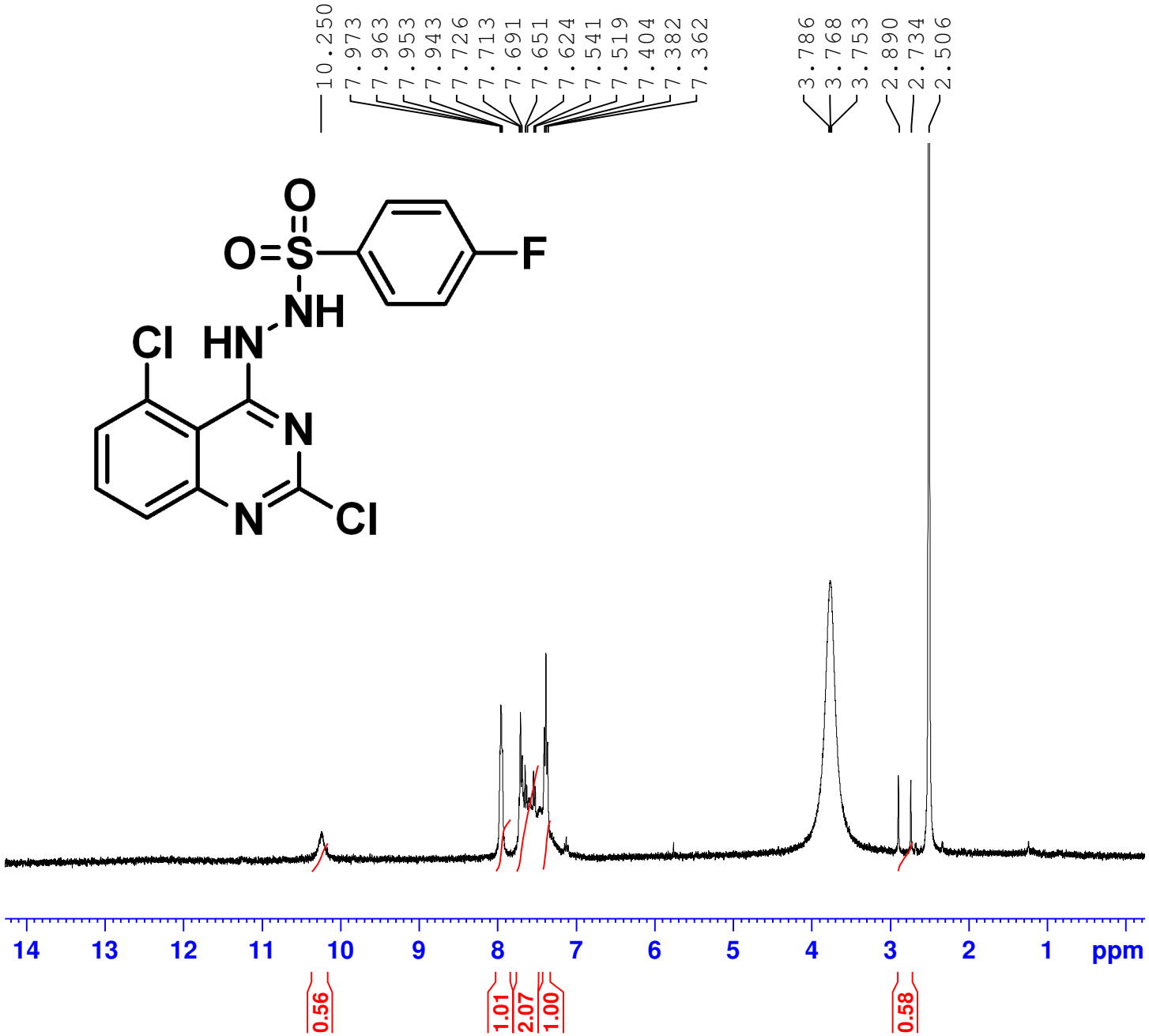


163.69
146.43
144.82
135.62
133.76
133.42
131.41
129.66
129.57
126.21
116.52
116.30
115.42
103.97
40.61 DMSO
40.40 DMSO
40.19 DMSO
39.78 DMSO
39.57 DMSO
39.36 DMSO

f1 (ppm)

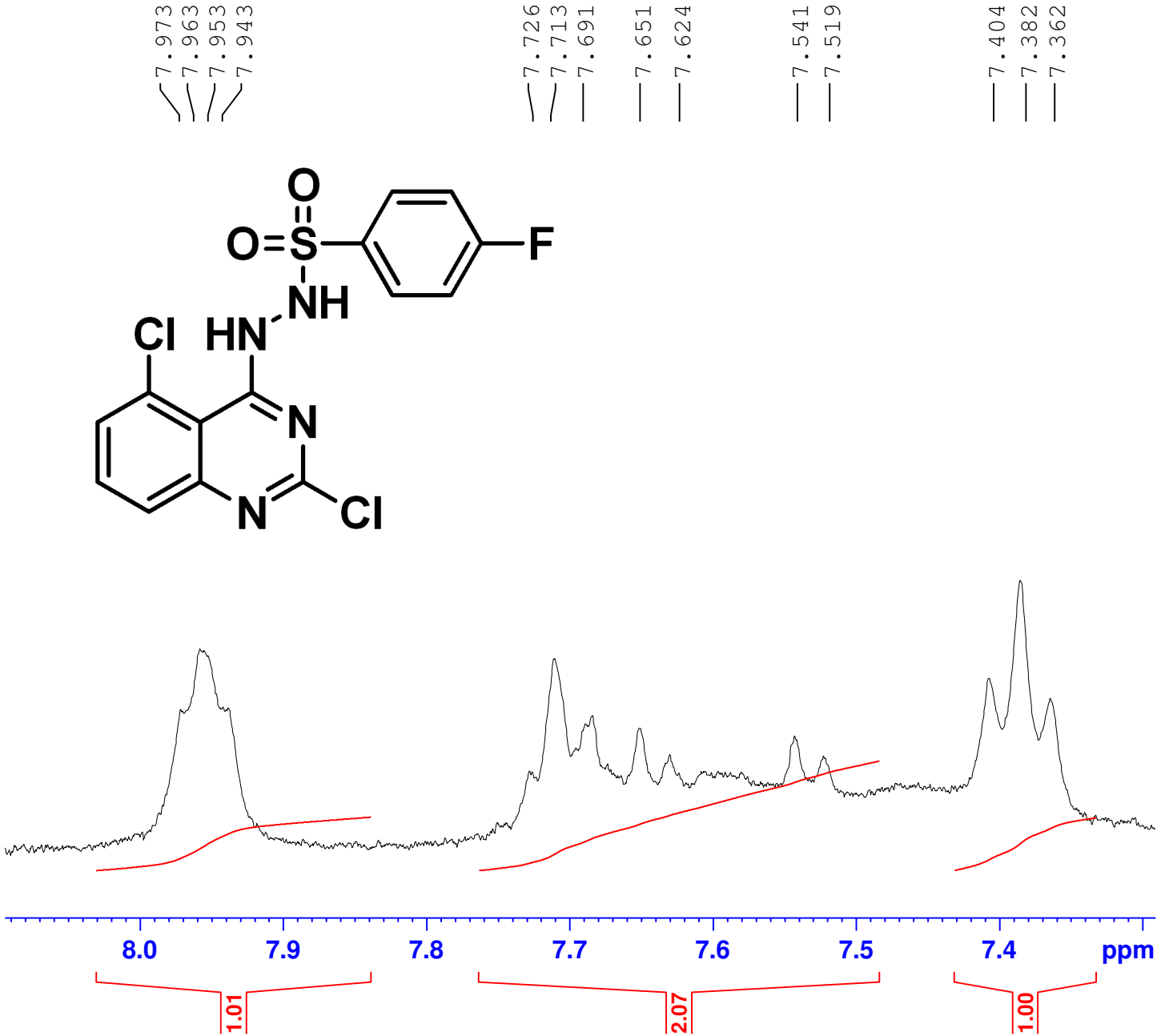
13C NMR of 19b





BRUKER

```
Current Data Parameters
NAME      albraa-ibrahim-AR17
EXPNO                      1
PROCNO                     1

F2 - Acquisition Parameters
Date_            20210401
Time                14.05
INSTRUM             spect
PROBHD    5 mm PABBO BB/
PULPROG              zg30
TD                  65536
SOLVENT              DMSO
NS                     64
DS                      2
SWH              8012.820 Hz
FIDRES           0.122266 Hz
AQ              4.0894465 sec
RG                 205.37
DW                 62.400 usec
DE                  6.50 usec
TE                 300.0 K
D1            1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
SFO1         400.1524711 MHz
NUC1                   1H
P1                 12.00 usec
PLW1         18.00000000 W

F2 - Processing parameters
SI                  65536
SF           400.1500000 MHz
WDW                    EM
SSB       0
LB                  0.30 Hz
GB        0
PC                  1.00
```

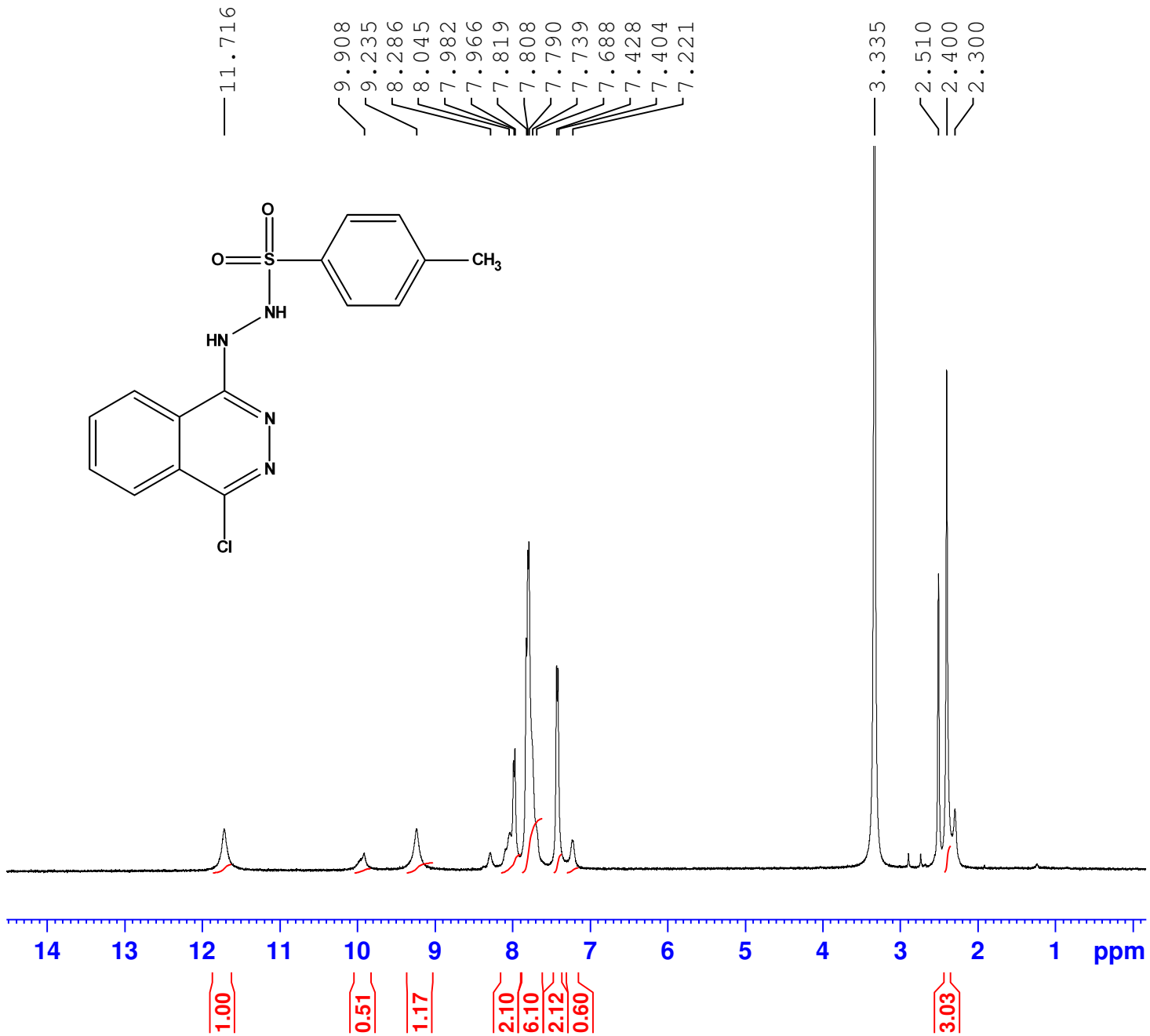


Current Data Parameters
NAME albraa-ibrahim-AR17
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210401
Time 14.05
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 64
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

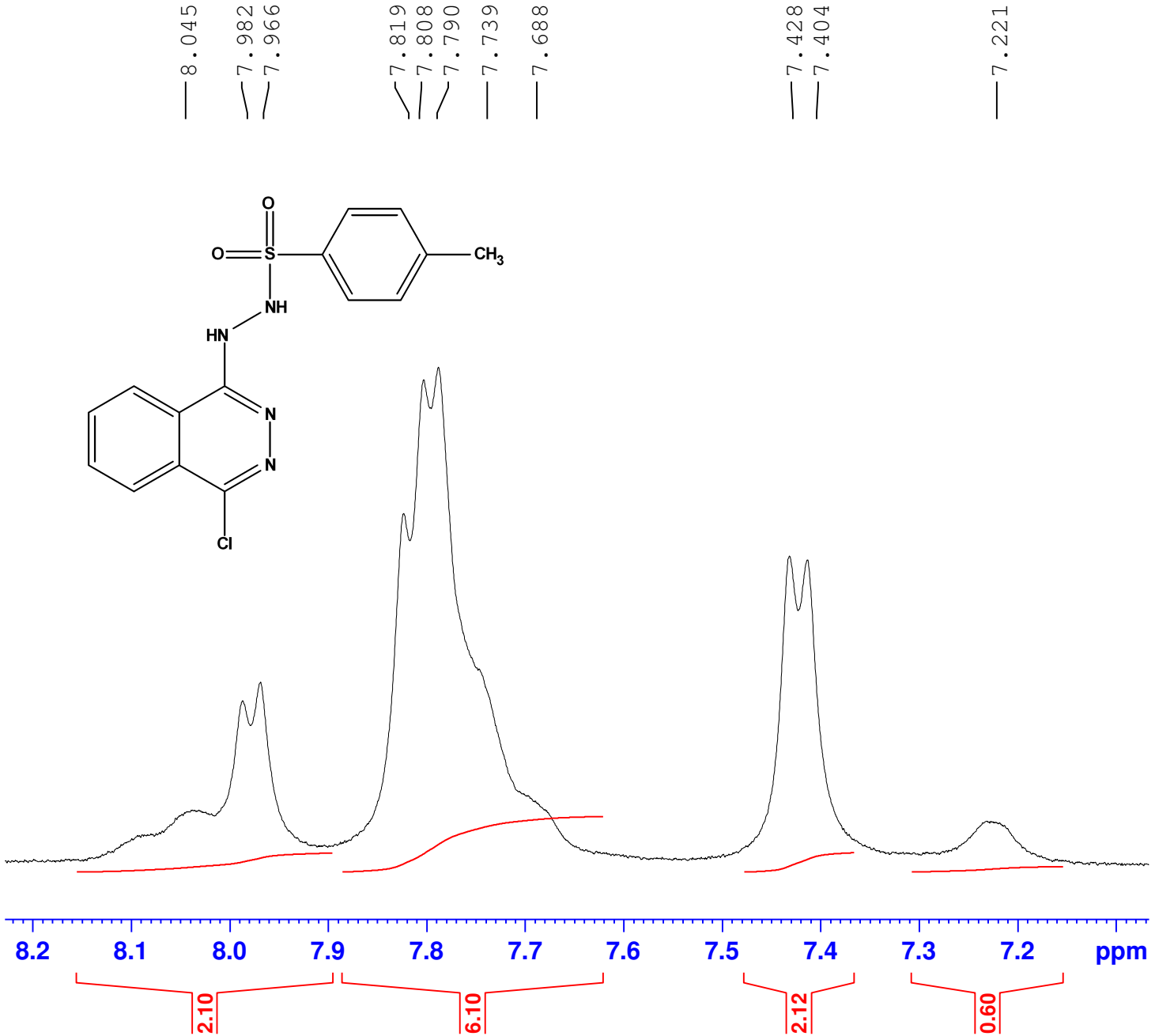
======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00



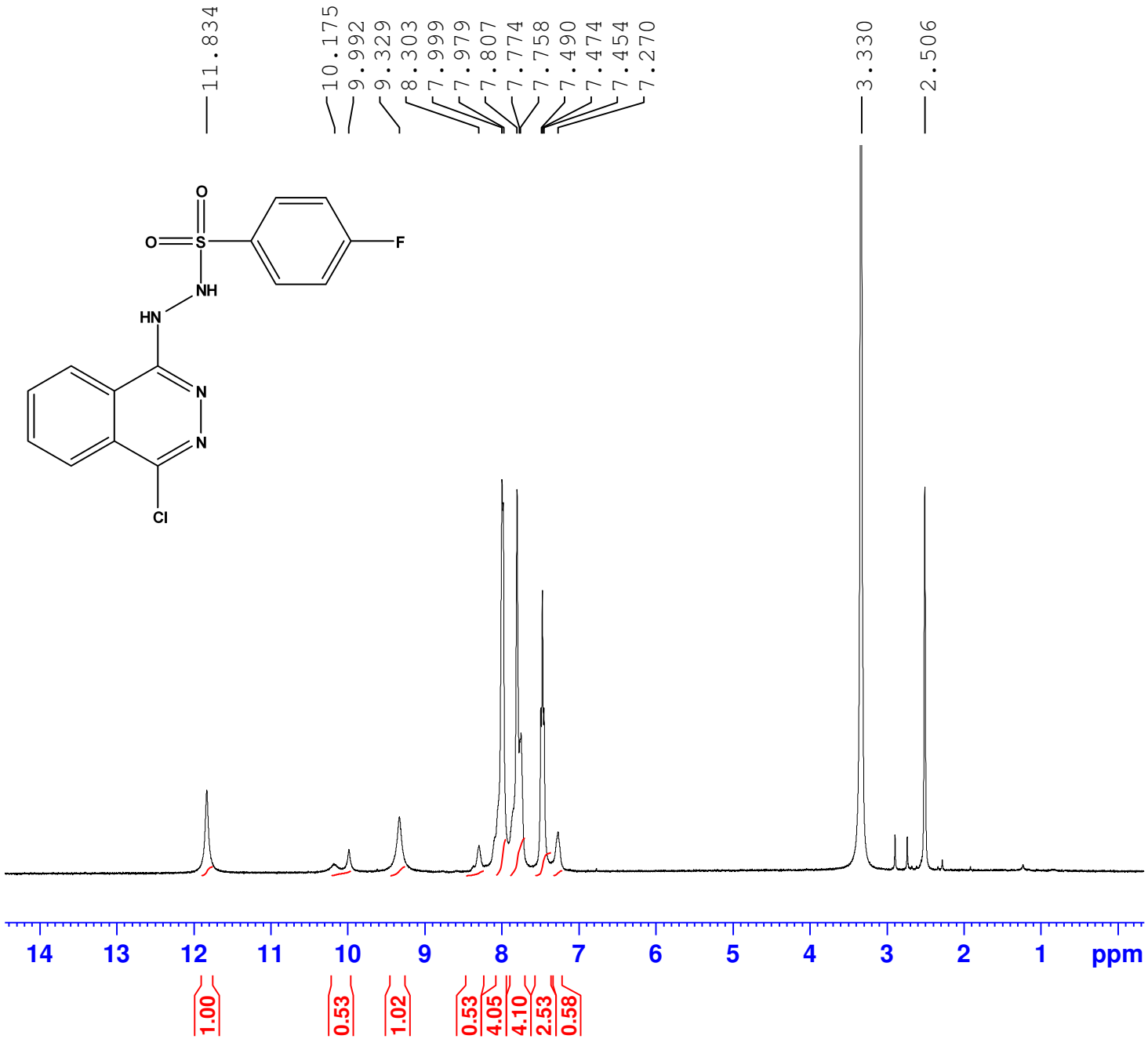
BRUKER

```
Current Data Parameters
NAME      Anas ramdan-AR11
EXPNO                    1
PROCNO                   1

F2 - Acquisition Parameters
Date_             20210228
Time                 15.37
INSTRUM              spect
PROBHD    5 mm PABBO BB/
PULPROG              zg30
TD                  65536
SOLVENT              DMSO
NS                     51
DS                      2
SWH              8012.820 Hz
FIDRES           0.122266 Hz
AQ               4.0894465 sec
RG                  205.37
DW                  62.400 usec
DE                   6.50 usec
TE                  300.0 K
D1             1.00000000 sec
TD0                     1

======== CHANNEL f1 ========
SFO1         400.1524711 MHz
NUC1                   1H
P1                  12.00 usec
PLW1         18.00000000 W

F2 - Processing parameters
SI                  65536
SF           400.1500000 MHz
WDW                    EM
SSB       0
LB                   0.30 Hz
GB        0
PC                   1.00
```



BRUKER

```
Current Data Parameters
NAME      Anas ramdan-AR11
EXPNO                    1
PROCNO                   1

F2 - Acquisition Parameters
Date_             20210228
Time                 15.37
INSTRUM              spect
PROBHD    5 mm PABBO BB/
PULPROG              zg30
TD                  65536
SOLVENT              DMSO
NS                     51
DS                      2
SWH              8012.820 Hz
FIDRES           0.122266 Hz
AQ               4.0894465 sec
RG                  205.37
DW                  62.400 usec
DE                   6.50 usec
TE                  300.0 K
D1              1.00000000 sec
TD0                      1

======== CHANNEL f1 ========
SFO1          400.1524711 MHz
NUC1                    1H
P1                  12.00 usec
PLW1          18.00000000 W

F2 - Processing parameters
SI                    65536
SF            400.1500000 MHz
WDW                     EM
SSB       0
LB                   0.30 Hz
GB        0
PC                   1.00
```

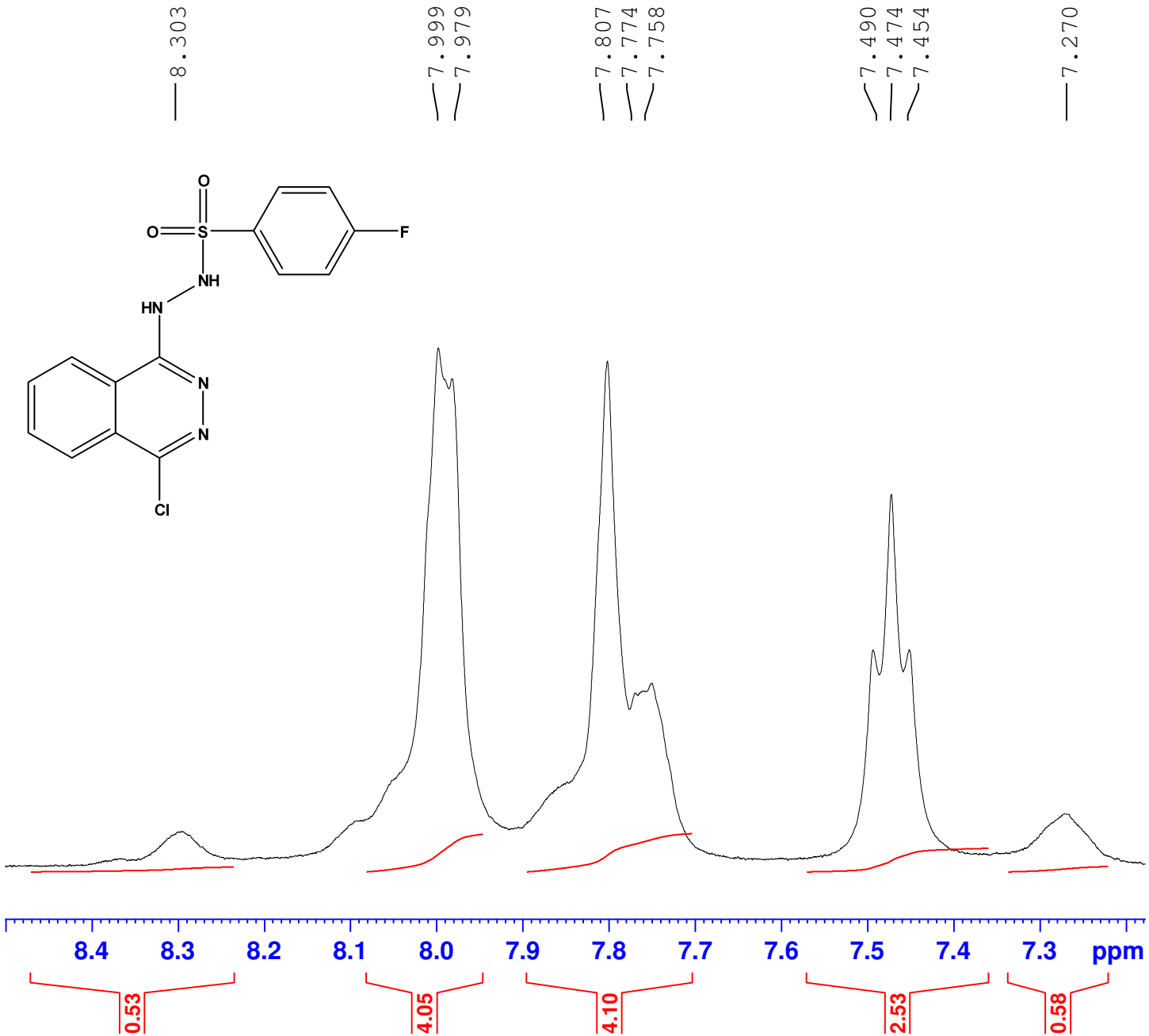


Current Data Parameters
NAME Anas ramdan-AR10
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210228
Time 15.29
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 102
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

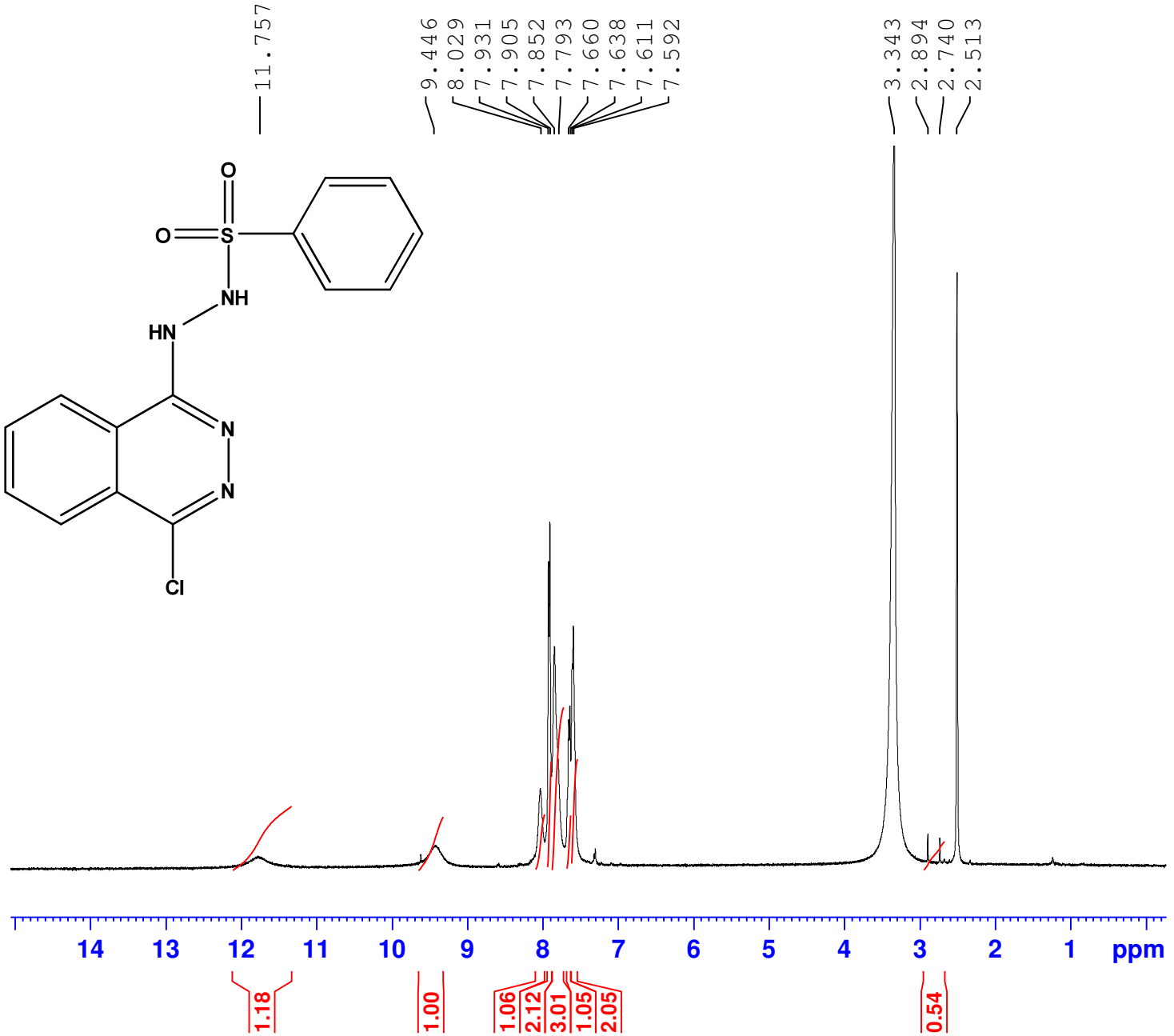


Current Data Parameters
NAME Anas ramdan-AR10
EXPNO 1
PROCNO 1

F2 - Acquisition Parameters
Date_ 20210228
Time 15.29
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 102
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

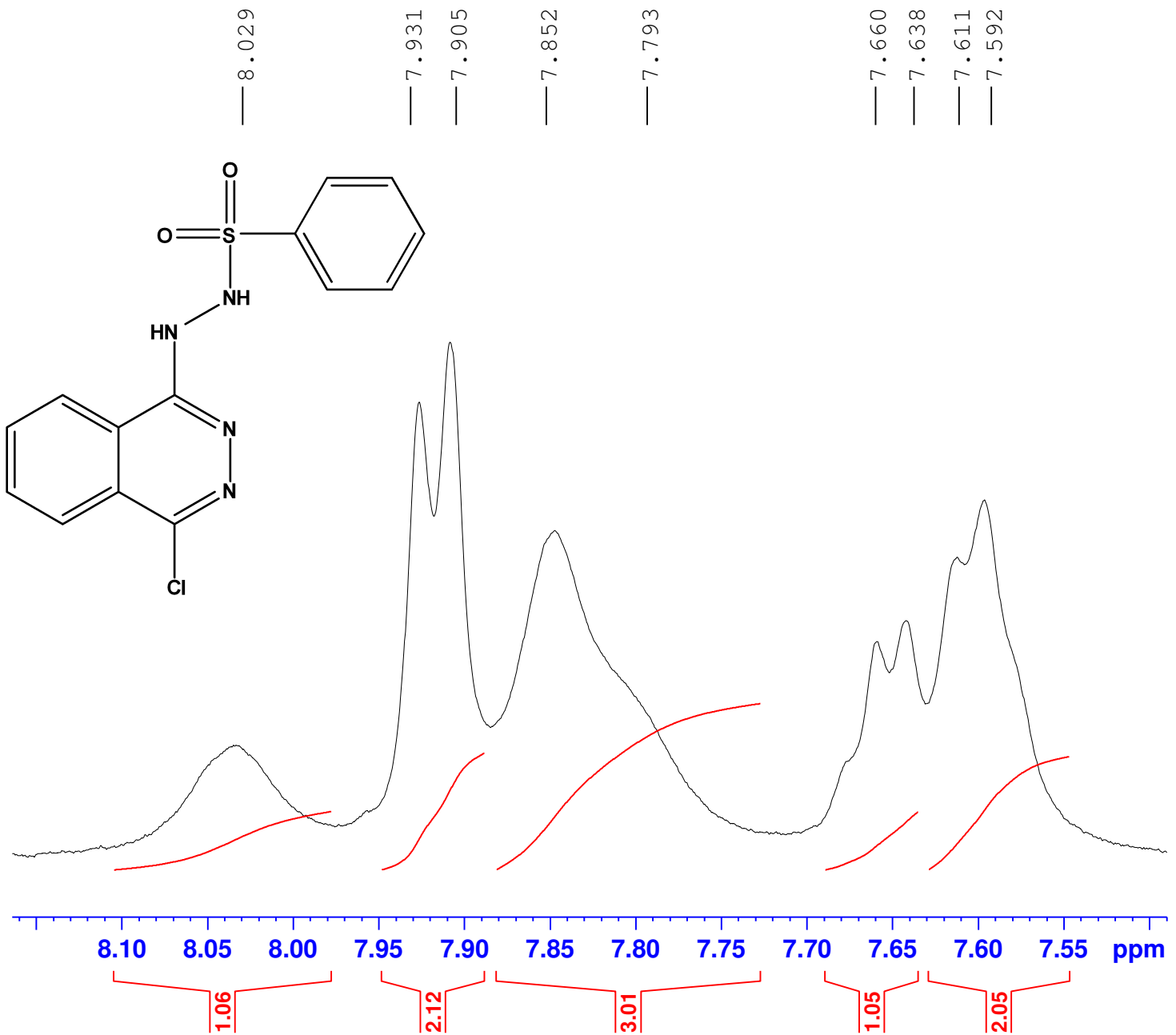
======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00



BRUKER

```
Current Data Parameters
NAME      Anas ramdan-AR12
EXPNO                    1
PROCNO                   1

F2 - Acquisition Parameters
Date_           20210228
Time               15.47
INSTRUM            spect
PROBHD   5 mm PABBO BB/
PULPROG             zg30
TD                 65536
SOLVENT             DMSO
NS                    86
DS                     2
SWH             8012.820 Hz
FIDRES          0.122266 Hz
AQ              4.0894465 sec
RG                205.37
DW                62.400 usec
DE                 6.50 usec
TE                300.0 K
D1           1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
SFO1         400.1524711 MHz
NUC1                   1H
P1                 12.00 usec
PLW1        18.00000000 W

F2 - Processing parameters
SI                 65536
SF          400.1500000 MHz
WDW                   EM
SSB       0
LB                 0.30 Hz
GB        0
PC                 1.00
```



Current Data Parameters
NAME Anas ramdan-AR12
EXPNO 1
PROCNO 1

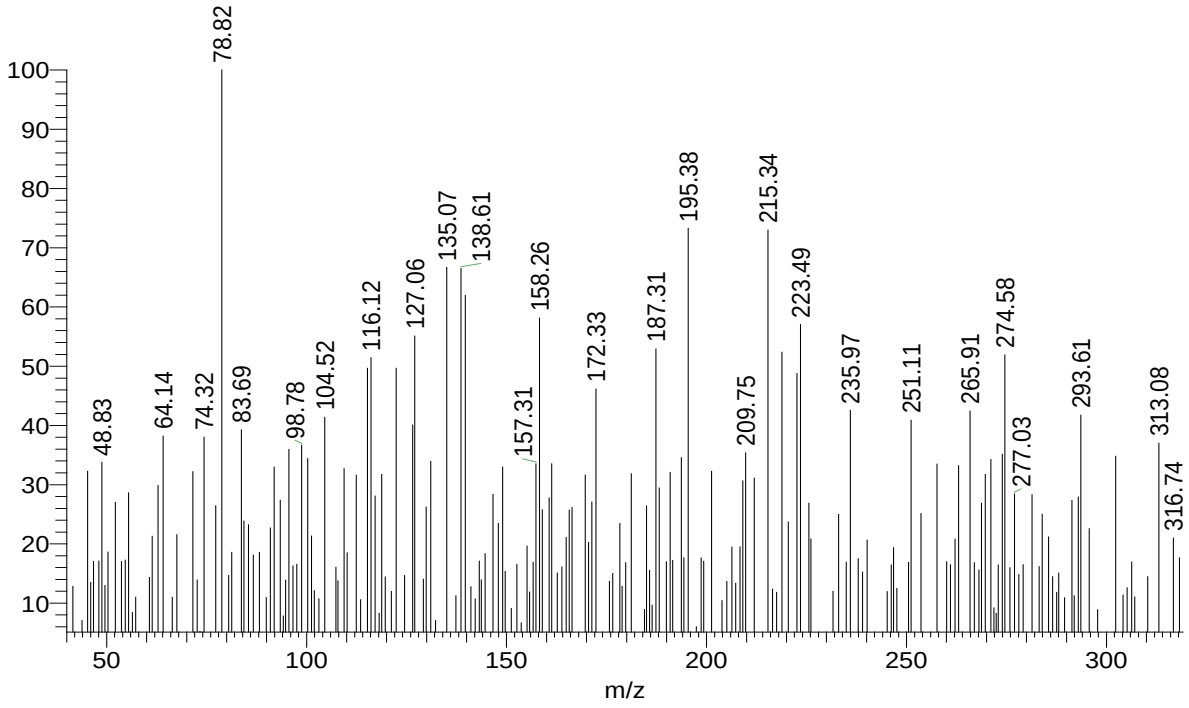
F2 - Acquisition Parameters
Date_ 20210228
Time 15.47
INSTRUM spect
PROBHD 5 mm PABBO BB/
PULPROG zg30
TD 65536
SOLVENT DMSO
NS 86
DS 2
SWH 8012.820 Hz
FIDRES 0.122266 Hz
AQ 4.0894465 sec
RG 205.37
DW 62.400 usec
DE 6.50 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1524711 MHz
NUC1 1H
P1 12.00 usec
PLW1 18.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1500000 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

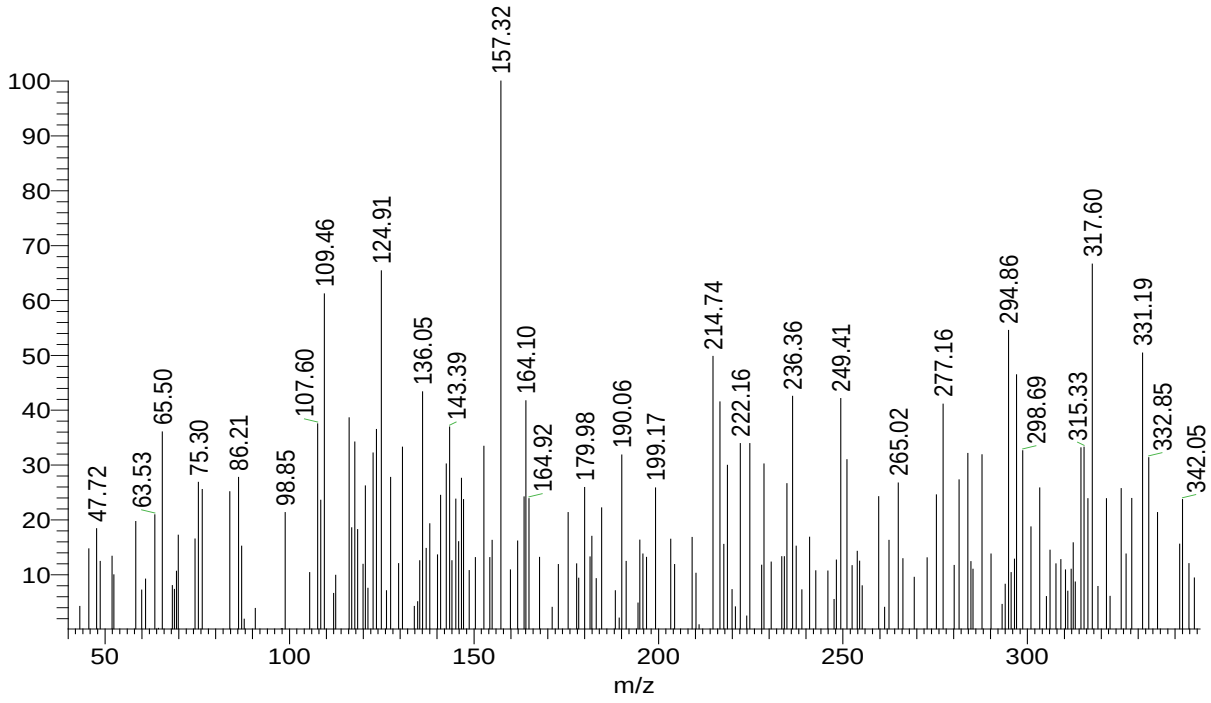
Mass spec. of 7b

Hazem-Osama-AR9 #54-63 RT: 0.92-1.07 AV: 10 SB: 2 2.44 , 2.44 NL: 8.08E1
T: + c EI Full ms [40.00-1000.00]



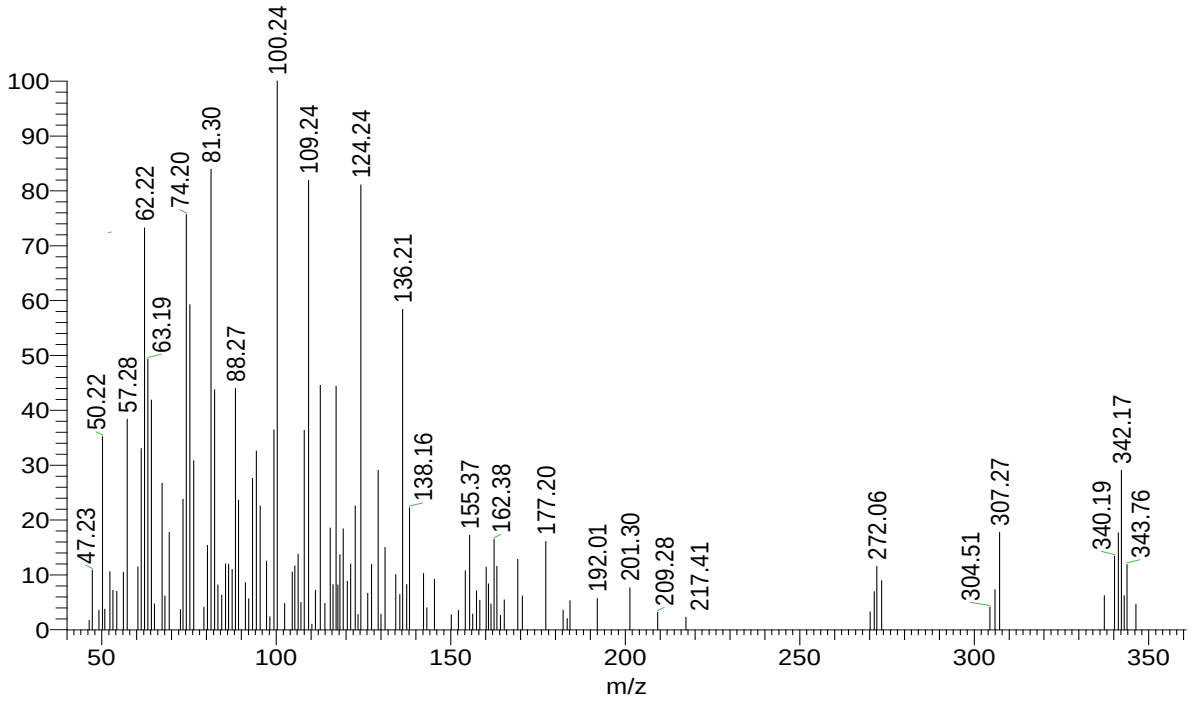
Mass spec. of 11a

Hazem-Osama-AR15 #125-129 RT: 2.11-2.18 AV: 5 SB: 6 2.73 , 2.48-2.54 NL: 1.65E2
T: + c EI Full ms [40.00-1000.00]



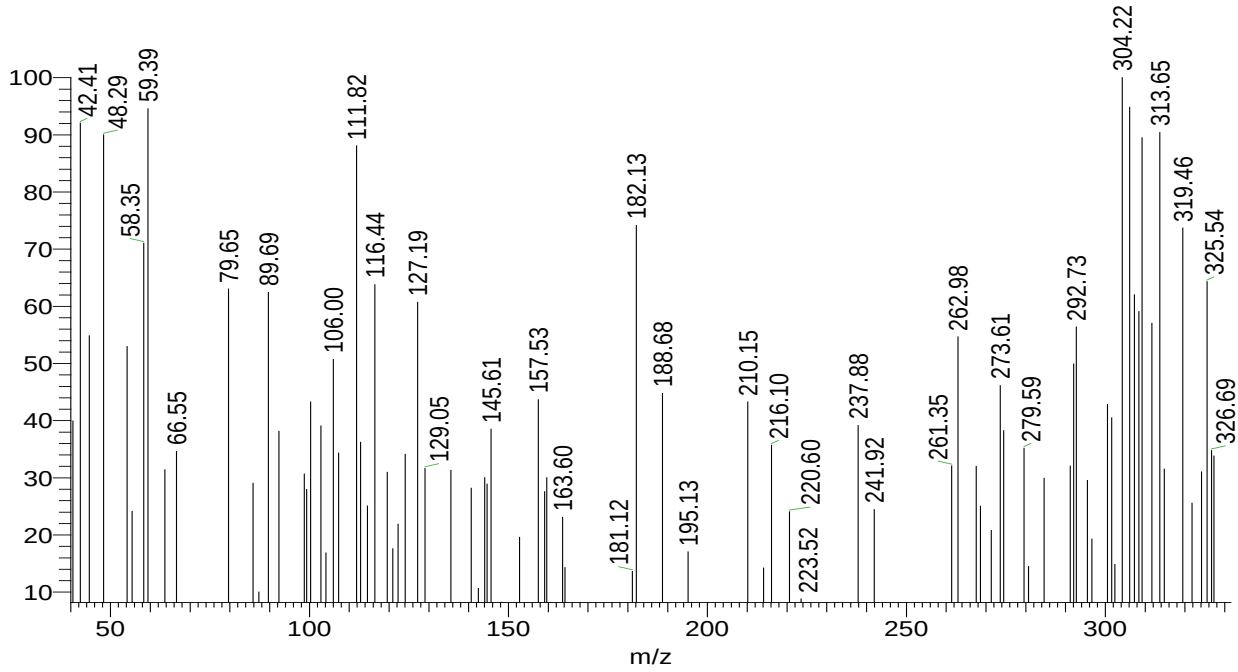
Mass spec. of 11b

Hazem-Osama-H2-3 #107 RT: 1.81 AV: 1 SB: 6 2.79 , 2.48-2.54 NL: 3.21E3
T: + c EI Full ms [40.00-1000.00]



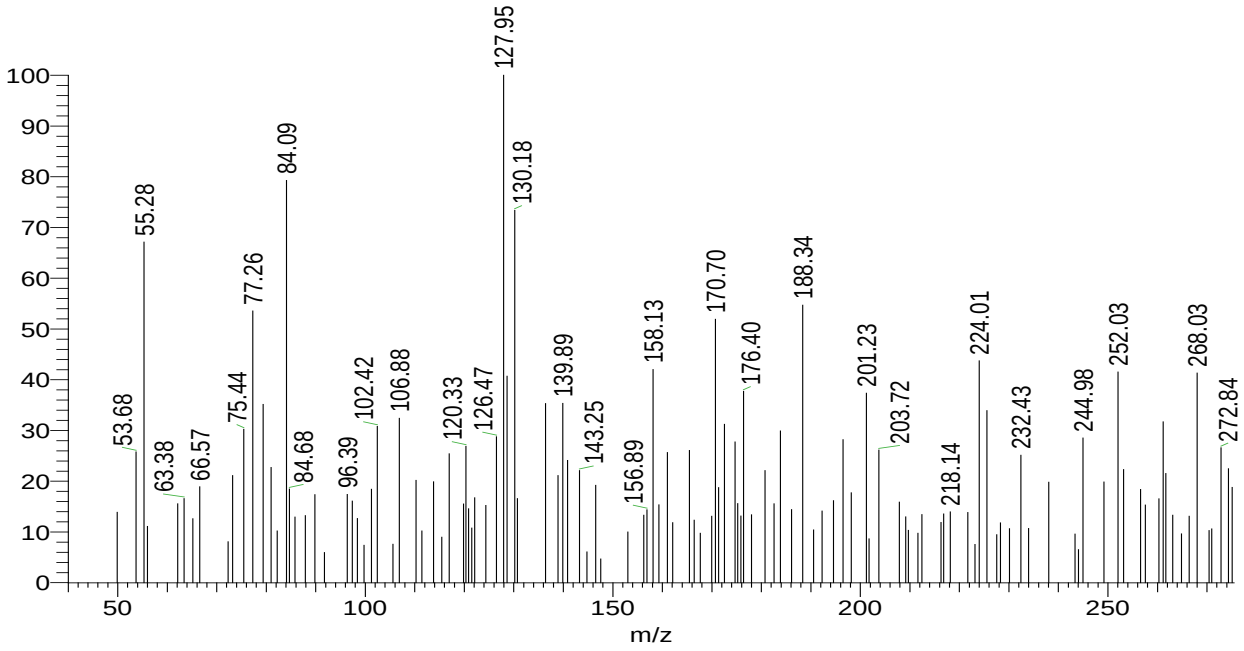
Mass spec. of 13b

Hazem-Osama-H2-1 #126 RT: 2.13 AV: 1 SB: 6 2.56 , 2.48-2.54 NL: 3.17E2
T: + c EI Full ms [40.00-1000.00]



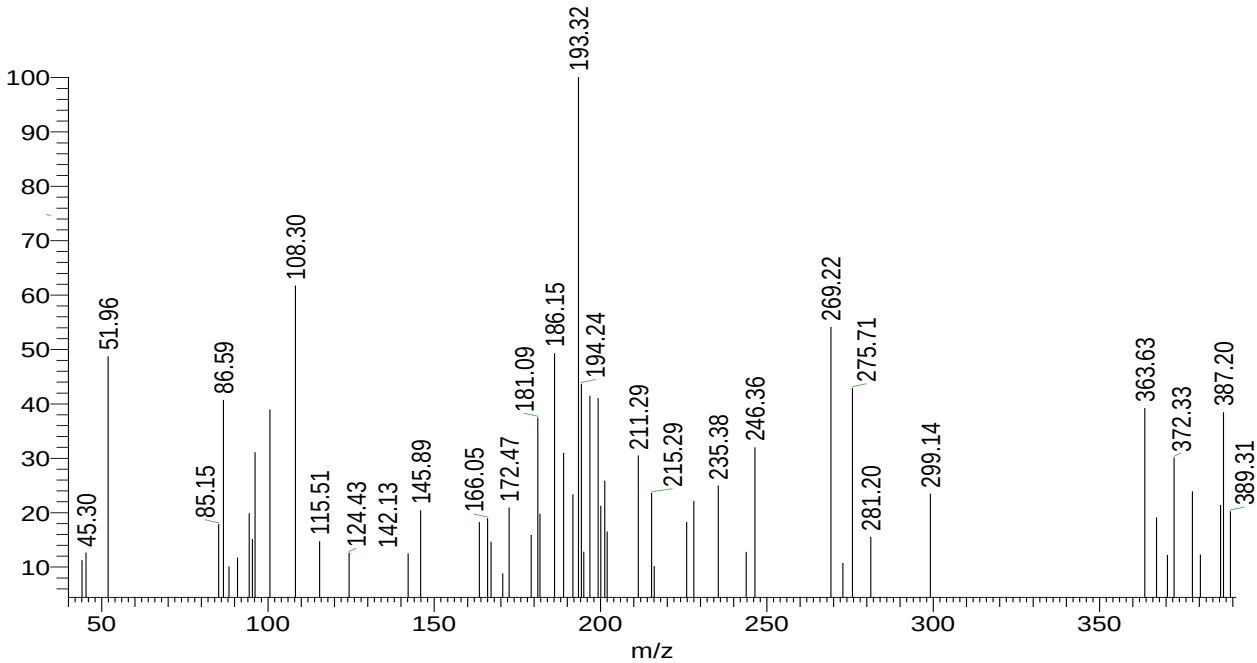
Mass spec. of 15b

Hazem-Osama-H2-2_20220731103618 #48-55 RT: 0.82-0.94 AV: 8 SB: 2 2.38 , 2.38 NL: 1.13E2
T: + c EI Full ms [40.00-1000.00]



Mass spec. of 19b

Hazem-Osama-AR17 #38 RT: 0.65 AV: 1 SB: 6 2.79 , 2.48-2.54 NL: 5.69E2
T: + c EI Full ms [40.00-1000.00]



Al-Azhar University
The Regional Center for Mycology and Biotechnology

Requester Data:

Name: **Dr. Anas Ramadan Abdel-Sattar Qotb**

Authority: Faculty of Pharmacy, Al-Azhar University

Sample Data:

Sixteen samples had been submitted for elemental analysis.

Analysis Report:

Sample Code	C%	H%	N%
AR1	58.41	5.40	14.61
AR2	53.72	5.49	16.98
AR3	55.35	5.98	16.30
AR5	59.63	5.81	13.97
AR6	65.18	4.96	12.19
AR7	54.73	2.85	13.81
AR8	59.38	4.12	18.32
AR9	60.91	4.42	17.87
AR10	47.85	2.97	15.62
AR11	51.89	3.85	16.32
AR12	50.34	3.45	16.90
AR13	48.28	3.34	17.41
AR14	39.88	2.75	25.53
AR15 (H4)	49.29	2.18	12.49
AR17	43.54	2.50	14.63
AR18	47.27	3.38	14.89

INVESTIGATOR **DIRECTOR**

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

Al-Azhar University
The Regional Center for Mycology and Biotechnology

Requester Data:

Name: **Dr. Anas Ramadan Abdel-Sattar Qotb**
Authority: Faculty of Pharmacy, Al-Azhar University

Sample Data:

Thirteen samples had been submitted for elemental analysis.

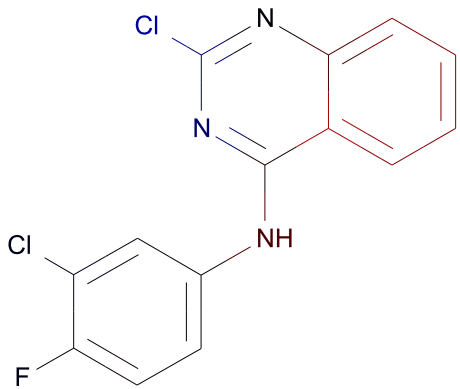
Analysis Report:

Sample Code	C%	H%	N%
H2-1	48.25	3.31	17.59
H2-2	39.90	2.78	25.91
H2-3	49.24	2.18	12.44
M1	49.71	4.27	13.96
M2	49.86	4.31	13.98
M3	52.29	4.50	14.47
M4	50.97	3.45	9.98
M5	50.99	3.48	9.93
M6	52.91	3.63	10.12
M7	52.79	3.68	10.04
Y1	54.60	4.41	10.28
Y2	54.56	4.49	10.02
Y3	56.78	4.62	10.68

INVESTIGATOR **DIRECTOR**

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

Toxicity report



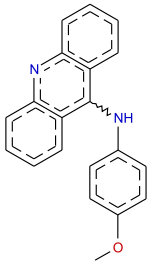
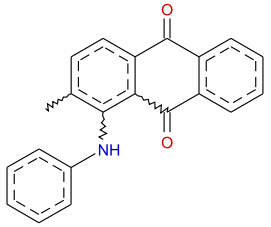
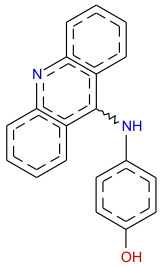
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mutagen
Probability: 0.75
Enrichment: 1.34
Bayesian Score: 0.611
Mahalanobis Distance: 7.68
Mahalanobis Distance p-value: 0.998

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

Structural Similar Compounds

Name	61421-82-7	4947-27-7	61421-83-8
Structure			
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.472	0.501	0.527
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

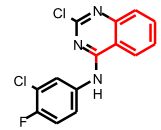
Model Applicability

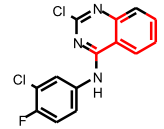
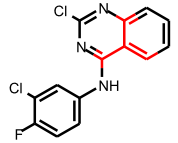
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

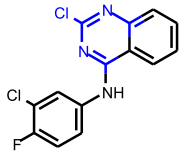
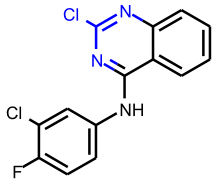
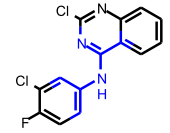
Feature Contribution

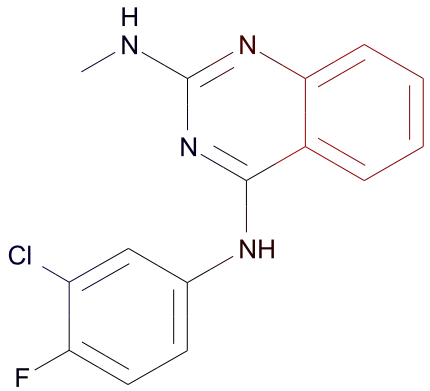
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.377	575 out of 690

SCFP_12	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.375	934 out of 1123
SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.887	2 out of 11
SCFP_12	1328855840	 [*]:[cH]:[c](N[c](:n: [*]):[c](:[*]):[*]): [cH]:[*]	-0.496	1 out of 4



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Mutagen

Probability: 0.781

Enrichment: 1.4

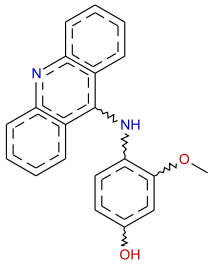
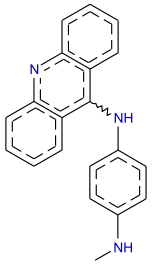
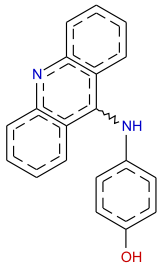
Bayesian Score: 2.66

Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.236

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

Structural Similar Compounds

Name	106063-42-7	75776-00-0	61421-83-8
Structure			
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.433	0.477	0.493
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

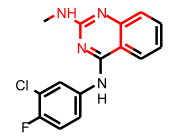
Model Applicability

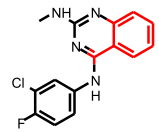
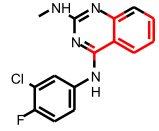
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

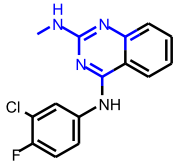
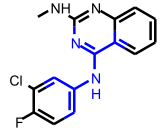
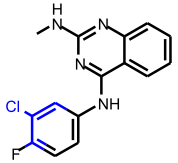
Feature Contribution

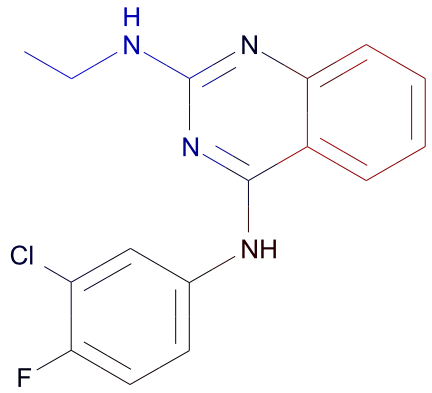
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1551584806	 [*]N[c]1:n:[*]:[c](:[*]):[c](:[cH]:[*]):n:1	0.464	46 out of 50

SCFP_12	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690
SCFP_12	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.375	934 out of 1123

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-760164567	 [*][c]1:[*]:[c](:[*]) :n:[c](NC):n:1	-1.35	0 out of 5
SCFP_12	1328855840	 [*]:[cH]:[c](N[c](:n: [*]):[c](:[*]):[*]): [cH]:[*]	-0.496	1 out of 4
SCFP_12	-601571304	 [*][c](:[*]):[c](Cl): [cH]:[*]	-0.465	154 out of 430



$C_{16}H_{14}ClFN_4$

Molecular Weight: 316.76056

ALogP: 5.025

Rotatable Bonds: 4

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.717

Enrichment: 1.28

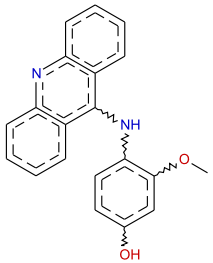
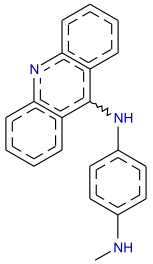
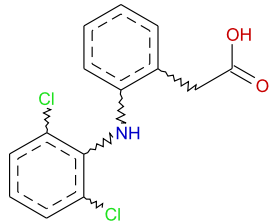
Bayesian Score: -1.04

Mahalanobis Distance: 9.34

Mahalanobis Distance p-value: 0.673

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

Structural Similar Compounds

Name	106063-42-7	75776-00-0	DICLOFENAC
Structure			
Actual Endpoint	Mutagen	Mutagen	Non-Mutagen
Predicted Endpoint	Mutagen	Mutagen	Non-Mutagen
Distance	0.482	0.505	0.543
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	PDR 1994

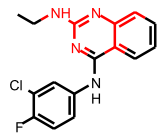
Model Applicability

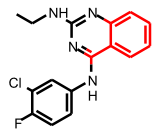
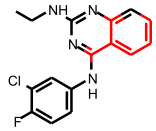
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

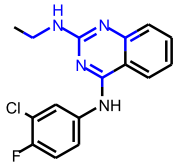
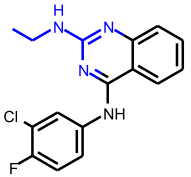
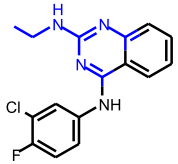
Feature Contribution

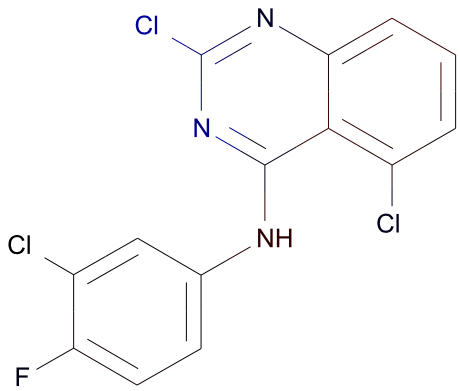
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1551584806	 [*]N[c]1:n:[*]:[c](:[*]):[c](:[cH]:[*]):n:1	0.464	46 out of 50

SCFP_12	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690
SCFP_12	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.375	934 out of 1123

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-760164567	 [*][c]1:[*]:[c](:[*]) :n:[c](NC):n:1	-1.35	0 out of 5
SCFP_12	-150051645	 [*]:n:[c](NCC):n:[*]	-0.998	0 out of 3
SCFP_12	-1301088181	 [*][c]1:[*]:[c](:[*]) :n:[c](NCC):n:1	-0.998	0 out of 3



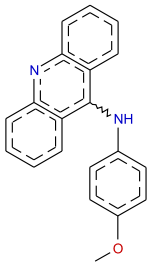
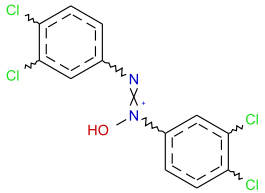
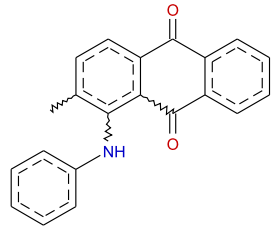
$C_{14}H_7Cl_3FN_3$
Molecular Weight: 342.58288
ALogP: 5.867
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.693
Enrichment: 1.24
Bayesian Score: -2.02
Mahalanobis Distance: 7.76
Mahalanobis Distance p-value: 0.997

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

Structural Similar Compounds

Name	61421-82-7	21232-47-3	4947-27-7
Structure			
Actual Endpoint	Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Mutagen	Non-Mutagen	Mutagen
Distance	0.545	0.547	0.570
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

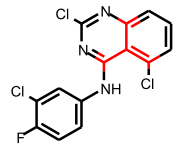
Model Applicability

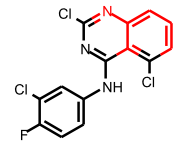
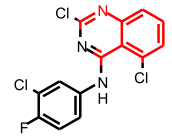
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

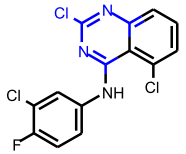
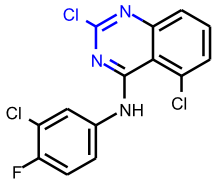
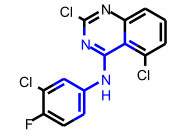
Feature Contribution

Top features for positive contribution

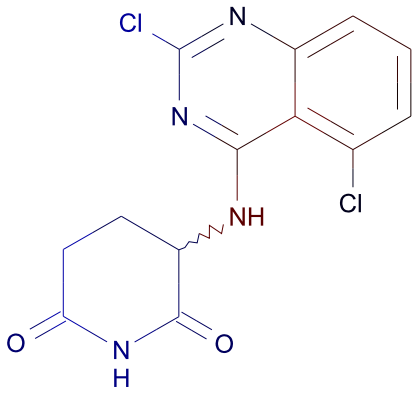
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263

SCFP_12	-1381862798	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.314	339 out of 433
SCFP_12	403834996	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.314	138 out of 176

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.887	2 out of 11
SCFP_12	1328855840	 [*]:[cH]:[c](N[c](:n: [*]):[c](:[*]):[*]): [cH]:[*]	-0.496	1 out of 4

TOPKAT_Ames_Mutagenicity



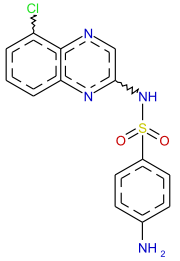
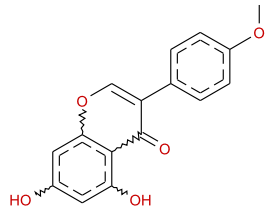
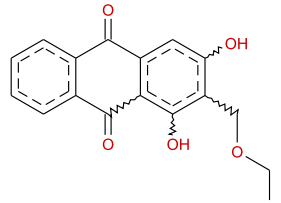
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen
Probability: 0.637
Enrichment: 1.14
Bayesian Score: -3.96
Mahalanobis Distance: 8.67
Mahalanobis Distance p-value: 0.923

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

Structural Similar Compounds

Name	97919-22-7	491-80-5	2-ETHOXYMETHYL-1;3-DIHYDROXYANTHRAQUINONE
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Distance	0.527	0.534	0.534
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	EMIC

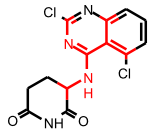
Model Applicability

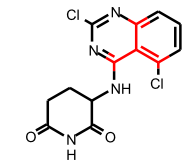
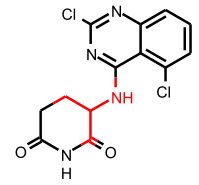
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

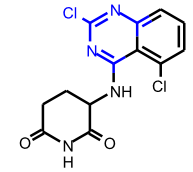
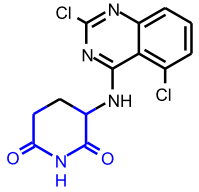
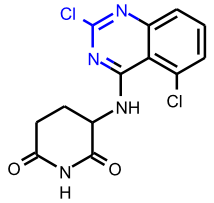
Feature Contribution

Top features for positive contribution

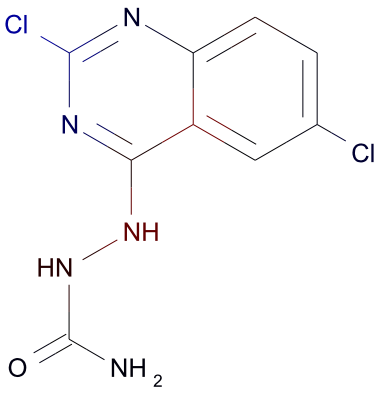
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1231050790	 [*]C([*])N[c]1:n:[c](:[*]):[*]:[c](:[*]):[c]:1:[c]([*]):[*]	0.388	3 out of 3

SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
SCFP_12	-1817806383	 [*]CC(N[*])C(=[*])[*]	0.337	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	395945879	 [*]C1[*]CC(=O)NC1=O	-1.19	0 out of 4
SCFP_12	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.887	2 out of 11

TOPKAT_Ames_Mutagenicity



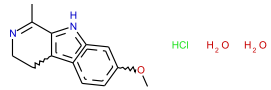
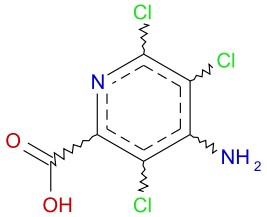
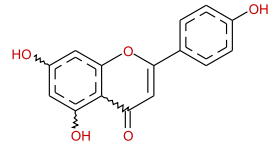
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09078
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Mutagen
Probability: 0.717
Enrichment: 1.28
Bayesian Score: -1.04
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.0482

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

Structural Similar Compounds

Name	6027-98-1	Picloram	520-36-5
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.560	0.561	0.564
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Mut. Res. 204: 17-115; 1988	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

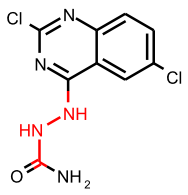
Model Applicability

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.

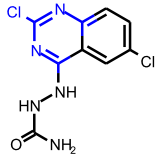
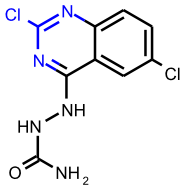
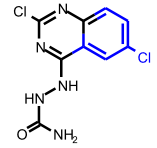
Feature Contribution

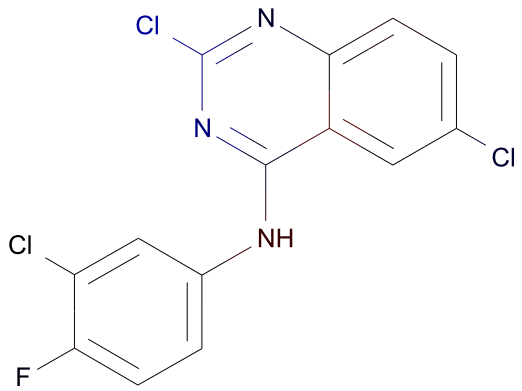
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1977951815	 [*]NNC(=[*])[*]	0.362	7 out of 8

SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
SCFP_12	-1381862798	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.314	339 out of 433

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.887	2 out of 11
SCFP_12	-116109291	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	-0.502	76 out of 221



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Non-Mutagen

Probability: 0.677

Enrichment: 1.21

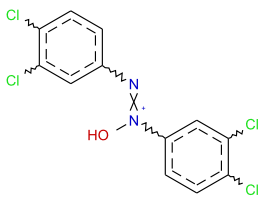
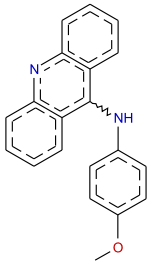
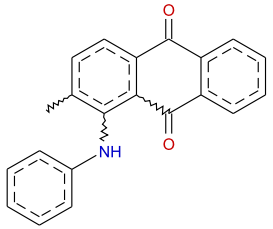
Bayesian Score: -2.64

Mahalanobis Distance: 7.76

Mahalanobis Distance p-value: 0.997

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

Structural Similar Compounds

Name	21232-47-3	61421-82-7	4947-27-7
Structure			
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Mutagen
Distance	0.548	0.557	0.575
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

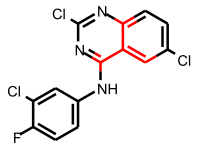
Model Applicability

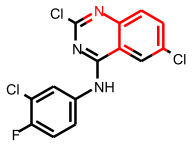
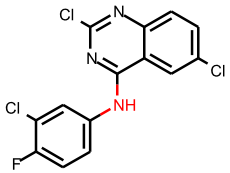
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

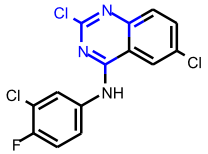
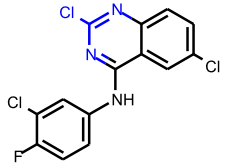
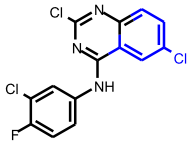
Feature Contribution

Top features for positive contribution

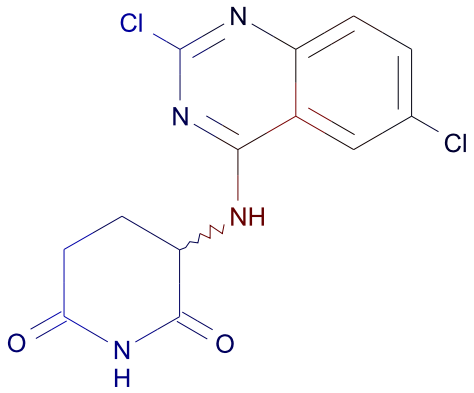
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263

SCFP_12	-1381862798	 [*]:n:[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.314	339 out of 433
SCFP_12	10	 [*]N[*]	0.306	1774 out of 2287

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.887	2 out of 11
SCFP_12	-116109291	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	-0.502	76 out of 221

TOPKAT_Ames_Mutagenicity



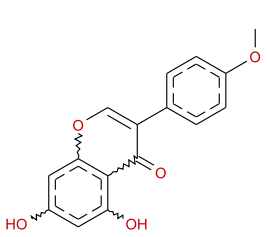
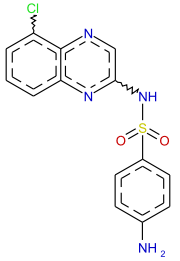
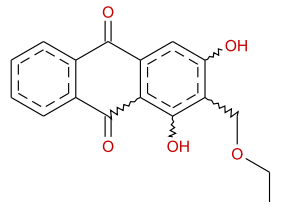
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen
Probability: 0.616
Enrichment: 1.1
Bayesian Score: -4.58
Mahalanobis Distance: 8.67
Mahalanobis Distance p-value: 0.923

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

Structural Similar Compounds

Name	491-80-5	97919-22-7	2-ETHOXYMETHYL-1;3-DIHYDROXYANTHRAQUINONE
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Distance	0.534	0.537	0.539
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	EMIC

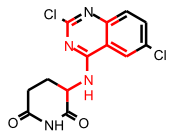
Model Applicability

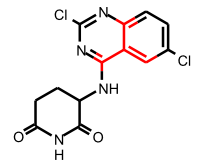
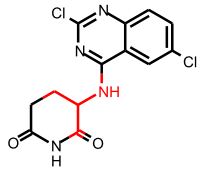
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

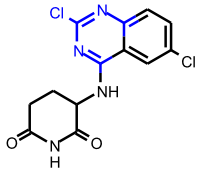
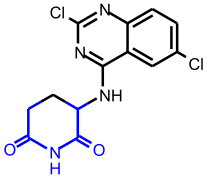
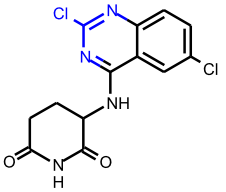
Feature Contribution

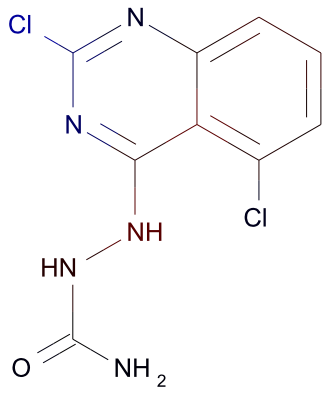
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1231050790	 [*]C([*])N[c]1:n:[c]([*]):[*]:[c](:[*]):[c]:1:[c]([*]):[*]	0.388	3 out of 3

SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
SCFP_12	-1817806383	 [*]CC(N[*])C(=[*])[*]	0.337	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	395945879	 [*]C1[*]CC(=O)NC1=O	-1.19	0 out of 4
SCFP_12	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.887	2 out of 11



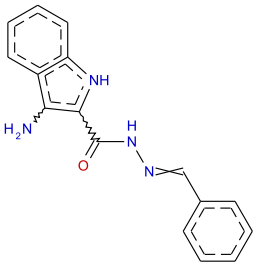
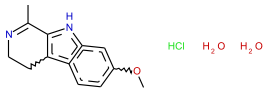
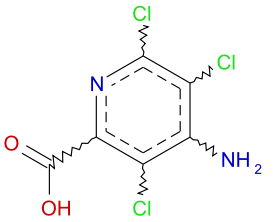
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Mutagen
Probability: 0.731
Enrichment: 1.31
Bayesian Score: -0.416
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.0482

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

Structural Similar Compounds

Name	2-Benzyliden-1-(3-aminoindol)-2-carbohydrazide	6027-98-1	Picloram
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Non-Mutagen
Distance	0.554	0.559	0.561
Reference	Mutagenesis 7(1):37-39; 1992	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Mut. Res. 204: 17-115; 1988

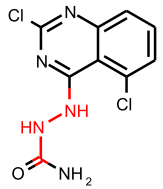
Model Applicability

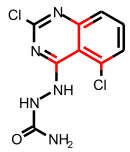
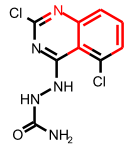
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

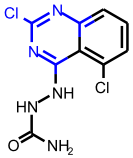
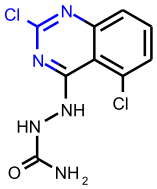
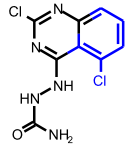
Feature Contribution

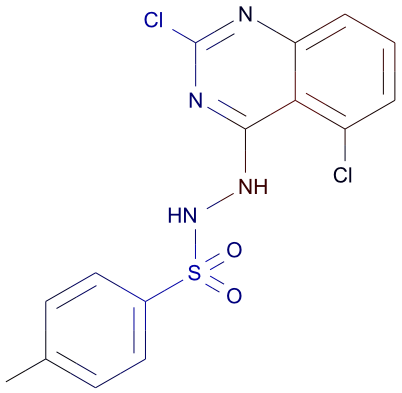
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1977951815	 [*]NNC(=[*])[*]	0.362	7 out of 8

SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263
SCFP_12	-1381862798	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.314	339 out of 433

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.887	2 out of 11
SCFP_12	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.474	101 out of 285



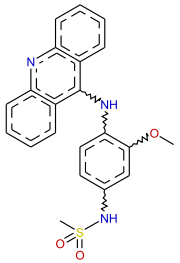
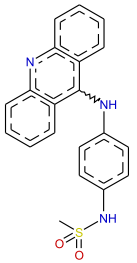
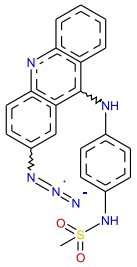
$C_{15}H_{12}Cl_2N_4O_2S$
Molecular Weight: 383.25237
ALogP: 4.443
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen
Probability: 0.623
Enrichment: 1.12
Bayesian Score: -4.38
Mahalanobis Distance: 9.26
Mahalanobis Distance p-value: 0.715

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

Structural Similar Compounds

Name	N/A	N/A	110021-94-8
Structure			
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.531	0.538	0.562
Reference	Mut. Res. 280:233-244; 1992	Mut. Res. 280:233-244; 1992	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

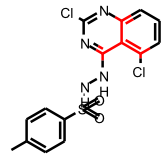
Model Applicability

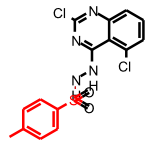
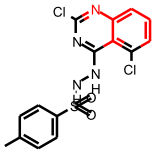
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

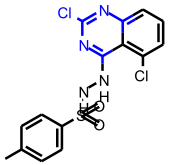
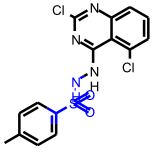
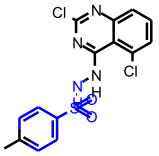
Feature Contribution

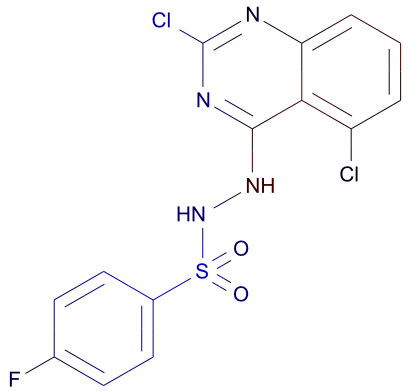
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263

SCFP_12	2054891299	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.322	13 out of 16
SCFP_12	-1381862798	 [*]:n:[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.314	339 out of 433

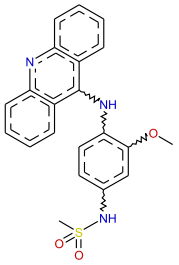
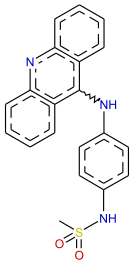
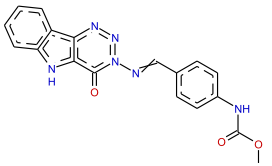
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	182991870	 [*]NS(=O)(=O)[c](:[*]):[*]	-1.15	8 out of 48
SCFP_12	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.01	6 out of 32



$C_{14}H_9Cl_2FN_4O_2S$
Molecular Weight: 387.21626
ALogP: 4.162
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Structural Similar Compounds

Name	N/A	N/A	3-(4'-Acetylaminobenzylidenamino)-5H-1;2;3-triazin-[5;4b]indol-4-one
Structure			
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.526	0.538	0.557
Reference	Mut. Res. 280:233-244; 1992	Mut. Res. 280:233-244; 1992	Mutagenesis 7(1):37-39; 1992

Model Prediction

Prediction: Non-Mutagen
Probability: 0.577
Enrichment: 1.03
Bayesian Score: -5.64
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.0424

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

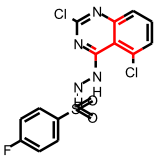
Model Applicability

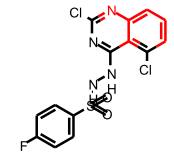
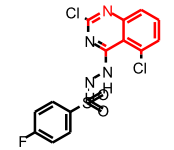
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

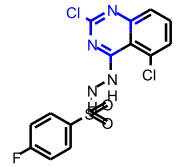
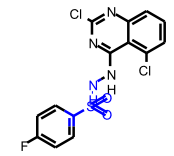
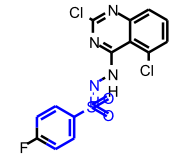
Feature Contribution

Top features for positive contribution

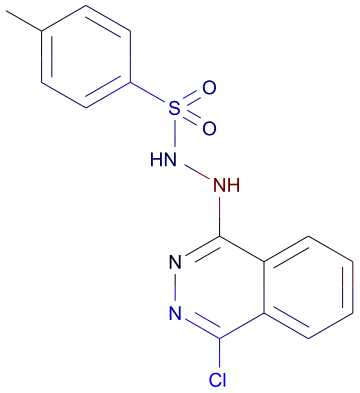
Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.36	1035 out of 1263

SCFP_12	-1381862798	 [*]:n:[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.314	339 out of 433
SCFP_12	403834996	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[cH]:[cH]:[cH]:[c]:2:n:1	0.314	138 out of 176

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-1.82	0 out of 9
SCFP_12	182991870	 [*]NS(=O)(=O)[c](:[*]):[*]	-1.15	8 out of 48
SCFP_12	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.01	6 out of 32

TOPKAT_Ames_Mutagenicity



$C_{15}H_{13}ClN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Mutagen

Probability: 0.61

Enrichment: 1.09

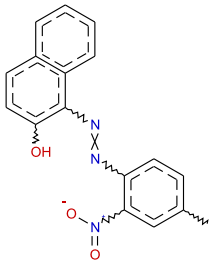
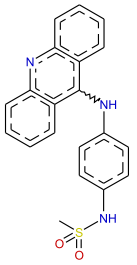
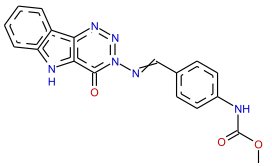
Bayesian Score: -4.74

Mahalanobis Distance: 9.19

Mahalanobis Distance p-value: 0.749

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

Structural Similar Compounds

Name	2425-85-6	N/A	3-(4'-Acetylaminobenzylidenamino)-5H-1;2;3-triazin-[5;4b]indol-4-one
Structure			
Actual Endpoint	Mutagen	Mutagen	Mutagen
Predicted Endpoint	Mutagen	Mutagen	Mutagen
Distance	0.537	0.548	0.549
Reference	Helma, C., Cramer, T., Kramer, S., and De Raedt, L., J. Chem. Inf. Comput. Sci., 2004, pp. 1402-1411	Mut. Res. 280:233-244; 1992	Mutagenesis 7(1):37-39; 1992

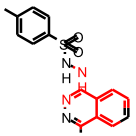
Model Applicability

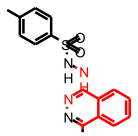
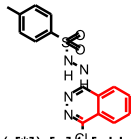
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

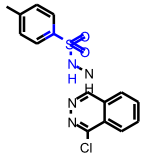
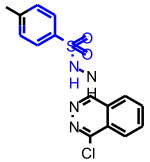
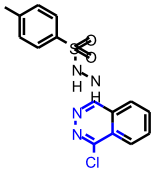
Feature Contribution

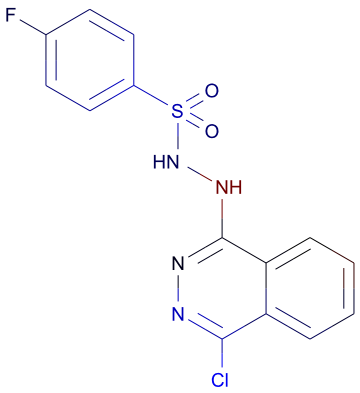
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1038366601	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4

SCFP_12	2097816882	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[cH]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4
SCFP_12	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.377	575 out of 690

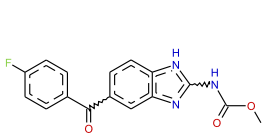
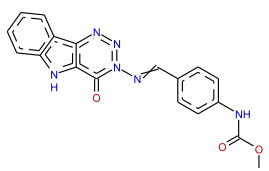
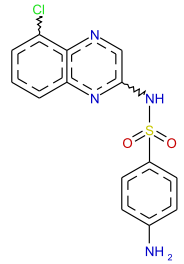
Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	182991870	 [*]NS(=O)(=O)[c](:[*]):[*]	-1.15	8 out of 48
SCFP_12	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.01	6 out of 32
SCFP_12	-2090779860	 [*][c]1:[*]:[c](:[*]):[c](Cl):n:n:1	-0.762	0 out of 2



$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Structural Similar Compounds

Name	31430-15-6	3-(4'-Acetylaminobenzylidenamino)-5H-1;2;3-triazin-[5;4b]indol-4-one	97919-22-7
Structure			
Actual Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Distance	0.550	0.560	0.561
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Mutagenesis 7(1):37-39; 1992	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

Model Prediction

Prediction: Non-Mutagen
Probability: 0.56
Enrichment: 1
Bayesian Score: -6.09
Mahalanobis Distance: 10.9
Mahalanobis Distance p-value: 0.0495

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

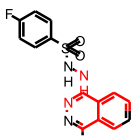
Model Applicability

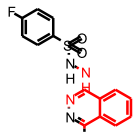
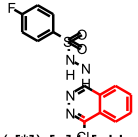
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

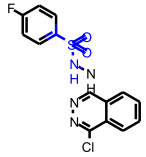
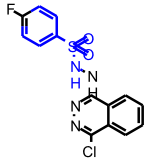
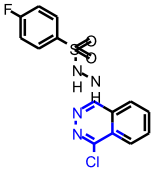
Feature Contribution

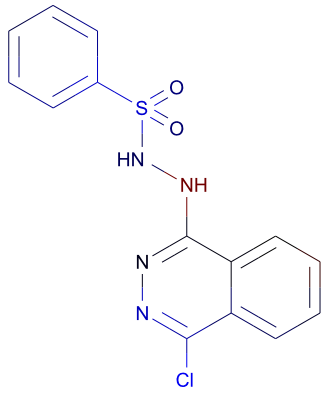
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	-1038366601	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4

SCFP_12	2097816882	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[cH]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4
SCFP_12	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.377	575 out of 690

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	182991870	 [*]NS(=O)(=O)[c](:[*]):[*]	-1.15	8 out of 48
SCFP_12	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.01	6 out of 32
SCFP_12	-2090779860	 [*][c]1:[*]:[c](:[*]):[c](Cl):n:n:1	-0.762	0 out of 2



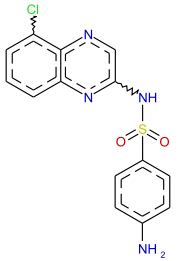
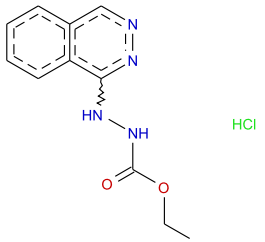
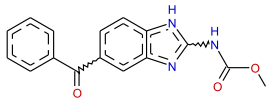
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Mutagen
Probability: 0.587
Enrichment: 1.05
Bayesian Score: -5.39
Mahalanobis Distance: 9.04
Mahalanobis Distance p-value: 0.812

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

Structural Similar Compounds

Name	97919-22-7	3778-76-5	31431-39-7
Structure			
Actual Endpoint	Non-Mutagen	Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Mutagen	Non-Mutagen
Distance	0.534	0.536	0.562
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

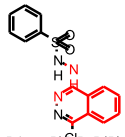
Model Applicability

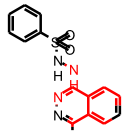
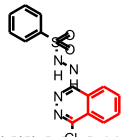
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

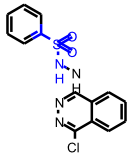
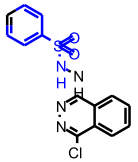
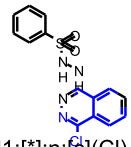
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	2097816882	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[cH]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4

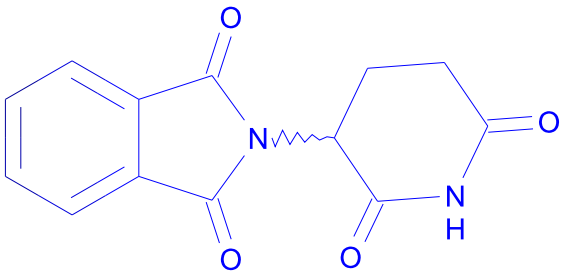
SCFP_12	-1038366601	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.42	4 out of 4
SCFP_12	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c] :1:[*]	0.377	575 out of 690

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	182991870	 [*]NS(=O)(=O)[c](:[*]):[*]	-1.15	8 out of 48
SCFP_12	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH] :1	-1.01	6 out of 32
SCFP_12	-1703874785	 [*][c]1:[*]:n:[c](Cl) :[c]2:[cH]:[cH]:[*]: [cH]:[c]:1:2	-0.762	0 out of 2

Thalidomide

TOPKAT_Ames_Mutagenicity



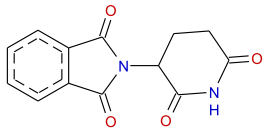
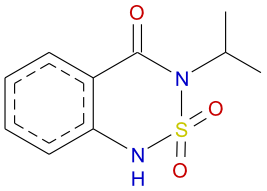
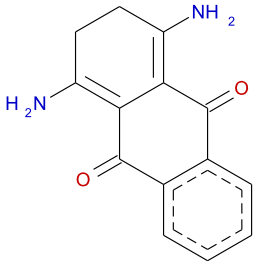
$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

Model Prediction

Prediction: Non-Mutagen
Probability: 0.188
Enrichment: 0.338
Bayesian Score: -14.5
Mahalanobis Distance: 7.7
Mahalanobis Distance p-value: 0.998

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

Structural Similar Compounds

Name	50-35-1	25057-89-0	81-63-0
Structure			
Actual Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Predicted Endpoint	Non-Mutagen	Non-Mutagen	Mutagen
Distance	0.000	0.509	0.539
Reference	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320	Kazius et. al., J. Med. Chem. (2005) 48, 312-320

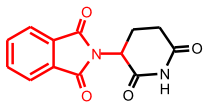
Model Applicability

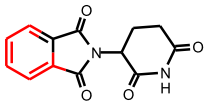
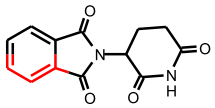
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

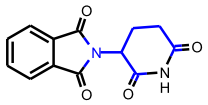
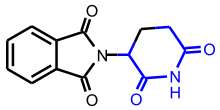
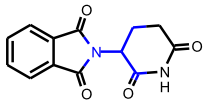
Feature Contribution

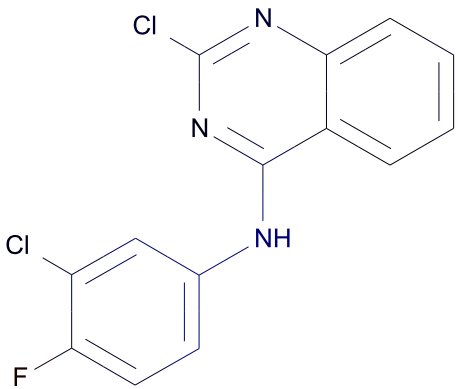
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	280592283	 [*]C([*])N1C(=O)[c]2:[cH]:[cH]:[cH]:[cH]:[c]:2C1=O	0.197	3 out of 4

SCFP_12	-1379591900	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0.108	1480 out of 2326
SCFP_12	-496409612	 [*]:[cH]:[cH]:[cH]:[*]	0.0771	2616 out of 4239

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Mutagen in training set
SCFP_12	1947881227	 [*]N([*])C1CCC(=[*])[*]C1=[*]	-1.9	0 out of 10
SCFP_12	395945879	 [*]C1[*]CC(=O)NC1=O	-1.19	0 out of 4
SCFP_12	-1946889102	 [*]CC(N([*])[*])C(=[*])[*]	-0.69	25 out of 89



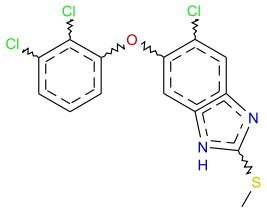
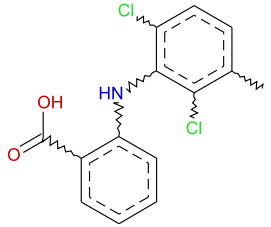
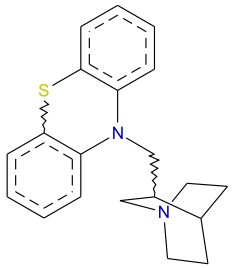
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.425
Enrichment: 0.809
Bayesian Score: -3.55
Mahalanobis Distance: 8.91
Mahalanobis Distance p-value: 0.204

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

Structural Similar Compounds

Name	Triclabendazole	Meclofenamate Sodium (Free acid form)	Mequitazine
Structure			
Actual Endpoint	Toxic	Non-Toxic	Toxic
Predicted Endpoint	Toxic	Non-Toxic	Toxic
Distance	0.587	0.595	0.621
Reference	Toxicology 43(3):283-287; 1987	Fundam Appl Toxicol 5:665-671; 1985	Oyo Yakuri 21:881-892; 1981

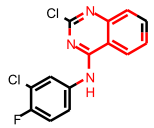
Model Applicability

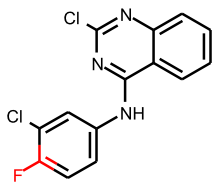
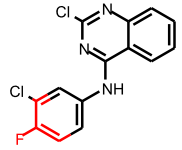
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

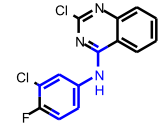
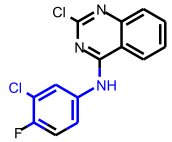
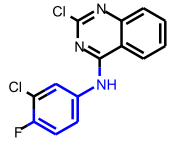
Feature Contribution

Top features for positive contribution

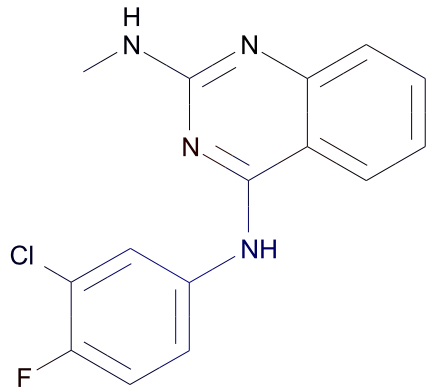
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	129215346	 [*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2

SCFP_6	-1794884847	 [*]:[c](:[*])F	0.202	6 out of 9
SCFP_6	-730654023	 [*][c](:[*]):[c](F):[cH]:[*]	0.202	6 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	 [*]N[c]1:[cH]:[*]:[c]([*]):[c](Cl):[cH]:1	-0.594	1 out of 5
SCFP_6	-1380909229	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.449	6 out of 19

TOPKAT_Developmental_Toxicity_Potential



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Toxic

Probability: 0.443

Enrichment: 0.842

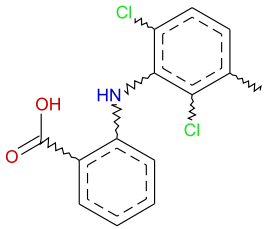
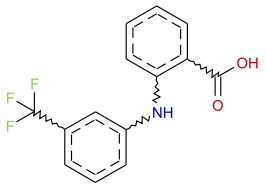
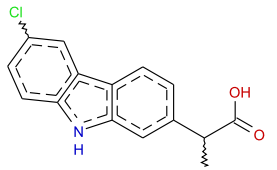
Bayesian Score: -2.99

Mahalanobis Distance: 9.37

Mahalanobis Distance p-value: 0.0899

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

Structural Similar Compounds

Name	Meclofenamate Sodium (Free acid form)	Flufenamic Acid	Carprofen
Structure			
Actual Endpoint	Non-Toxic	Non-Toxic	Non-Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Non-Toxic
Distance	0.505	0.557	0.589
Reference	Fundam Appl Toxicol 5:665-671; 1985	Kiso to Rinsho 13:3302-3313; 1979	Toxicol Appl Pharmacol 56(3):376-82; 1980

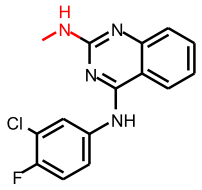
Model Applicability

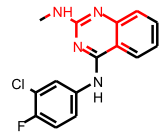
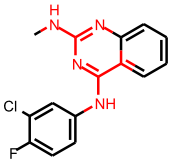
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

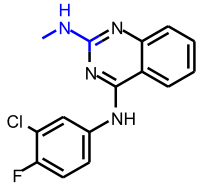
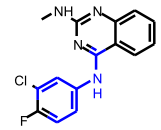
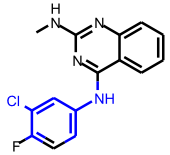
Feature Contribution

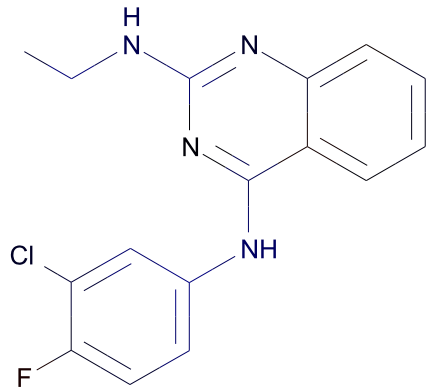
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	136418580	 [*]NC	0.381	2 out of 2

SCFP_6	-1551584806	 [*]N[c]1:n:[*]:[c](:[*]):[c](:[cH]:[*]):n:1	0.381	2 out of 2
SCFP_6	-1545804258	 [*]N[c]1:n:[*]:[c](:[*]):[c](N[*]):n:1	0.381	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	 [*]:[c](:[*])NC	-0.945	0 out of 3
SCFP_6	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	 [*]N[c]1:[cH]:[*]:[c]([*]):[c](Cl):[cH]:1	-0.594	1 out of 5



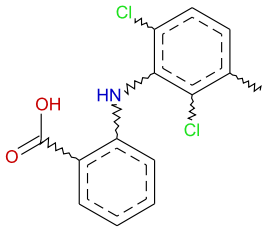
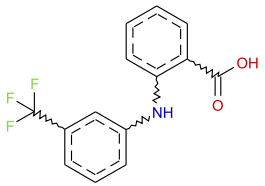
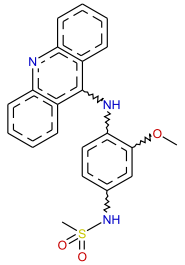
$C_{16}H_{14}ClFN_4$
Molecular Weight: 316.76056
ALogP: 5.025
Rotatable Bonds: 4
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.395
Enrichment: 0.751
Bayesian Score: -4.59
Mahalanobis Distance: 10.1
Mahalanobis Distance p-value: 0.0147

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

Structural Similar Compounds

Name	Meclofenamate Sodium (Free acid form)	Flufenamic Acid	Amsacrine
Structure			
Actual Endpoint	Non-Toxic	Non-Toxic	Toxic
Predicted Endpoint	Non-Toxic	Non-Toxic	Toxic
Distance	0.534	0.575	0.602
Reference	Fundam Appl Toxicol 5:665-671; 1985	Kiso to Rinsho 13:3302-3313; 1979	Fundam Appl Toxicol 7(2):214-20; 1986

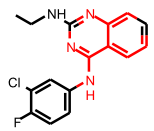
Model Applicability

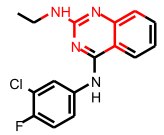
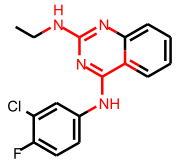
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

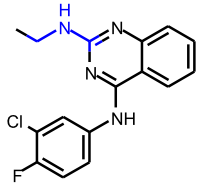
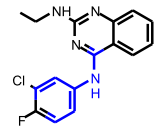
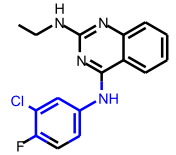
Feature Contribution

Top features for positive contribution

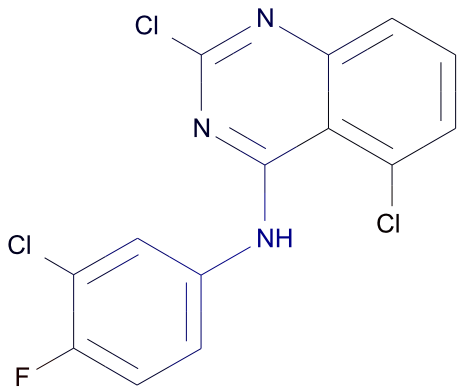
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	129215346	 [*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2

SCFP_6	-1551584806	 [*]N[c]1:n:[*]:[c](:[*]):[c](:[cH]:[*]):n:1	0.381	2 out of 2
SCFP_6	-1545804258	 [*]N[c]1:n:[*]:[c](:[*]):[c](N[*]):n:1	0.381	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	 [*]:[c](:[*])NC	-0.945	0 out of 3
SCFP_6	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	 [*]N[c]1:[cH]:[*]:[c]([*]):[c](Cl):[cH]:1	-0.594	1 out of 5

TOPKAT_Developmental_Toxicity_Potential



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Non-Toxic

Probability: 0.411

Enrichment: 0.781

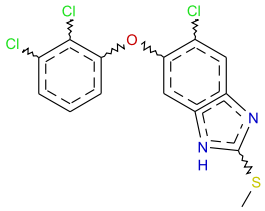
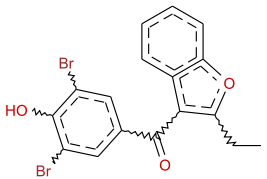
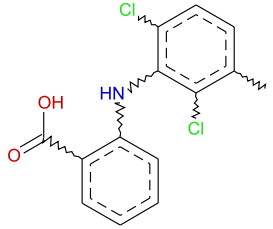
Bayesian Score: -4.05

Mahalanobis Distance: 9.09

Mahalanobis Distance p-value: 0.152

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

Structural Similar Compounds

Name	Triclabendazole	Benzbromarone	Meclofenamate Sodium (Free acid form)
Structure			
Actual Endpoint	Toxic	Toxic	Non-Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.555	0.624	0.626
Reference	Toxicology 43(3):283-287; 1987	Shinryo to Shinaku 16:1521-1545; 1979	Fundam Appl Toxicol 5:665-671; 1985

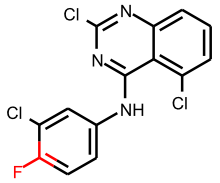
Model Applicability

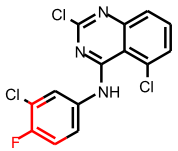
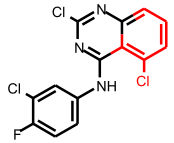
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

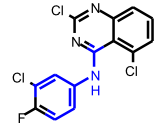
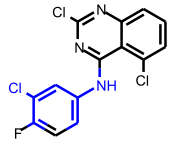
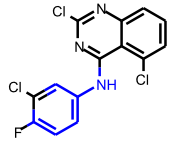
Feature Contribution

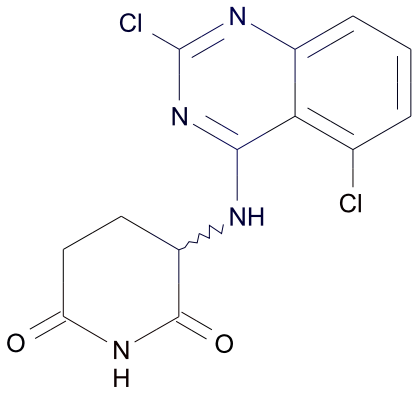
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1794884847	 [*]:[c](:[*])F	0.202	6 out of 9

SCFP_6	-730654023	 [*][c](:[*]):[c](F):[cH]:[*]	0.202	6 out of 9
SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	0.137	17 out of 28

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	 [*]N[c]1:[cH]:[*]:[c]([*]):[c](Cl):[cH]:1	-0.594	1 out of 5
SCFP_6	-1380909229	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.449	6 out of 19



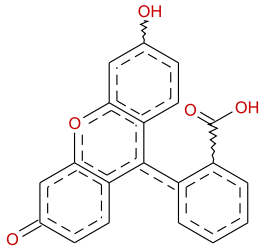
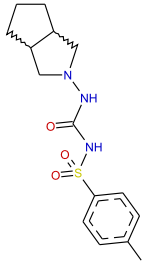
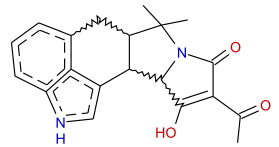
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.452
Enrichment: 0.859
Bayesian Score: -2.71
Mahalanobis Distance: 10.3
Mahalanobis Distance p-value: 0.0103

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

Structural Similar Compounds

Name	D&C Yellow 8	Sulfonylurea Gliclazide	Cyclopiazonic Acid
Structure			
Actual Endpoint	Non-Toxic	Toxic	Non-Toxic
Predicted Endpoint	Non-Toxic	Toxic	Non-Toxic
Distance	0.528	0.562	0.611
Reference	Food Chem Toxicol 24:819-823; 1986	Yakuri to Chiryo 9:3551-3571; 1981	J Toxicol Environ Health 14:585-594; 1984

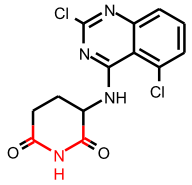
Model Applicability

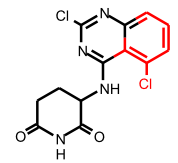
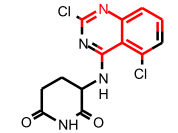
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

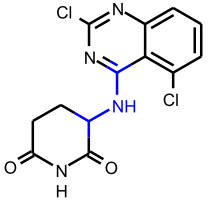
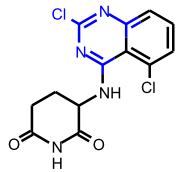
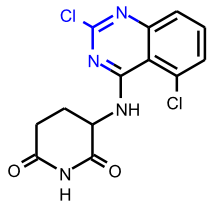
Feature Contribution

Top features for positive contribution

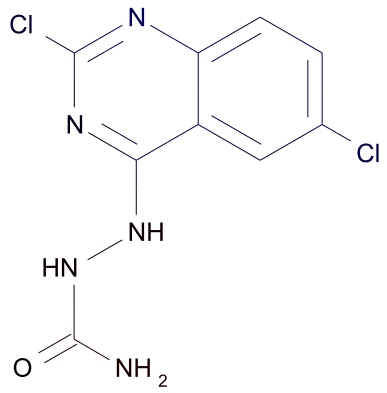
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1631785938	 [*]C(=[*])NC(=[*])[*]	0.255	3 out of 4

SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	0.137	17 out of 28
SCFP_6	622342378	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.114	6 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	 [*]:[c](:[*])NC	-0.945	0 out of 3
SCFP_6	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-0.422	0 out of 1
SCFP_6	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.422	0 out of 1

TOPKAT_Developmental_Toxicity_Potential



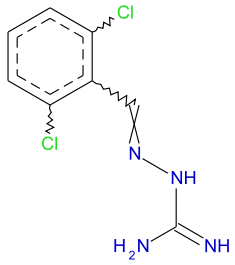
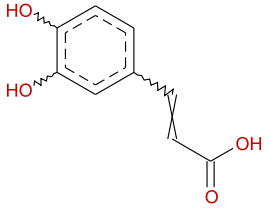
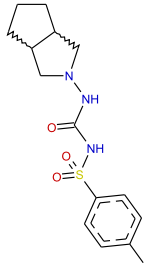
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09078
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Toxic
Probability: 0.465
Enrichment: 0.885
Bayesian Score: -2.3
Mahalanobis Distance: 8.39
Mahalanobis Distance p-value: 0.42

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

Structural Similar Compounds

Name	Guanabenz	Caffeic Acid	Sulfonylurea Gliclazide
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.609	0.624	0.666
Reference	Journal of Toxic Sciences 11:107-119; 1982	Toxicol Appl Pharmacol 36(2):227-37; 1976	Yakuri to Chiryo 9:3551-3571; 1981

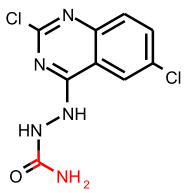
Model Applicability

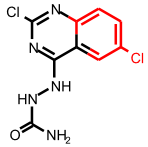
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

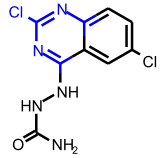
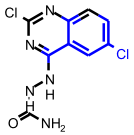
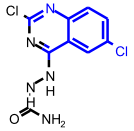
Feature Contribution

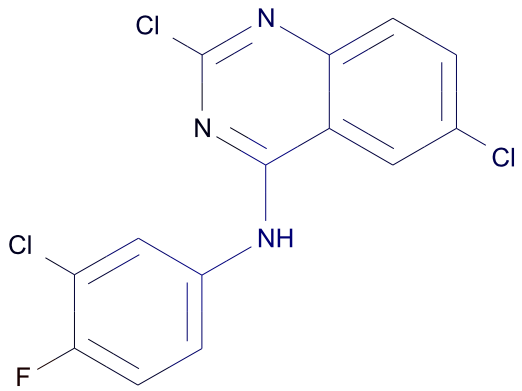
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1357949052	 [*]C(=[*])N	0.453	8 out of 9

SCFP_6	129215346	 [*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2
SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	0.137	17 out of 28

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	60071962	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	-0.422	0 out of 1
SCFP_6	1808332909	 [*][c](:[*]):[c]1:[cH]:[c](Cl):[cH]:[*]:[c]:1:[*]	-0.422	0 out of 1
SCFP_6	-182622233	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[c](Cl):[c H]:[cH]:[c]:2:n:1	-0.422	0 out of 1



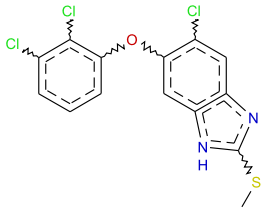
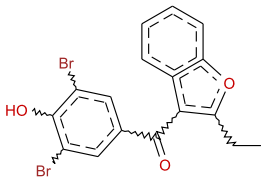
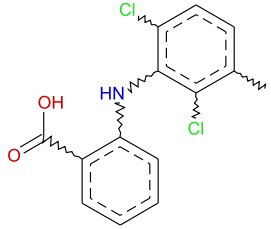
$C_{14}H_7Cl_3FN_3$
Molecular Weight: 342.58288
ALogP: 5.867
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.387
Enrichment: 0.736
Bayesian Score: -4.88
Mahalanobis Distance: 8.98
Mahalanobis Distance p-value: 0.184

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

Structural Similar Compounds

Name	Triclabendazole	Benzbromarone	Meclofenamate Sodium (Free acid form)
Structure			
Actual Endpoint	Toxic	Toxic	Non-Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.561	0.630	0.633
Reference	Toxicology 43(3):283-287; 1987	Shinryo to Shinaku 16:1521-1545; 1979	Fundam Appl Toxicol 5:665-671; 1985

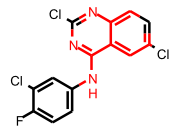
Model Applicability

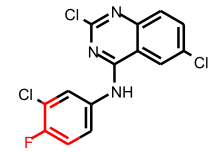
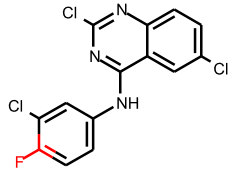
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

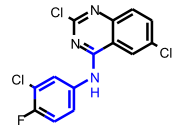
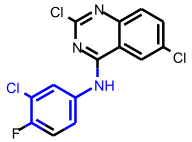
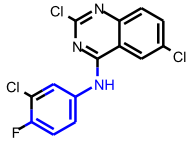
Feature Contribution

Top features for positive contribution

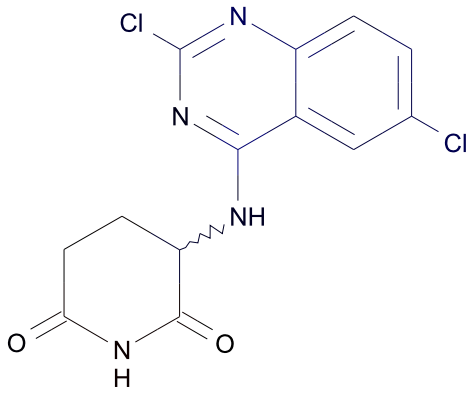
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	129215346	 [*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2

SCFP_6	-730654023	 [*][c](:[*]):[c](F):[cH]:[*]	0.202	6 out of 9
SCFP_6	-1794884847	 [*]:[c](:[*])F	0.202	6 out of 9

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.718	0 out of 2
SCFP_6	1807097289	 [*]N[c]1:[cH]:[*]:[c]([*]):[c](Cl):[cH]:1	-0.594	1 out of 5
SCFP_6	-1380909229	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.449	6 out of 19

TOPKAT_Developmental_Toxicity_Potential



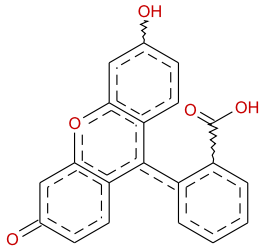
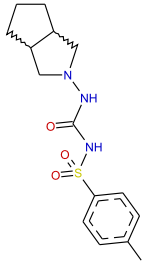
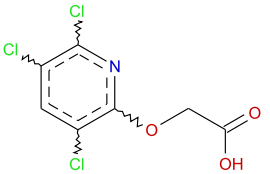
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.425
Enrichment: 0.809
Bayesian Score: -3.55
Mahalanobis Distance: 9.55
Mahalanobis Distance p-value: 0.0613

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

Structural Similar Compounds

Name	D&C Yellow 8	Sulfonylurea Gliclazide	Triclopyr
Structure			
Actual Endpoint	Non-Toxic	Toxic	Non-Toxic
Predicted Endpoint	Non-Toxic	Toxic	Non-Toxic
Distance	0.533	0.562	0.618
Reference	Food Chem Toxicol 24:819-823; 1986	Yakuri to Chiryo 9:3551-3571; 1981	Fundam Appl Toxicol 4:872-882; 1984

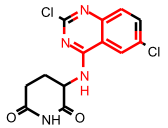
Model Applicability

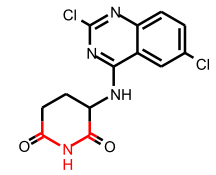
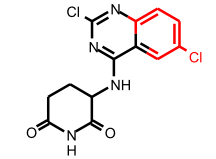
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

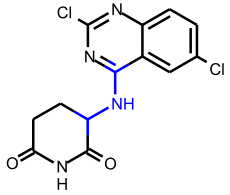
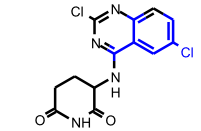
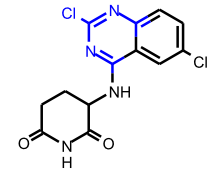
Feature Contribution

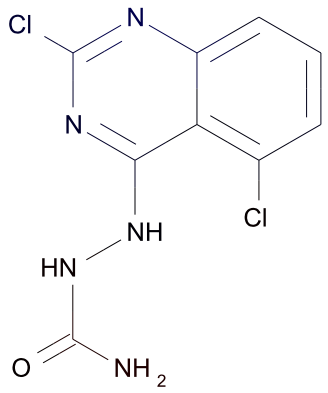
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	129215346	 [*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	0.381	2 out of 2

SCFP_6	1631785938	 [*]C(=[*])NC(=[*])[*]	0.255	3 out of 4
SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	0.137	17 out of 28

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	18117904	 [*]:[c](:[*])NC	-0.945	0 out of 3
SCFP_6	1808332909	 [*][c](:[*]):[c]1:[cH]:[c](Cl):[cH]:[*]:[c]:1:[*]	-0.422	0 out of 1
SCFP_6	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-0.422	0 out of 1



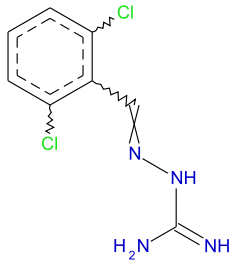
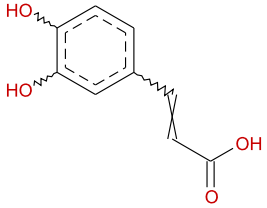
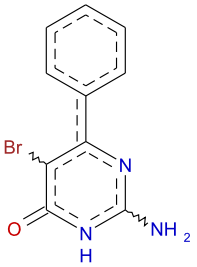
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Toxic
Probability: 0.494
Enrichment: 0.94
Bayesian Score: -1.46
Mahalanobis Distance: 9.07
Mahalanobis Distance p-value: 0.158

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

Structural Similar Compounds

Name	Guanabenz	Caffeic Acid	Bropirimine
Structure			
Actual Endpoint	Toxic	Toxic	Non-Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.600	0.624	0.661
Reference	Journal of Toxic Sciences 11:107-119; 1982	Toxicol Appl Pharmacol 36(2):227-37; 1976	Teratology 38(1):7-14; 1988

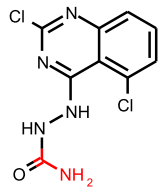
Model Applicability

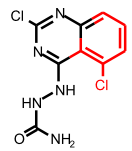
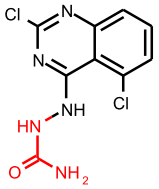
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

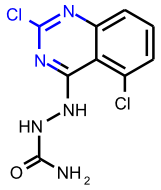
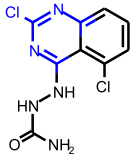
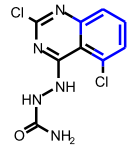
Feature Contribution

Top features for positive contribution

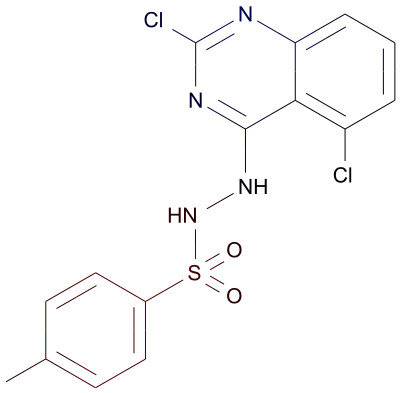
Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1357949052	 [*]C(=[*])N	0.453	8 out of 9

SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	0.137	17 out of 28
SCFP_6	1256786467	 [*]NC(=O)N	0.12	9 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.422	0 out of 1
SCFP_6	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-0.422	0 out of 1
SCFP_6	-1379591900	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84

TOPKAT_Developmental_Toxicity_Potential



$C_{15}H_{12}Cl_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Toxic

Probability: 0.525

Enrichment: 0.998

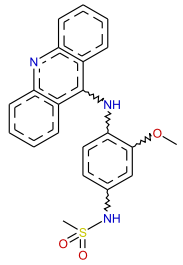
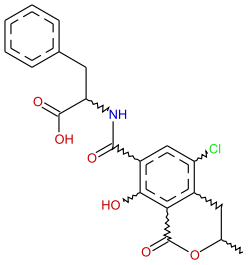
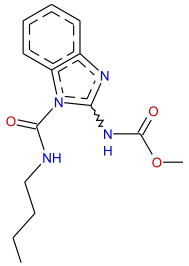
Bayesian Score: -0.61

Mahalanobis Distance: 11.8

Mahalanobis Distance p-value: 6.09e-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	Amsacrine	Ochratoxin a	Benomyl
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.515	0.642	0.651
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	J Toxicol Environ Health 17:405-417; 1986

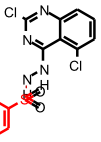
Model Applicability

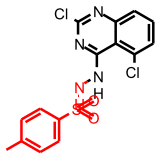
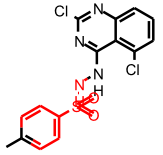
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

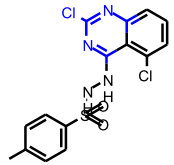
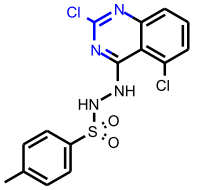
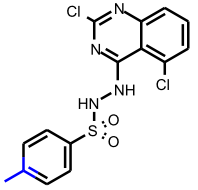
Feature Contribution

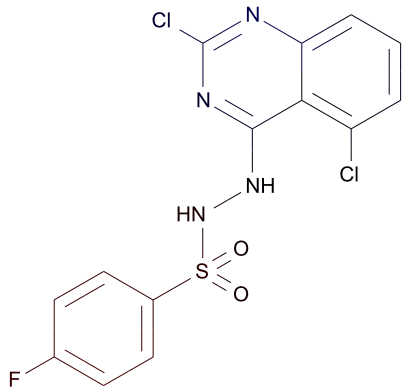
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	2054891299	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.271	1 out of 1

SCFP_6	1892882306	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.271	1 out of 1
SCFP_6	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[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	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-0.422	0 out of 1
SCFP_6	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.422	0 out of 1
SCFP_6	136686699	 [*]:[c](:[*])C	-0.316	7 out of 19



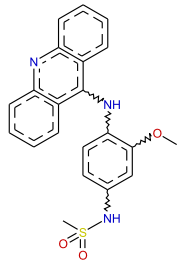
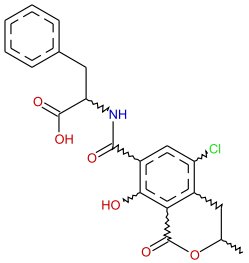
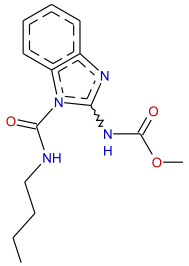
$C_{14}H_9Cl_2FN_4O_2S$
Molecular Weight: 387.21626
ALogP: 4.162
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Toxic
Probability: 0.545
Enrichment: 1.04
Bayesian Score: -0.06
Mahalanobis Distance: 12.1
Mahalanobis Distance p-value: 2.32e-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	Amsacrine	Ochratoxin a	Benomyl
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.515	0.641	0.644
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	Toxicol Appl Pharmacol 37(2):331-8; 1976	J Toxicol Environ Health 17:405-417; 1986

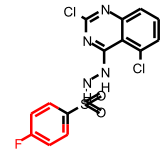
Model Applicability

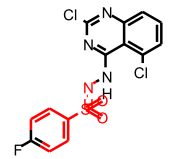
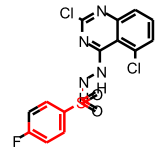
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

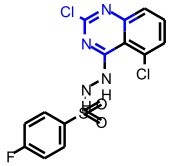
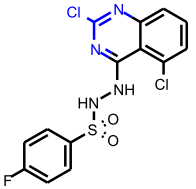
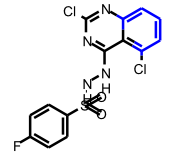
Feature Contribution

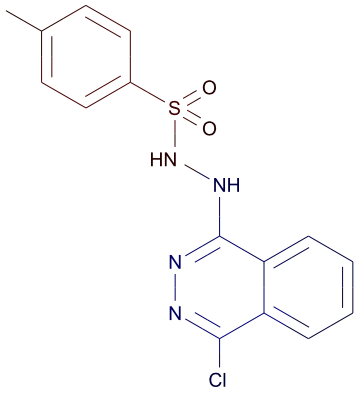
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-783770208	 F[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.322	4 out of 5

SCFP_6	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.271	1 out of 1
SCFP_6	-1380395165	 [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*])(=[*])[*]	0.255	3 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	60071962	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	-0.422	0 out of 1
SCFP_6	-1286592310	 [*]:n:[c](Cl):n:[*]	-0.422	0 out of 1
SCFP_6	-1379591900	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84



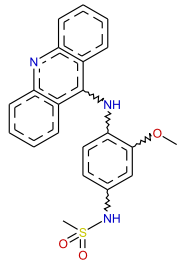
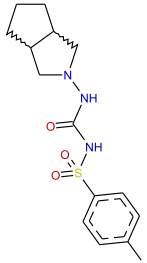
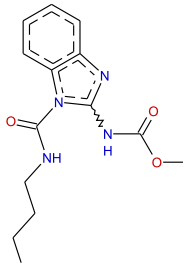
$C_{15}H_{13}ClN_4O_2S$
Molecular Weight: 348.80731
ALogP: 3.529
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.456
Enrichment: 0.867
Bayesian Score: -2.59
Mahalanobis Distance: 15.9
Mahalanobis Distance p-value: 2.43e-013

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

Structural Similar Compounds

Name	Amsacrine	Sulfonylurea Gliclazide	Benomyl
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.547	0.565	0.594
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	Yakuri to Chiryo 9:3551-3571; 1981	J Toxicol Environ Health 17:405-417; 1986

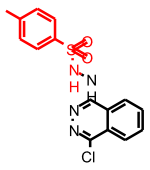
Model Applicability

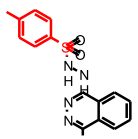
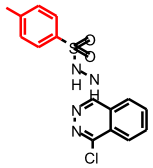
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

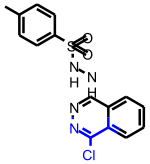
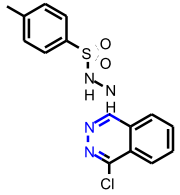
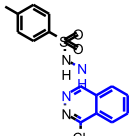
Feature Contribution

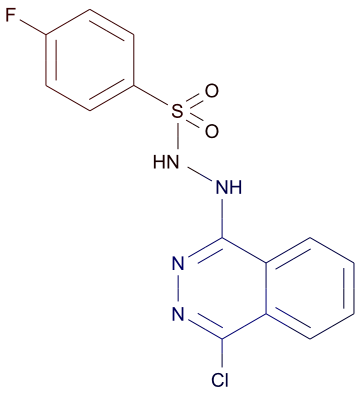
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1892882306	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.271	1 out of 1

SCFP_6	2054891299	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[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	-601362767	 [*]:n:[c](Cl):[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	 [*][c](:[*]):n:n:[*]	-0.448	5 out of 16
SCFP_6	2097816882	 [*]N[c]1:n:[*]:[c](:[*]):[c]2:[cH]:[cH]:[cH]:[cH]:[c]:1:2	-0.422	0 out of 1



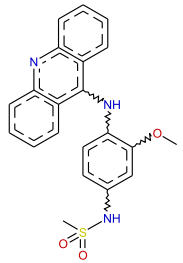
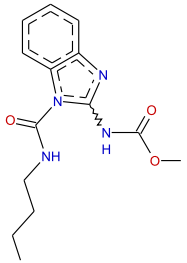
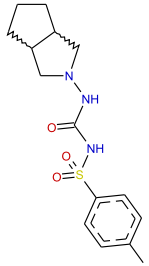
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.462
Enrichment: 0.879
Bayesian Score: -2.39
Mahalanobis Distance: 16
Mahalanobis Distance p-value: 1.37e-013

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

Structural Similar Compounds

Name	Amsacrine	Benomyl	Sulfonylurea Gliclazide
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.568	0.598	0.602
Reference	Fundam Appl Toxicol 7(2):214-20; 1986	J Toxicol Environ Health 17:405-417; 1986	Yakuri to Chiryo 9:3551-3571; 1981

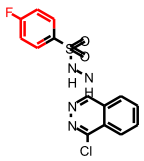
Model Applicability

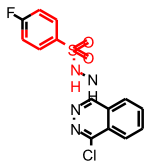
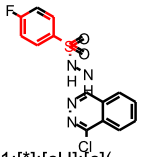
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

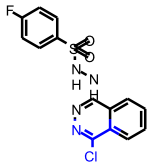
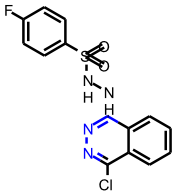
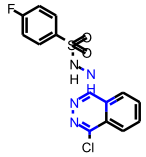
Feature Contribution

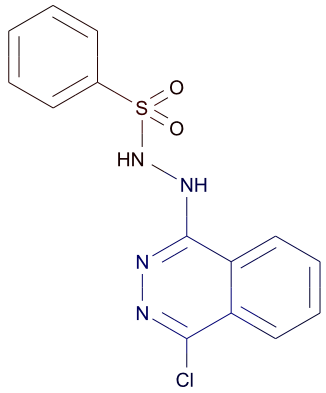
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-783770208	 F[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.322	4 out of 5

SCFP_6	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.271	1 out of 1
SCFP_6	-1380395165	 [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*])(=[*])[*]	0.255	3 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-601362767	 [*]:n:[c](Cl):[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	 [*][c](:[*]):n:n:[*]	-0.448	5 out of 16
SCFP_6	-705181820	 [*]N[c]1:n:n:[c]([*]):[*]:[c]:1:[*]	-0.422	0 out of 1



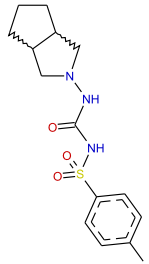
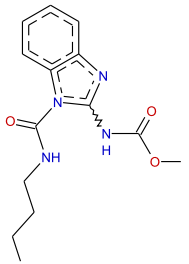
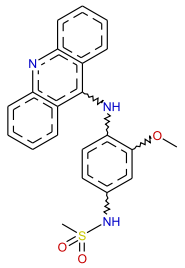
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Toxic
Probability: 0.426
Enrichment: 0.81
Bayesian Score: -3.52
Mahalanobis Distance: 15.7
Mahalanobis Distance p-value: 8.89e-013

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

Structural Similar Compounds

Name	Sulfonylurea Gliclazide	Benomyl	Amsacrine
Structure			
Actual Endpoint	Toxic	Toxic	Toxic
Predicted Endpoint	Toxic	Toxic	Toxic
Distance	0.584	0.586	0.592
Reference	Yakuri to Chiryo 9:3551-3571; 1981	J Toxicol Environ Health 17:405-417; 1986	Fundam Appl Toxicol 7(2):214-20; 1986

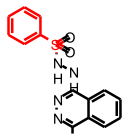
Model Applicability

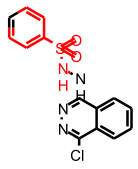
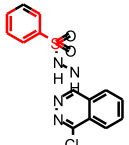
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

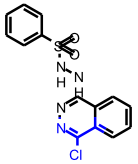
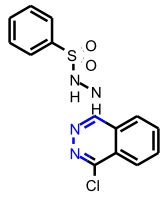
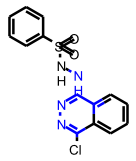
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1655803608	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.271	1 out of 1

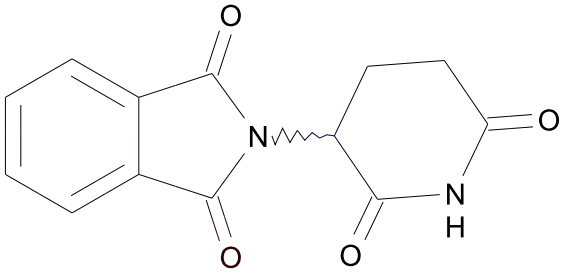
SCFP_6	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.271	1 out of 1
SCFP_6	-1380395165	 [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*])(=[*])[*]	0.255	3 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	-601362767	 [*]:n:[c](Cl):[c](:[*]):[*]	-0.945	0 out of 3
SCFP_6	149212520	 [*][c](:[*]):n:n:[*]	-0.448	5 out of 16
SCFP_6	-705181820	 [*]N[c]1:n:n:[c]([*]):[*]:[c]:1:[*]	-0.422	0 out of 1

Thalidomide

TOPKAT_Developmental_Toxicity_Potential



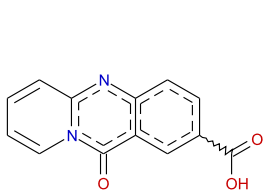
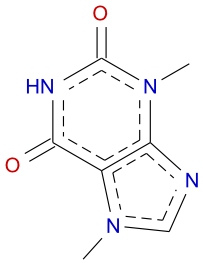
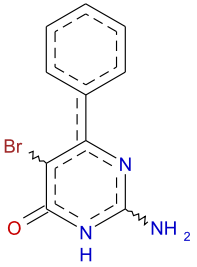
$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

Model Prediction

Prediction: Non-Toxic
Probability: 0.536
Enrichment: 1.02
Bayesian Score: -0.317
Mahalanobis Distance: 7.73
Mahalanobis Distance p-value: 0.731

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

Structural Similar Compounds

Name	11-oxo-11H-Pyrido(2;1-b)quinazoline-2-carboxylic Acid	Theobromine	Bropirimine
Structure			
Actual Endpoint	Toxic	Toxic	Non-Toxic
Predicted Endpoint	Toxic	Toxic	Non-Toxic
Distance	0.539	0.577	0.598
Reference	Teratology 38(4):351-67; 1988	Food Chem Toxicol 24(5):375-82; 1986	Teratology 38(1):7-14; 1988

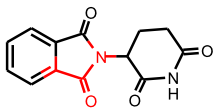
Model Applicability

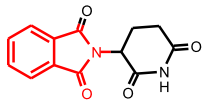
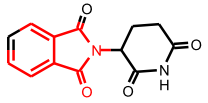
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

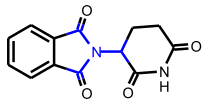
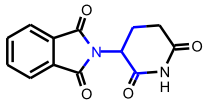
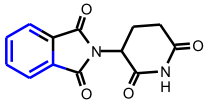
Feature Contribution

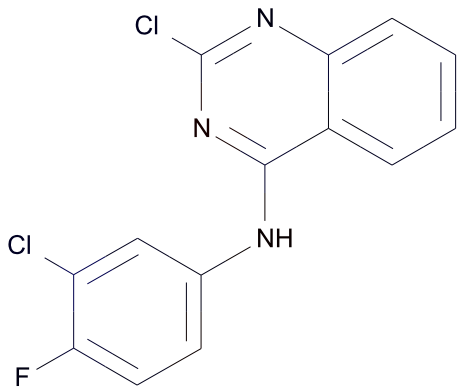
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	1257084377	 [*]N1[*][*]:[c](:[*])C1=O	0.362	14 out of 18

SCFP_6	1725056539	 [*]N1C(=[*])[c]2:[cH] :[cH]:[cH]:[cH]:[c]: 2C1=O	0.271	1 out of 1
SCFP_6	2049973914	 [*]N1C(=[*])[c]2:[cH] :[*]:[cH]:[cH]:[c]:2 C1=O	0.271	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Toxic in training set
SCFP_6	399659969	 [*]C([*])N1C(=[*])[*] :[*]C1=[*]	-0.526	3 out of 11
SCFP_6	-1946889102	 [*]CC(N([*])[*])C(=[*])[*]	-0.512	4 out of 14
SCFP_6	-1379591900	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.282	33 out of 84



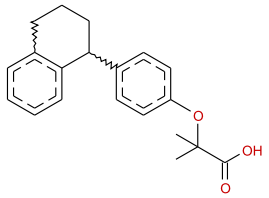
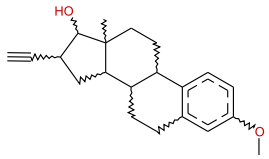
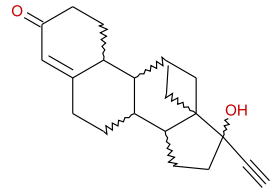
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.227
Enrichment: 0.707
Bayesian Score: -2.35
Mahalanobis Distance: 9.9
Mahalanobis Distance p-value: 0.491

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

Structural Similar Compounds

Name	Nafenopin	Mestranol	Levonorgestrel
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.602	0.603	0.606
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

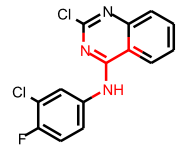
Model Applicability

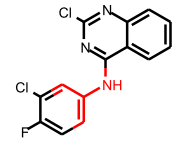
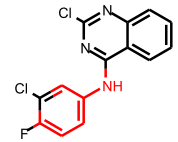
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

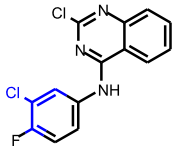
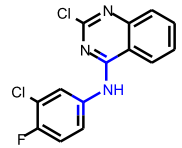
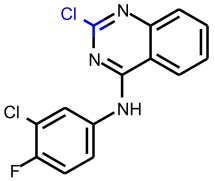
Feature Contribution

Top features for positive contribution

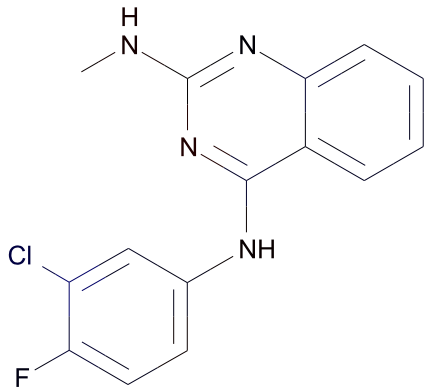
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

ECFP_6	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9
ECFP_6	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.234	3 out of 7

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	-1242906247	 [*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2
ECFP_6	99947387	 [*]:[c](:[*])Cl	-0.461	9 out of 48

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.228

Enrichment: 0.712

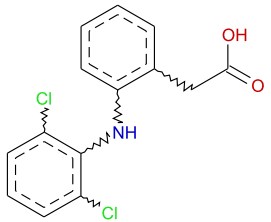
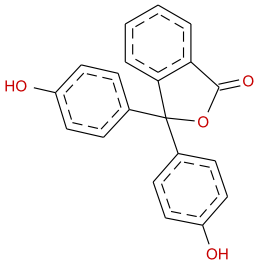
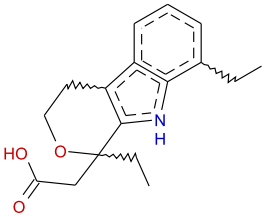
Bayesian Score: -2.2

Mahalanobis Distance: 10.3

Mahalanobis Distance p-value: 0.302

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

Structural Similar Compounds

Name	Diclofenac	Phenolphthalein	Etodolac
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.553	0.588	0.613
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

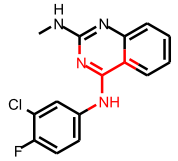
Model Applicability

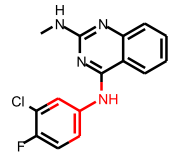
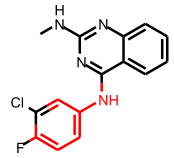
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1240852525: [*]:[c](:[*])NC

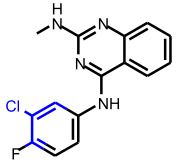
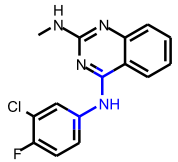
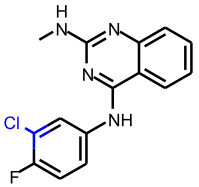
Feature Contribution

Top features for positive contribution

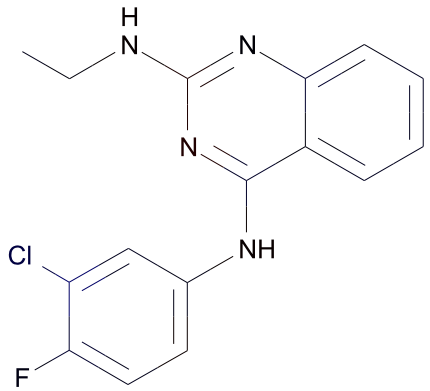
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

ECFP_6	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9
ECFP_6	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.234	3 out of 7

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	-1242906247	 [*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2
ECFP_6	99947387	 [*]:[c](:[*])Cl	-0.461	9 out of 48

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{16}H_{14}ClFN_4$

Molecular Weight: 316.76056

ALogP: 5.025

Rotatable Bonds: 4

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.221

Enrichment: 0.691

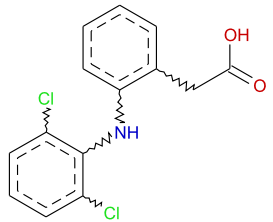
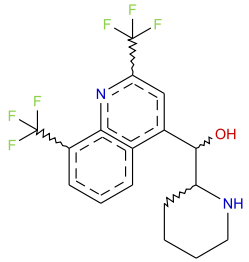
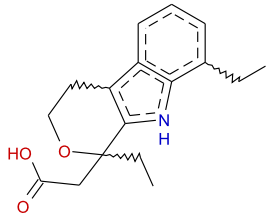
Bayesian Score: -2.84

Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.12

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

Structural Similar Compounds

Name	Diclofenac	Mefloquine	Etodolac
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.552	0.611	0.624
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

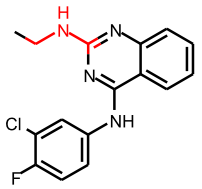
Model Applicability

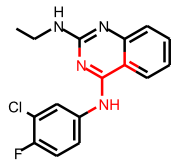
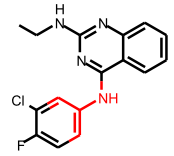
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC25 out of range. Value: -3.1285. Training min, max, SD, explained variance: -2.879, 2.6681, 1.105, 0.0124.

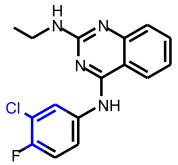
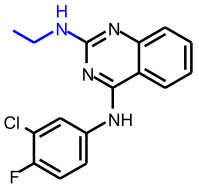
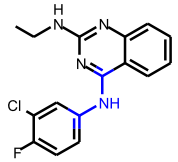
Feature Contribution

Top features for positive contribution

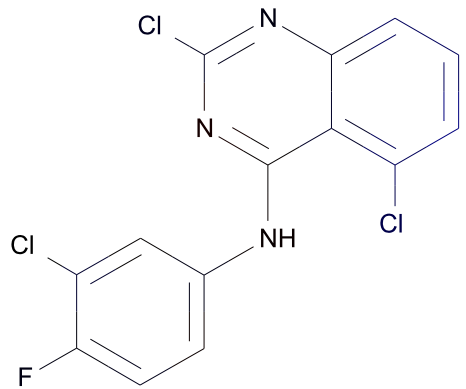
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	491100606	 [*]CN[c](:[*]):[*]	0.424	1 out of 1

ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1
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	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.669	3 out of 22
ECFP_6	-950223878	 [*]NCC	-0.482	0 out of 2
ECFP_6	-1242906247	 [*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.216

Enrichment: 0.673

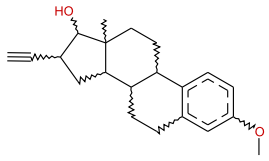
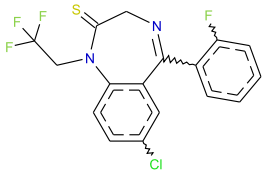
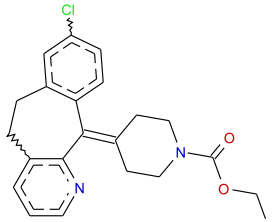
Bayesian Score: -3.49

Mahalanobis Distance: 10.5

Mahalanobis Distance p-value: 0.245

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

Structural Similar Compounds

Name	Mestranol	Quazepam	Loratidine
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.631	0.638	0.643
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

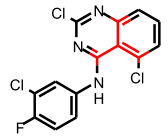
Model Applicability

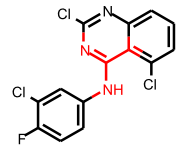
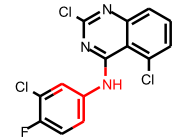
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

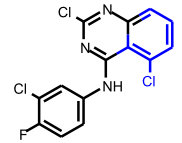
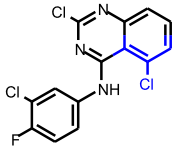
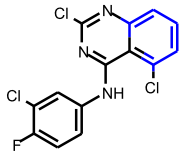
Feature Contribution

Top features for positive contribution

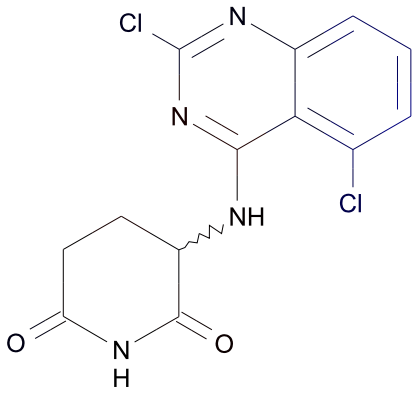
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.442	2 out of 3

ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1
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	1641317964	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.789	1 out of 11
ECFP_6	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.669	3 out of 22
ECFP_6	2007300961	 [*][c]1:[*]:[c](:[*]):[cH]:[cH]:[cH]:1	-0.652	5 out of 34

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227

Enrichment: 0.709

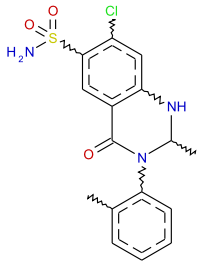
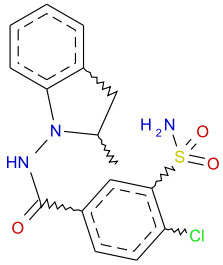
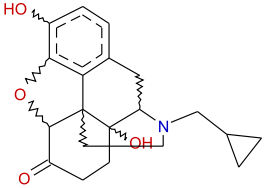
Bayesian Score: -2.28

Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.0156

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

Structural Similar Compounds

Name	Metolazone	Indapamide	Naltrexone
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.559	0.578	0.579
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

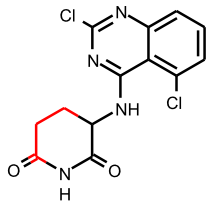
Model Applicability

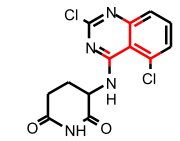
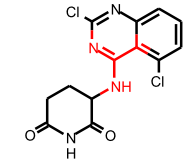
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

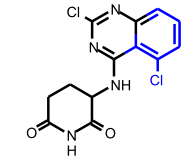
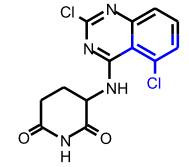
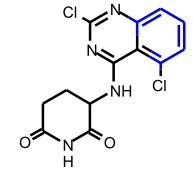
Feature Contribution

Top features for positive contribution

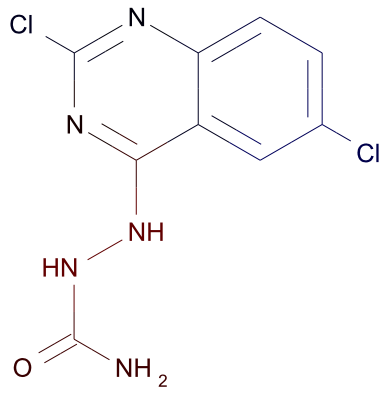
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	51876938	 [*]CCC(=[*])[*]	0.473	16 out of 31

ECFP_6	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.442	2 out of 3
ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1641317964	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.789	1 out of 11
ECFP_6	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.669	3 out of 22
ECFP_6	2007300961	 [*][c]1:[*]:[c](:[*]):[cH]:[cH]:[cH]:1	-0.652	5 out of 34

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



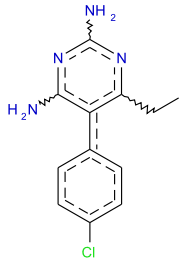
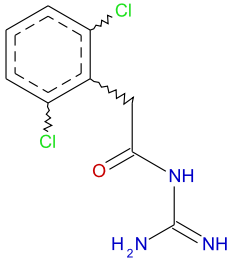
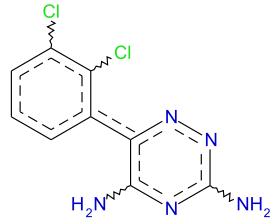
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09078
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.238
Enrichment: 0.744
Bayesian Score: -1.43
Mahalanobis Distance: 10.1
Mahalanobis Distance p-value: 0.392

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

Structural Similar Compounds

Name	Pyrimethamine	Guanfacine	Lamotrigine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.617	0.620	0.626
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

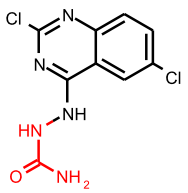
Model Applicability

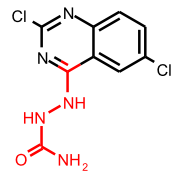
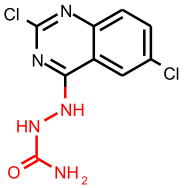
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

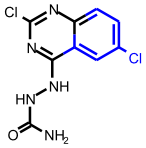
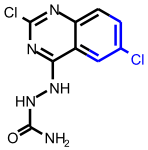
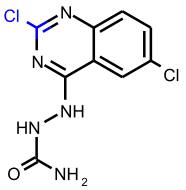
Feature Contribution

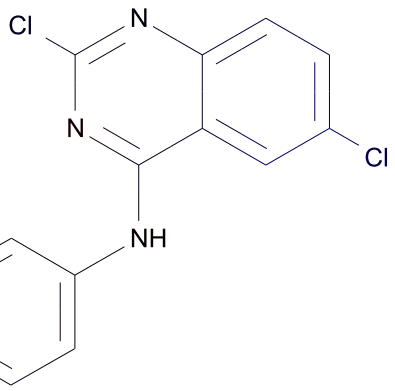
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	432952415	 [*]NC(=O)N	0.442	2 out of 3

ECFP_6	1626972527	 [*]:[c](:[*])NNC(=O)N	0.424	1 out of 1
ECFP_6	-469184004	 [*]NNC(=O)N	0.424	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	577592657	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	-0.586	3 out of 20
ECFP_6	-176494269	 [*]:[cH]:[c](Cl):[cH]:[*]	-0.476	5 out of 28
ECFP_6	99947387	 [*]:[c](:[*])Cl	-0.461	9 out of 48



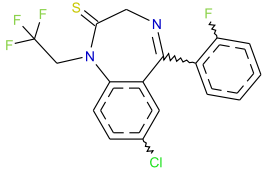
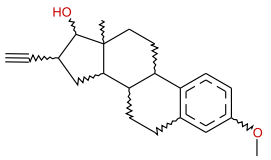
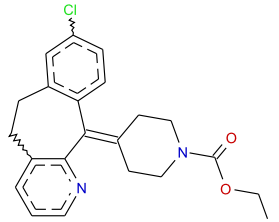
$C_{14}H_7Cl_3FN_3$
Molecular Weight: 342.58288
ALogP: 5.867
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.213
Enrichment: 0.664
Bayesian Score: -3.86
Mahalanobis Distance: 11.5
Mahalanobis Distance p-value: 0.036

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

Structural Similar Compounds

Name	Quazepam	Mestranol	Loratidine
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.627	0.631	0.634
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

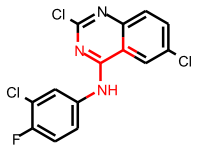
Model Applicability

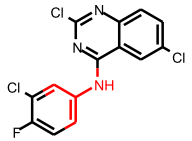
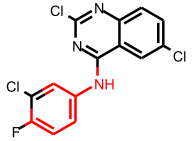
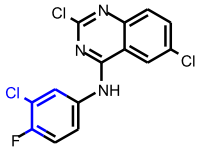
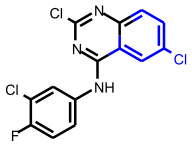
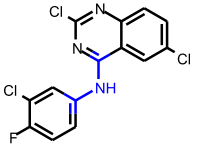
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

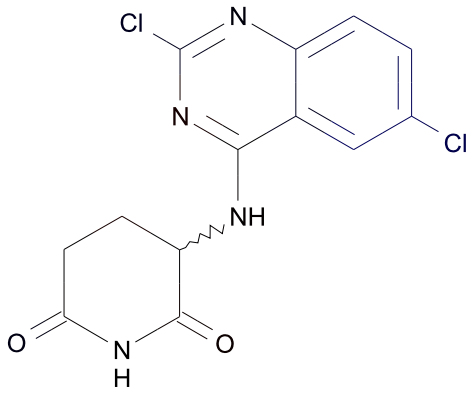
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

ECFP_6	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.279	4 out of 9
ECFP_6	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.234	3 out of 7
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	577592657	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	-0.586	3 out of 20
ECFP_6	-1242906247	 [*]:[c](:[*])N[c](:[*]):[*]	-0.482	0 out of 2

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



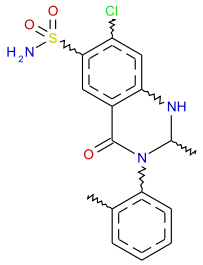
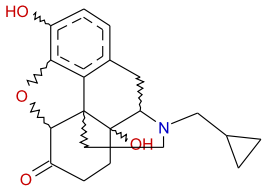
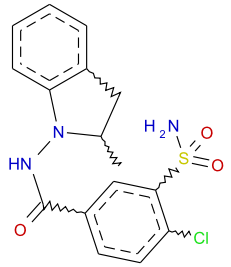
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.229
Enrichment: 0.714
Bayesian Score: -2.15
Mahalanobis Distance: 12.8
Mahalanobis Distance p-value: 0.00107

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

Structural Similar Compounds

Name	Metolazone	Naltrexone	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.564	0.579	0.583
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

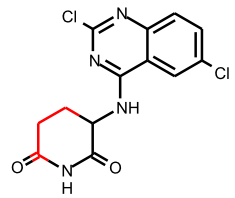
Model Applicability

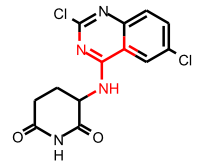
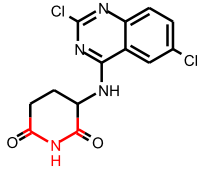
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

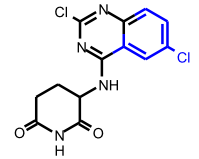
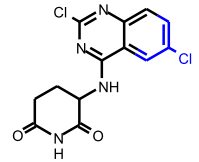
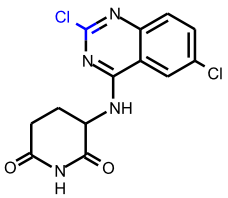
Feature Contribution

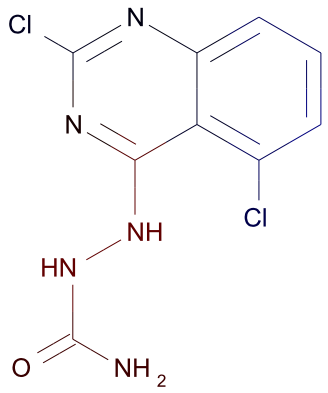
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	51876938	 [*]CCC(=[*])[*]	0.473	16 out of 31

ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1
ECFP_6	-1699286547	 [*]C(=[*])NC(=[*])[*]	0.297	12 out of 28

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	577592657	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	-0.586	3 out of 20
ECFP_6	-176494269	 [*]:[cH]:[c](Cl):[cH]:[*]	-0.476	5 out of 28
ECFP_6	99947387	 [*]:[c](:[*])Cl	-0.461	9 out of 48



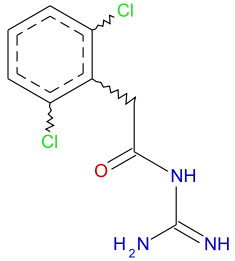
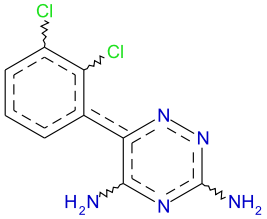
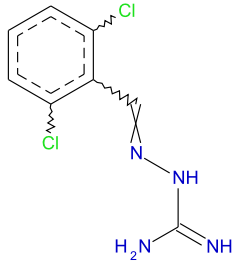
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.236
Enrichment: 0.738
Bayesian Score: -1.57
Mahalanobis Distance: 8.74
Mahalanobis Distance p-value: 0.918

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

Structural Similar Compounds

Name	Guanfacine	Lamotrigine	Guanabenz
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.586	0.597	0.608
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

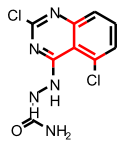
Model Applicability

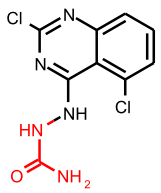
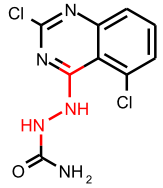
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

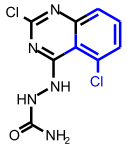
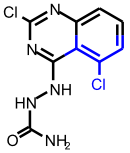
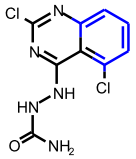
Feature Contribution

Top features for positive contribution

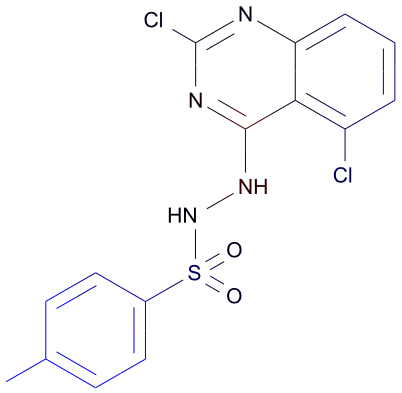
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.442	2 out of 3

ECFP_6	432952415	 [*]NC(=O)N	0.442	2 out of 3
ECFP_6	-1238415266	 [*]NN[c](:[*]):[*]	0.424	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1641317964	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.789	1 out of 11
ECFP_6	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.669	3 out of 22
ECFP_6	2007300961	 [*][c]1:[*]:[c](:[*]):[cH]:[cH]:[cH]:1	-0.652	5 out of 34

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



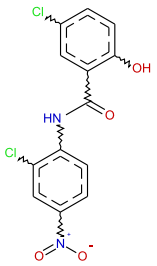
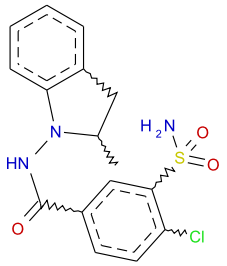
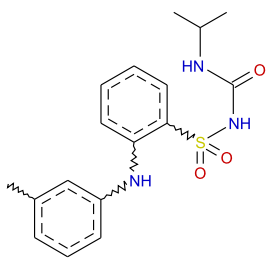
$C_{15}H_{12}Cl_2N_4O_2S$
Molecular Weight: 383.25237
ALogP: 4.443
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.213
Enrichment: 0.664
Bayesian Score: -8.77
Mahalanobis Distance: 9.67
Mahalanobis Distance p-value: 0.599

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

Structural Similar Compounds

Name	Niclosamide	Indapamide	Torsemide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.634	0.646	0.651
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

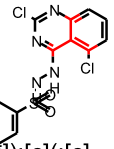
Model Applicability

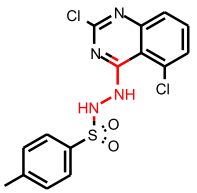
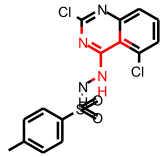
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

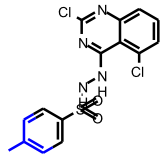
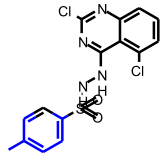
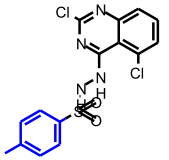
Feature Contribution

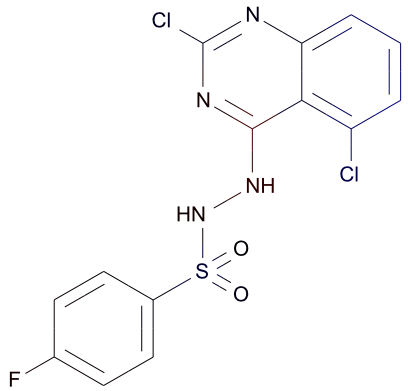
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.442	2 out of 3

ECFP_6	-1238415266	 [*]NN[c](:[*]):[*]	0.424	1 out of 1
ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

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	-1926229349	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.05	0 out of 6



$C_{14}H_9Cl_2FN_4O_2S$

Molecular Weight: 387.21626

ALogP: 4.162

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.22

Enrichment: 0.686

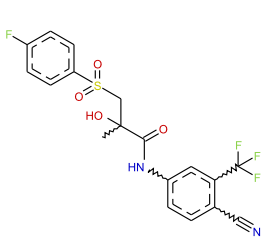
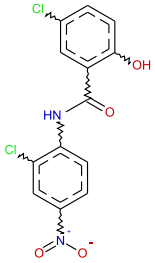
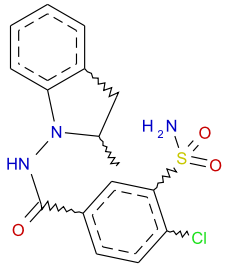
Bayesian Score: -2.98

Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.0333

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Niclosamide	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.609	0.624	0.629
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

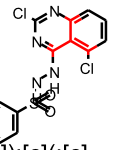
Model Applicability

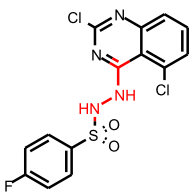
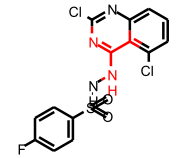
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

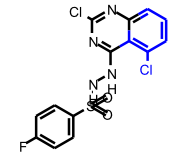
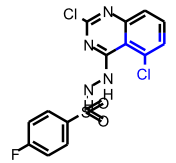
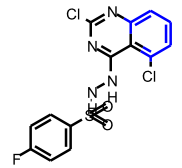
Feature Contribution

Top features for positive contribution

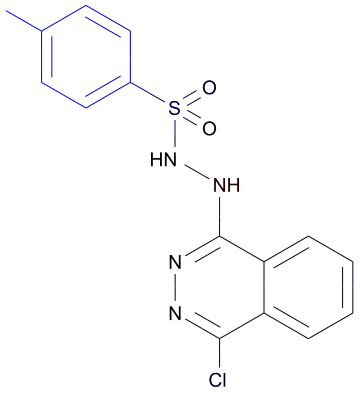
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.442	2 out of 3

ECFP_6	-1238415266	 [*]NN[c](:[*]):[*]	0.424	1 out of 1
ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1641317964	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.789	1 out of 11
ECFP_6	1335691903	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.669	3 out of 22
ECFP_6	2007300961	 [*][c]1:[*]:[c](:[*]):[cH]:[cH]:[cH]:1	-0.652	5 out of 34

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen



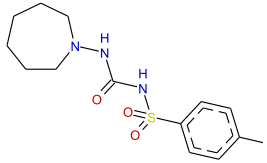
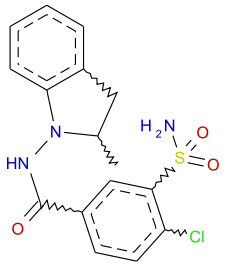
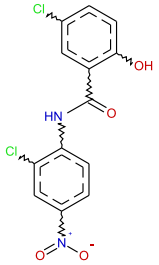
$C_{15}H_{13}ClN_4O_2S$
Molecular Weight: 348.80731
ALogP: 3.529
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.21
Enrichment: 0.654
Bayesian Score: -8.24
Mahalanobis Distance: 9.28
Mahalanobis Distance p-value: 0.766

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

Structural Similar Compounds

Name	Tolazamide	Indapamide	Niclosamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.562	0.578	0.580
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

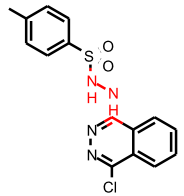
Model Applicability

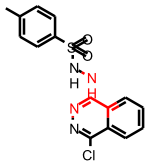
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

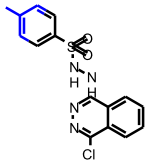
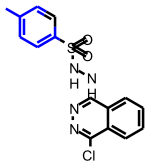
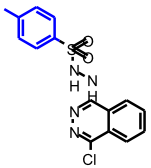
Feature Contribution

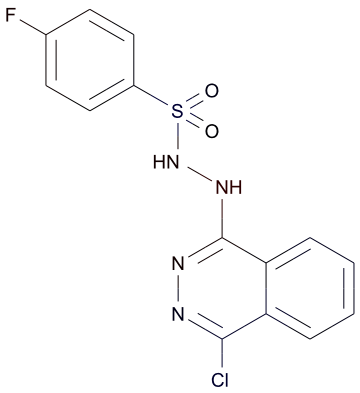
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1238415266	 [*]NN[c](:[*]):[*]	0.424	1 out of 1

ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

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	-1926229349	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.05	0 out of 6



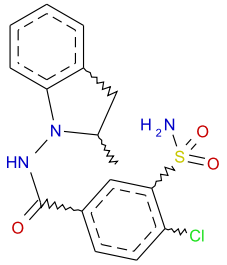
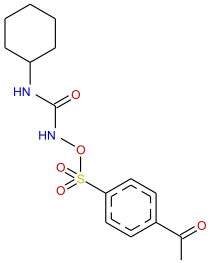
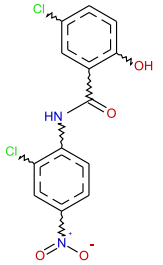
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.233
Enrichment: 0.725
Bayesian Score: -1.85
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.0895

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

Structural Similar Compounds

Name	Indapamide	Acetohexamide	Niclosamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.567	0.583	0.583
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

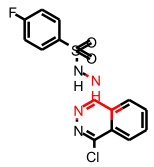
Model Applicability

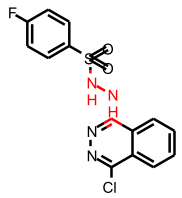
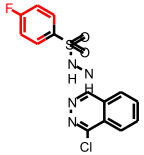
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

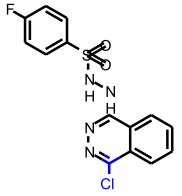
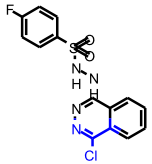
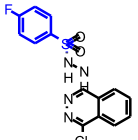
Feature Contribution

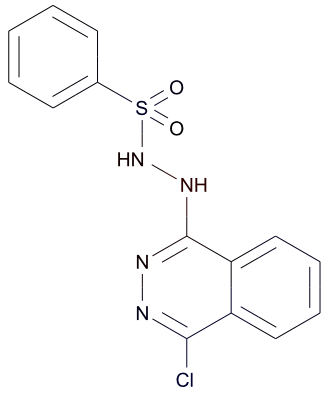
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

ECFP_6	-1238415266	 [*]NN[c](:[*]):[*]	0.424	1 out of 1
ECFP_6	-296909061	 F[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	0.141	3 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	99947387	 [*]:[c](:[*])Cl	-0.461	9 out of 48
ECFP_6	1050351974	 [*]:n:[c](Cl):[c](:[*]):[*]	-0.27	0 out of 1
ECFP_6	74189795	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	-0.27	0 out of 1



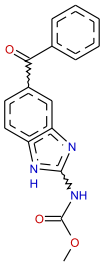
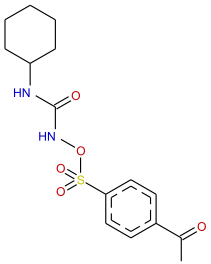
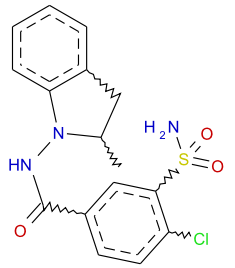
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.227
Enrichment: 0.709
Bayesian Score: -2.28
Mahalanobis Distance: 9.73
Mahalanobis Distance p-value: 0.569

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

Structural Similar Compounds

Name	Mebendazole	Acetohexamide	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.562	0.566	0.569
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

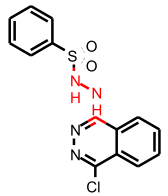
Model Applicability

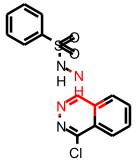
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

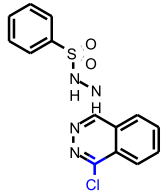
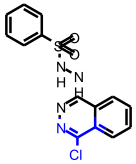
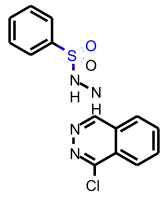
Feature Contribution

Top features for positive contribution

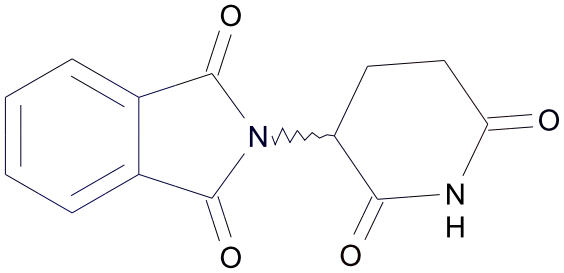
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-1238415266	 [*]NN[c](:[*]):[*]	0.424	1 out of 1

ECFP_6	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.424	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	99947387	 [*]:[c](:[*])Cl	-0.461	9 out of 48
ECFP_6	1050351974	 [*]:n:[c](Cl):[c](:[*]):[*]	-0.27	0 out of 1
ECFP_6	2102150379	 [*]S(=[*])(=O)[*]	-0.263	5 out of 22

Thalidomide



$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

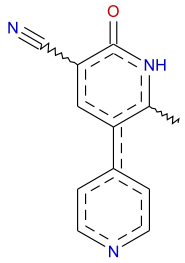
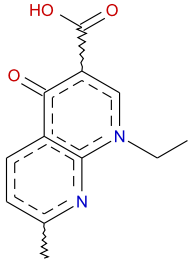
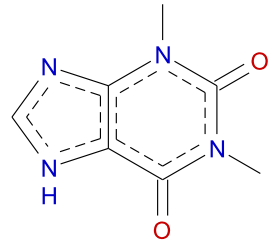
Model Prediction

Prediction: Non-Carcinogen
Probability: 0.244
Enrichment: 0.761
Bayesian Score: -1.06
Mahalanobis Distance: 11
Mahalanobis Distance p-value: 0.107

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

TOPKAT_Mouse_Female_FDA_None_vs_Carcinogen

Structural Similar Compounds

Name	Milrinone	Nalidixic acid	Theophylline
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.597	0.627	0.630
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

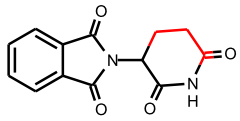
Model Applicability

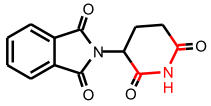
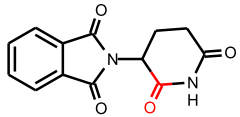
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

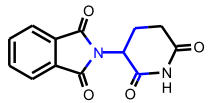
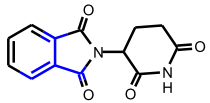
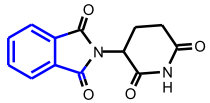
Feature Contribution

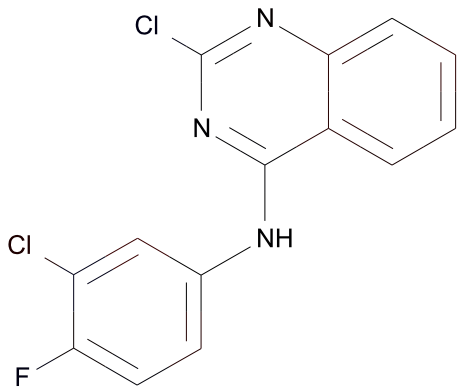
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	51876938	 [*]CCC(=[*])[*]	0.473	16 out of 31

ECFP_6	-1699286547	 [*]C(=[*])NC(=[*])[*]	0.297	12 out of 28
ECFP_6	2106656448	 [*]C(=O)[*]	0.254	31 out of 77

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_6	-484970154	 [*]CC(N([*])[*])C(=[*])[*]	-0.805	0 out of 4
ECFP_6	-427397688	 [*]=C1[*][*][c](:[*]) :[c]1:[cH]:[*]	-0.476	5 out of 28
ECFP_6	1984293270	 [*]=C1[*][*][c]2:[cH] :[cH]:[cH]:[cH]:[c]1 :2	-0.311	2 out of 10



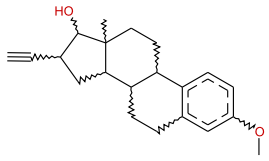
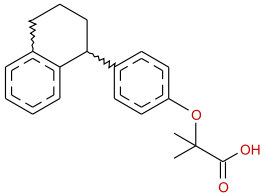
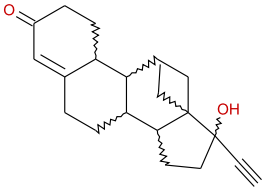
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.353
Enrichment: 1.2
Bayesian Score: 1.62
Mahalanobis Distance: 13
Mahalanobis Distance p-value: 0.000298

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

Structural Similar Compounds

Name	Mestranol	Nafenopin	Levonorgestrel
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.605	0.609	0.609
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

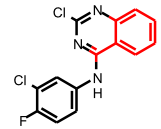
Model Applicability

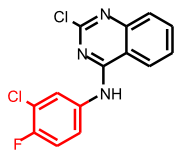
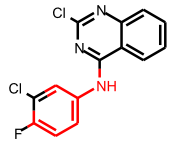
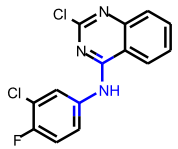
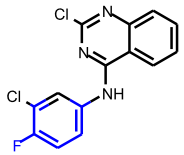
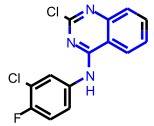
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

Feature Contribution

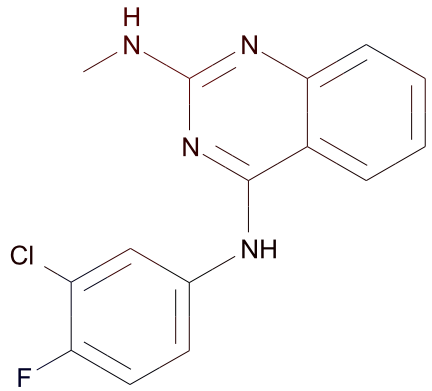
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-387072142	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.477	4 out of 8

FCFP_6	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.46	1 out of 1
FCFP_6	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.409	10 out of 24
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Carcinogen in training set**
FCFP_6	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	 [*]N[c]1:n:[*]:n:[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	-0.423	0 out of 2

7a

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.346

Enrichment: 1.17

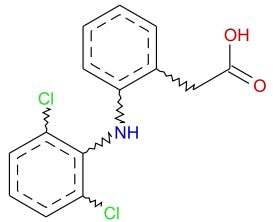
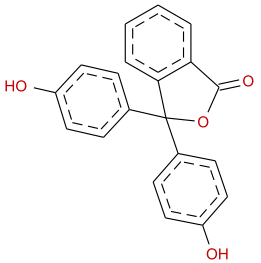
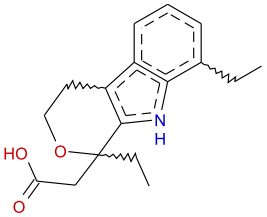
Bayesian Score: 1.37

Mahalanobis Distance: 12.4

Mahalanobis Distance p-value: 0.00178

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

Structural Similar Compounds

Name	Diclofenac	Phenolphthalein	Etodolac
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.530	0.590	0.614
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

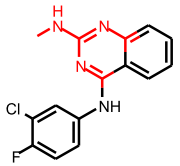
Model Applicability

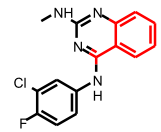
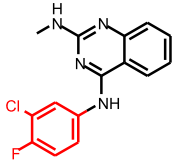
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

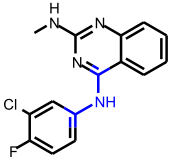
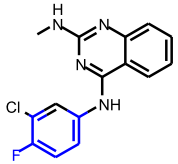
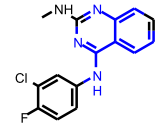
Feature Contribution

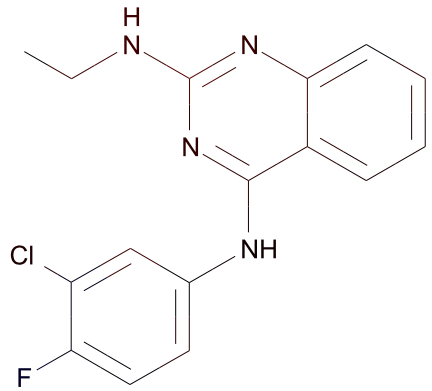
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	960546407	 [*][c]1:[*]:[c](:[*]):n:[c](NC):n:1	0.517	2 out of 3

FCFP_6	-387072142	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.477	4 out of 8
FCFP_6	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[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	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	 [*]N[c]1:n:[*]:n:[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	-0.423	0 out of 2



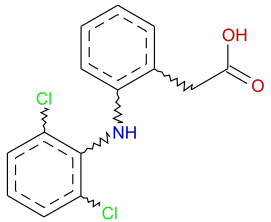
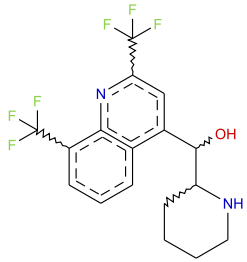
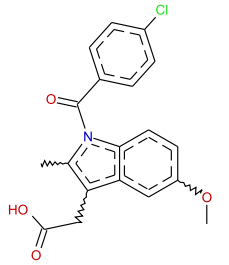
$C_{16}H_{14}ClFN_4$
Molecular Weight: 316.76056
ALogP: 5.025
Rotatable Bonds: 4
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.358
Enrichment: 1.22
Bayesian Score: 1.77
Mahalanobis Distance: 13.3
Mahalanobis Distance p-value: 0.000109

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

Structural Similar Compounds

Name	Diclofenac	Mefloquine	Indomethacin
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.536	0.592	0.621
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

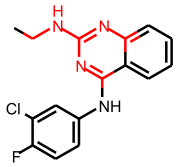
Model Applicability

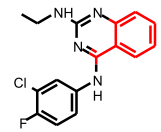
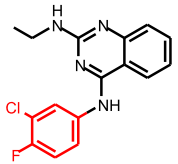
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

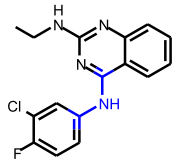
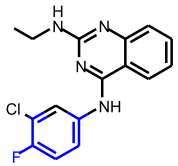
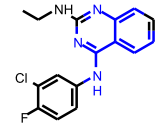
Feature Contribution

Top features for positive contribution

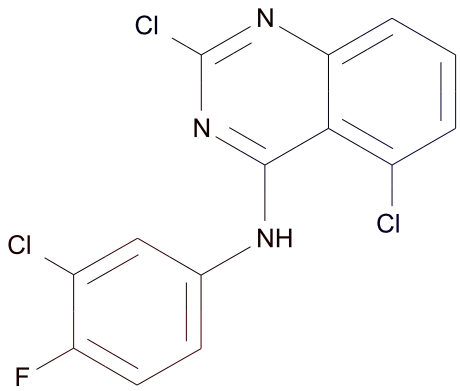
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	960546407	 [*][c]1:[*]:[c](:[*]):n:[c](NC):n:1	0.517	2 out of 3

FCFP_6	-387072142	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c] :1:[*]	0.477	4 out of 8
FCFP_6	-2132756387	 [*][c]1:[cH]:[cH]:[c] (F):[c](Cl):[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	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]): [c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	 [*]N[c]1:n:[*]:n:[c]2 :[cH]:[*]:[cH]:[cH]: [c]:1:2	-0.423	0 out of 2

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.286

Enrichment: 0.972

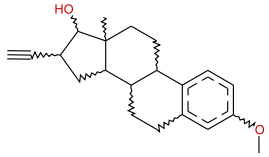
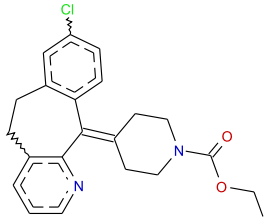
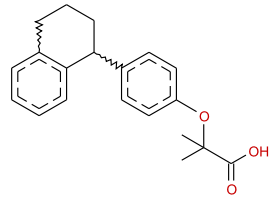
Bayesian Score: -0.75

Mahalanobis Distance: 13

Mahalanobis Distance p-value: 0.0003

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

Structural Similar Compounds

Name	Mestranol	Loratidine	Nafenopin
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.633	0.638	0.646
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

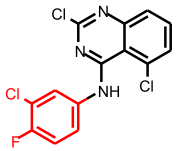
Model Applicability

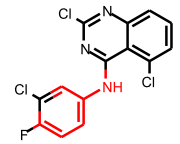
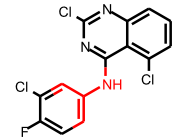
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

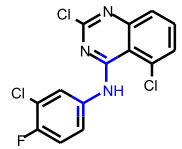
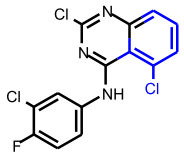
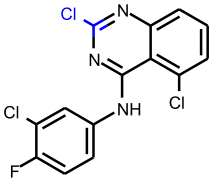
Feature Contribution

Top features for positive contribution

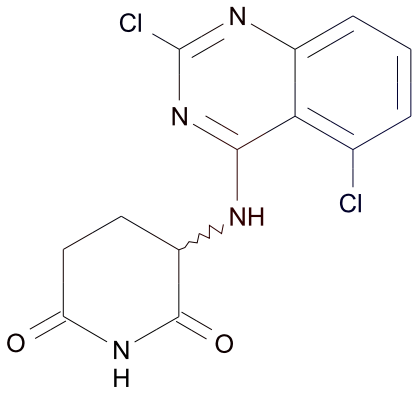
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.46	1 out of 1

FCFP_6	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.409	10 out of 24
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.369	13 out of 33

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



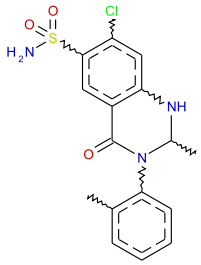
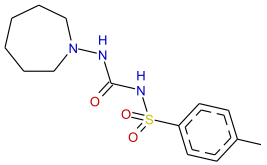
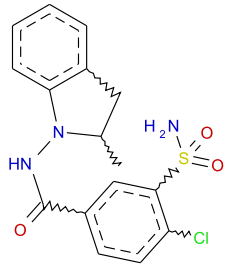
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.275
Enrichment: 0.935
Bayesian Score: -1.19
Mahalanobis Distance: 12.8
Mahalanobis Distance p-value: 0.000579

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

Structural Similar Compounds

Name	Metolazone	Tolazamide	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.550	0.566	0.568
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

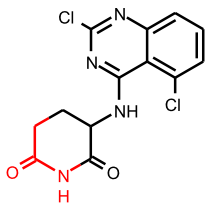
Model Applicability

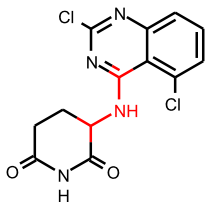
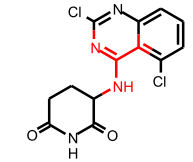
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

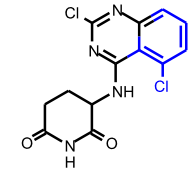
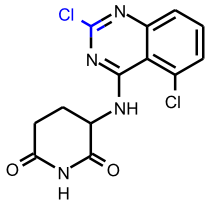
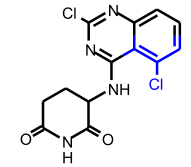
Feature Contribution

Top features for positive contribution

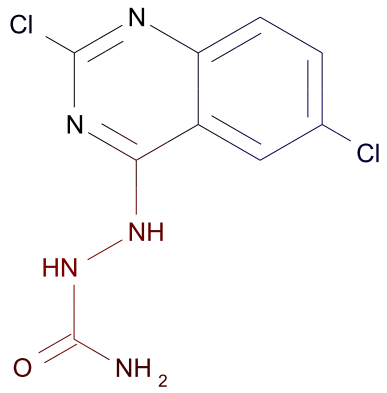
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	566058135	 [*]CC(=O)N[*]	0.447	17 out of 40

FCFP_6	1294255210	 [*]:[c](:[*])NC	0.441	12 out of 28
FCFP_6	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.348	6 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59
FCFP_6	367998008	 [*][c](:[*]):[c](F):[cH]:[*]	-0.374	10 out of 57

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



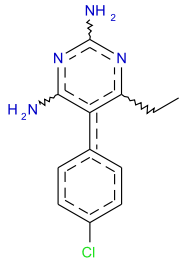
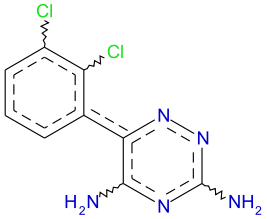
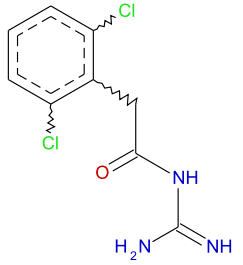
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09078
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Carcinogen
Probability: 0.34
Enrichment: 1.15
Bayesian Score: 1.17
Mahalanobis Distance: 10.6
Mahalanobis Distance p-value: 0.15

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

Structural Similar Compounds

Name	Pyrimethamine	Lamotrigine	Guanfacine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.595	0.607	0.611
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

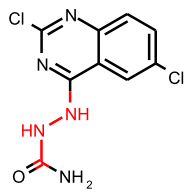
Model Applicability

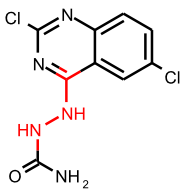
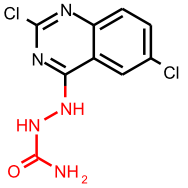
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

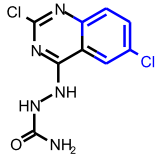
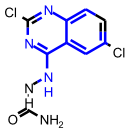
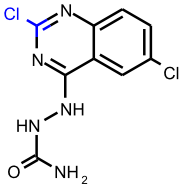
Feature Contribution

Top features for positive contribution

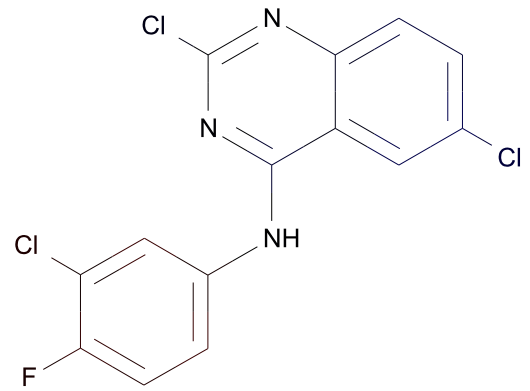
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885461129	 [*]NNC(=[*])[*]	0.547	3 out of 5

FCFP_6	1294344583	 [*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	-1995759737	 [*]NNC(=O)N	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	 [*]N[c]1:n:[*]:n:[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	-0.423	0 out of 2
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.273

Enrichment: 0.927

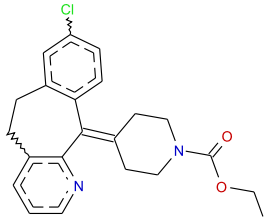
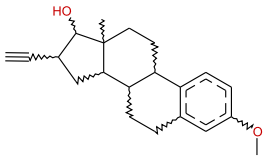
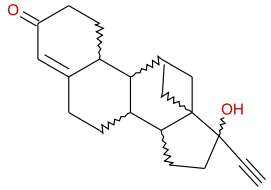
Bayesian Score: -1.29

Mahalanobis Distance: 12

Mahalanobis Distance p-value: 0.00584

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

Structural Similar Compounds

Name	Loratidine	Mestranol	Levonorgestrel
Structure			
Actual Endpoint	Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.633	0.633	0.649
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

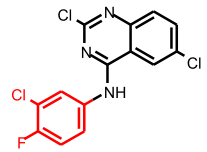
Model Applicability

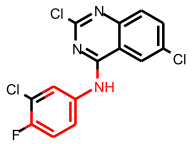
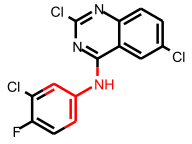
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

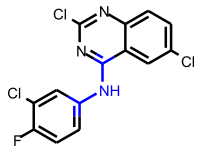
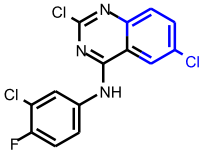
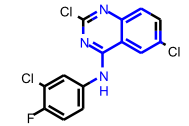
Feature Contribution

Top features for positive contribution

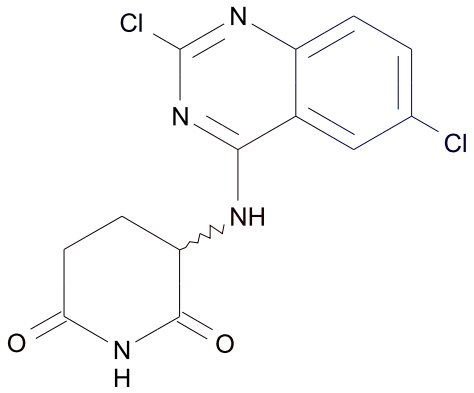
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.46	1 out of 1

FCFP_6	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.409	10 out of 24
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.369	13 out of 33

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.719	0 out of 4
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	 [*]N[c]1:n:[*]:n:[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	-0.423	0 out of 2

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



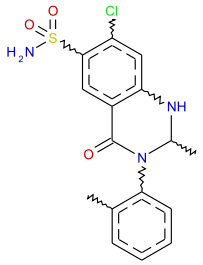
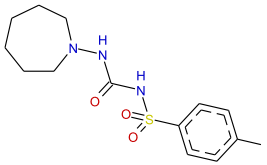
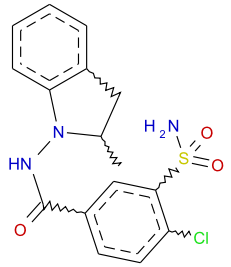
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.262
Enrichment: 0.892
Bayesian Score: -1.73
Mahalanobis Distance: 12.2
Mahalanobis Distance p-value: 0.00356

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

Structural Similar Compounds

Name	Metolazone	Tolazamide	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.556	0.566	0.574
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

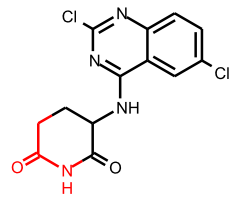
Model Applicability

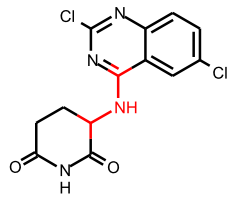
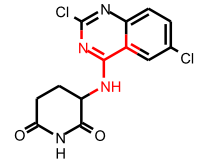
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

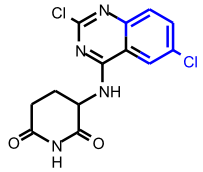
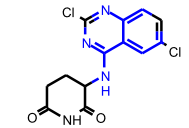
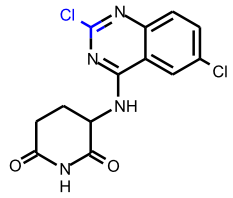
Feature Contribution

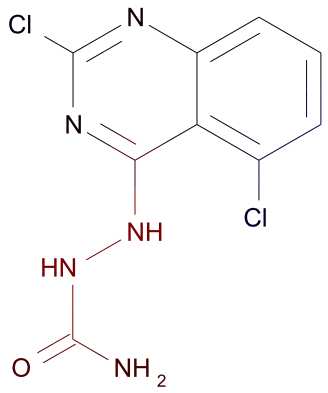
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	566058135	 [*]CC(=O)N[*]	0.447	17 out of 40

FCFP_6	1294255210	 [*]:[c](:[*])NC	0.441	12 out of 28
FCFP_6	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.348	6 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	1122741451	 [*]N[c]1:n:[*]:n:[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	-0.423	0 out of 2
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59



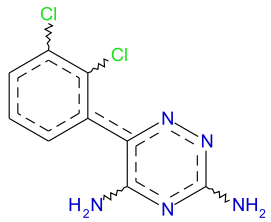
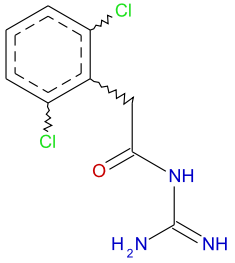
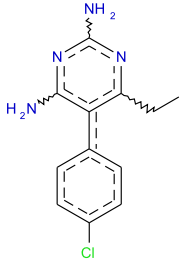
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Carcinogen
Probability: 0.356
Enrichment: 1.21
Bayesian Score: 1.71
Mahalanobis Distance: 12.3
Mahalanobis Distance p-value: 0.00285

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

Structural Similar Compounds

Name	Lamotrigine	Guanfacine	Pyrimethamine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.599	0.602	0.604
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

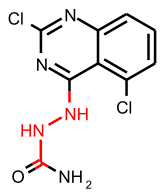
Model Applicability

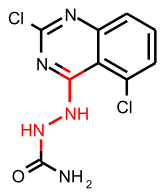
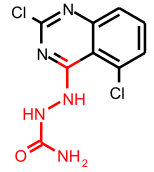
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

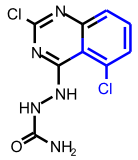
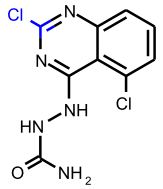
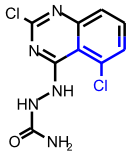
Feature Contribution

Top features for positive contribution

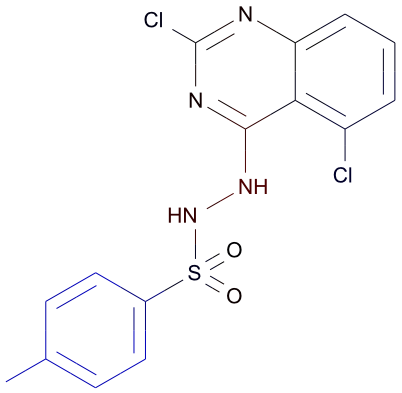
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885461129	 [*]NNC(=[*])[*]	0.547	3 out of 5

FCFP_6	1294344583	 [*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	1572236312	 [*]:[c](:[*])NNC(=O)N	0.46	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59
FCFP_6	367998008	 [*][c](:[*]):[c](F):[cH]:[*]	-0.374	10 out of 57

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{15}H_{12}Cl_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227

Enrichment: 0.773

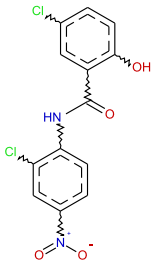
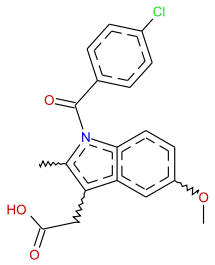
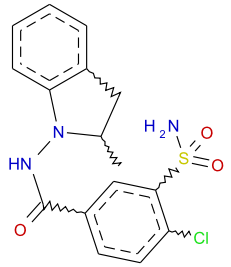
Bayesian Score: -3.4

Mahalanobis Distance: 12.7

Mahalanobis Distance p-value: 0.000761

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

Structural Similar Compounds

Name	Niclosamide	Indomethacin	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.604	0.611	0.624
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

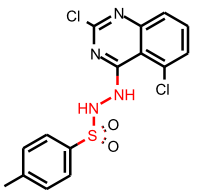
Model Applicability

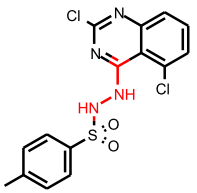
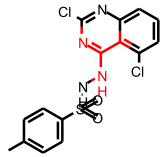
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

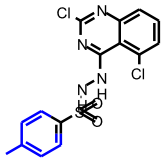
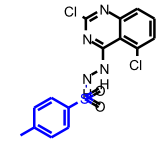
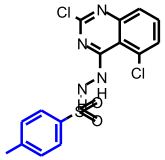
Feature Contribution

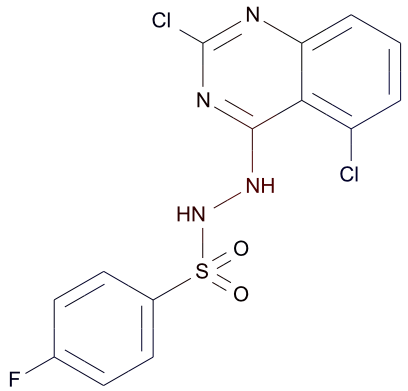
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885461129	 [*]NNC(=[*])[*]	0.547	3 out of 5

FCFP_6	1294344583	 [*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.348	6 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 10
FCFP_6	632767364	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.04	0 out of 7
FCFP_6	2109043264	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.947	0 out of 6



$C_{14}H_9Cl_2FN_4O_2S$

Molecular Weight: 387.21626

ALogP: 4.162

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.307

Enrichment: 1.04

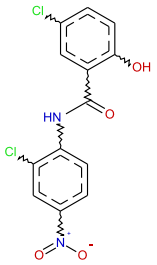
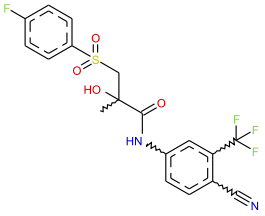
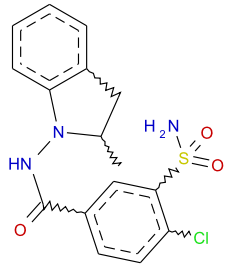
Bayesian Score: 0.0577

Mahalanobis Distance: 9.9

Mahalanobis Distance p-value: 0.421

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

Structural Similar Compounds

Name	Niclosamide	Bicalutamide	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.586	0.601	0.601
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

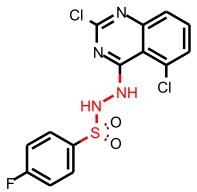
Model Applicability

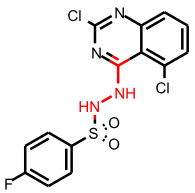
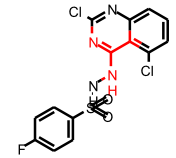
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

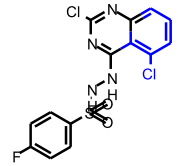
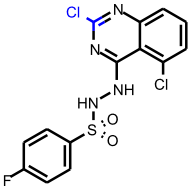
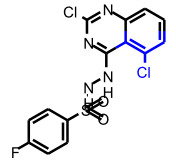
Feature Contribution

Top features for positive contribution

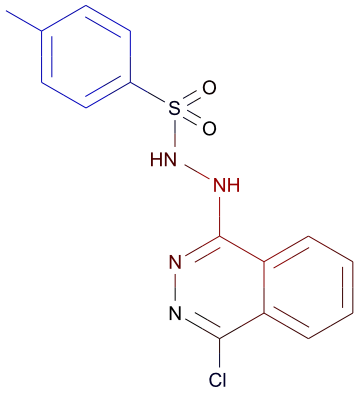
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885461129	 [*]NNC(=[*])[*]	0.547	3 out of 5

FCFP_6	1294344583	 [*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.348	6 out of 15

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59
FCFP_6	367998008	 [*][c](:[*]):[c](F):[cH]:[*]	-0.374	10 out of 57

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{15}H_{13}ClN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.289

Enrichment: 0.983

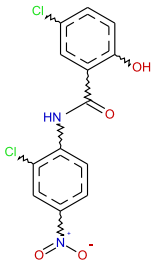
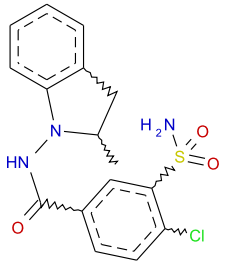
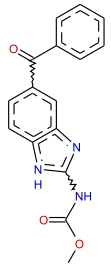
Bayesian Score: -0.629

Mahalanobis Distance: 18.8

Mahalanobis Distance p-value: 4.69e-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	Niclosamide	Indapamide	Mebendazole
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.556	0.567	0.575
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

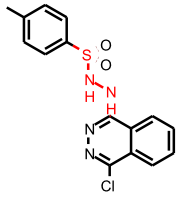
Model Applicability

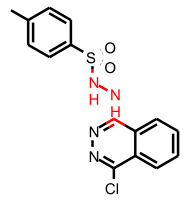
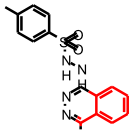
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

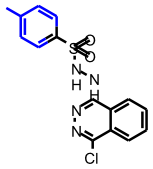
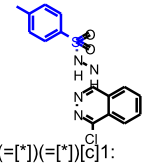
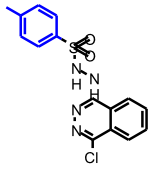
Feature Contribution

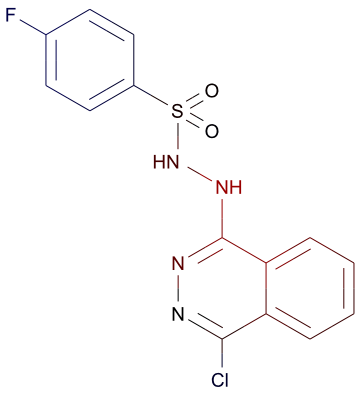
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885461129	 [*]NNC(=[*])[*]	0.547	3 out of 5

FCFP_6	1294344583	 [*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	-387072142	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.477	4 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1773728142	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.29	0 out of 10
FCFP_6	632767364	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.04	0 out of 7
FCFP_6	2109043264	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.947	0 out of 6



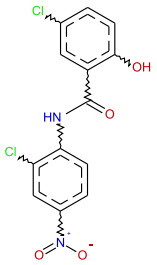
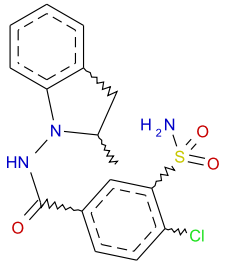
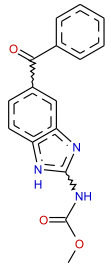
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.344
Enrichment: 1.17
Bayesian Score: 1.32
Mahalanobis Distance: 10.5
Mahalanobis Distance p-value: 0.189

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

Structural Similar Compounds

Name	Niclosamide	Indapamide	Mebendazole
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.539	0.539	0.576
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

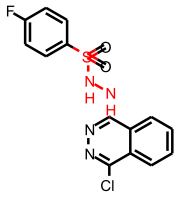
Model Applicability

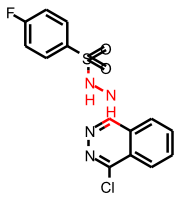
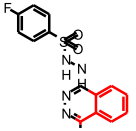
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

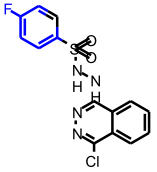
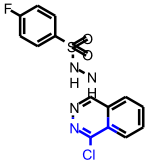
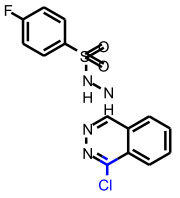
Feature Contribution

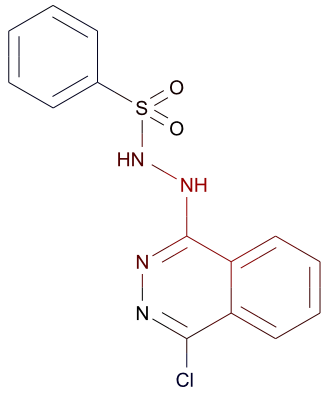
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885461129	 [*]NNC(=[*])[*]	0.547	3 out of 5

FCFP_6	1294344583	 [*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	-387072142	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.477	4 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	551850122	 [*][c]1:[*]:[c]([*]):[c](F):[cH]:[cH]:1	-0.433	8 out of 49
FCFP_6	-1374812327	 [*]:n:[c](Cl):[c](:[*]):[*]	-0.423	0 out of 2
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59



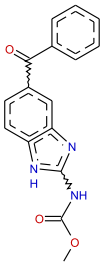
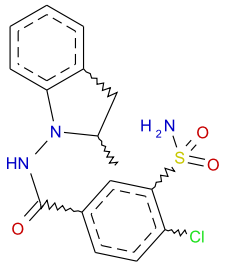
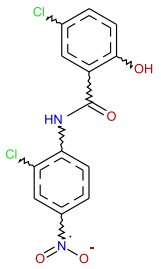
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.322
Enrichment: 1.09
Bayesian Score: 0.569
Mahalanobis Distance: 19.8
Mahalanobis Distance p-value: 2.3e-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	Mebendazole	Indapamide	Niclosamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.542	0.547	0.554
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

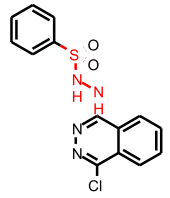
Model Applicability

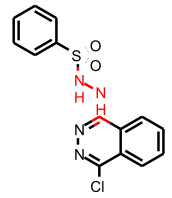
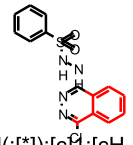
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

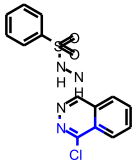
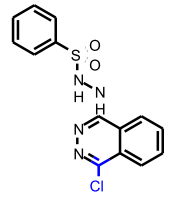
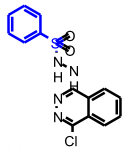
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-885461129	 [*]NNC(=[*])[*]	0.547	3 out of 5

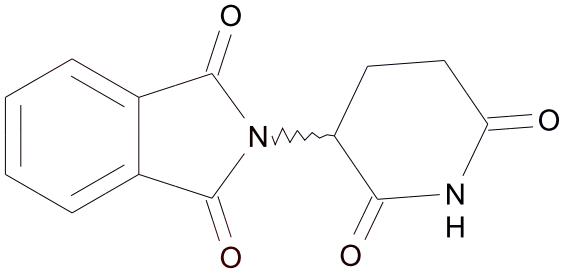
FCFP_6	1294344583	 [*]NN[c](:[*]):[*]	0.517	2 out of 3
FCFP_6	-387072142	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.477	4 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1374812327	 [*]:n:[c](Cl):[c](:[*]):[*]	-0.423	0 out of 2
FCFP_6	71476542	 [*]:[c](:[*])Cl	-0.406	10 out of 59
FCFP_6	-1698724694	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.22	15 out of 72

Thalidomide

TOPKAT_Mouse_Male_FDA_None_vs_Carcinogen



$C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002

Rotatable Bonds: 1

Acceptors: 4

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.317

Enrichment: 1.08

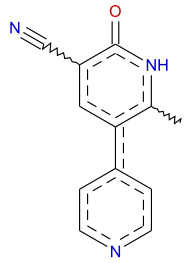
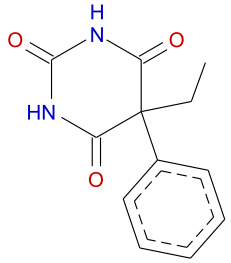
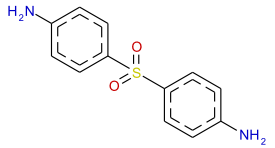
Bayesian Score: 0.389

Mahalanobis Distance: 8.54

Mahalanobis Distance p-value: 0.928

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

Structural Similar Compounds

Name	Milrinone	Phenobarbital	Dapsone
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Carcinogen	Carcinogen
Distance	0.575	0.587	0.618
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

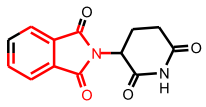
Model Applicability

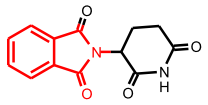
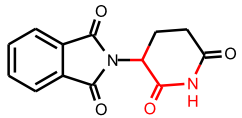
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

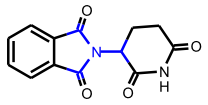
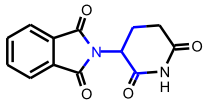
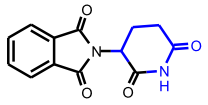
Feature Contribution

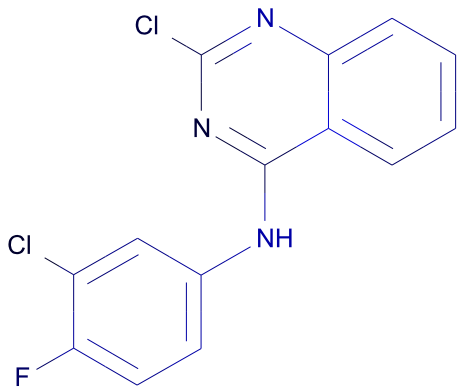
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	2139882011	 [*]N1C(=[*])[c]2:[cH]:[*]:[cH]:[cH]:[c]:2C1=O	0.46	1 out of 1

FCFP_6	1393189956	 [*]N1C(=[*])[c]2:[cH]:[cH]:[cH]:[cH]:[c]:2C1=O	0.46	1 out of 1
FCFP_6	566058135	 [*]CC(=O)N[*]	0.447	17 out of 40

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
FCFP_6	-1553874037	 [*]C([*])N1C(=[*])[*]:[*]C1=[*]	-0.45	5 out of 32
FCFP_6	-1043310069	 [*]CC(N([*])[*])C(=[*])[*]	-0.406	10 out of 59
FCFP_6	-228300541	 [*]C1[*]NC(=O)CC1	-0.233	0 out of 1



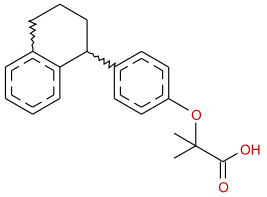
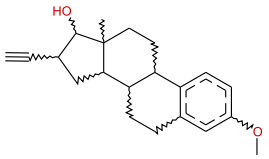
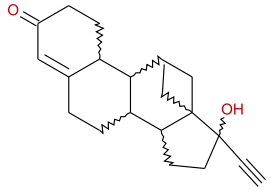
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.147
Enrichment: 0.49
Bayesian Score: -10.8
Mahalanobis Distance: 13
Mahalanobis Distance p-value: 9.61e-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	Nafenopin	Mestranol	Levonorgestrel
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.622	0.623	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

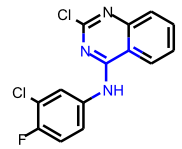
Model Applicability

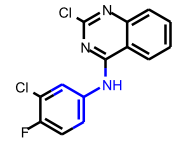
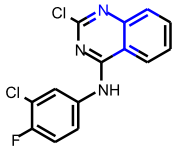
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC7 out of range. Value: -4.0266. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]
3. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

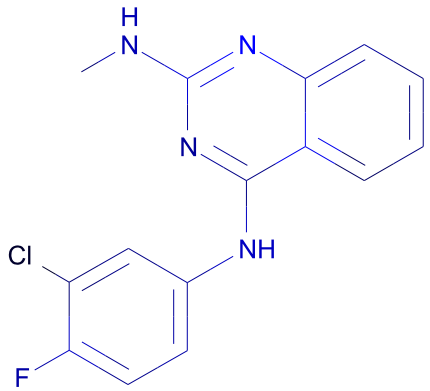
Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	-1.11	0 out of 6

FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	-0.994	0 out of 5

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.143

Enrichment: 0.474

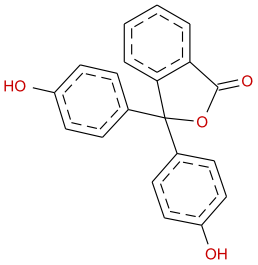
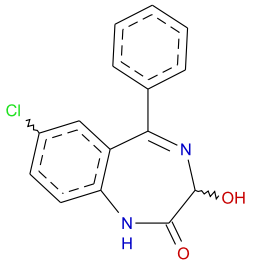
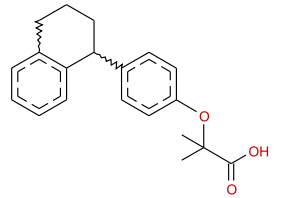
Bayesian Score: -14.3

Mahalanobis Distance: 11.9

Mahalanobis Distance p-value: 0.000707

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

Structural Similar Compounds

Name	Phenolphthalein	Oxazepam	Nafenopin
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.593	0.684	0.710
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

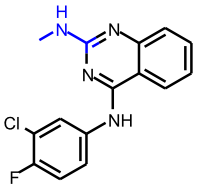
Model Applicability

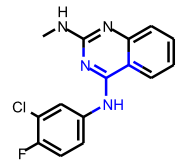
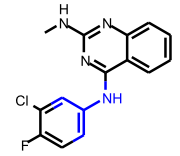
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

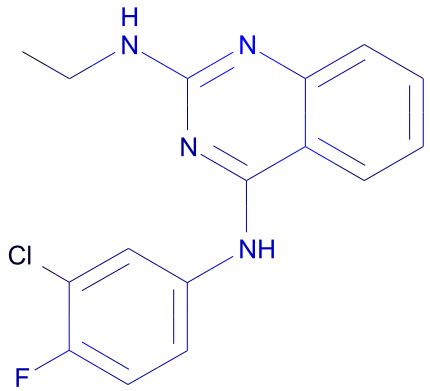
Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]:[c](:[*])NC	-1.63	0 out of 12

FCFP_12	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	-1.11	0 out of 6
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.998	1 out of 13

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{16}H_{14}ClFN_4$

Molecular Weight: 316.76056

ALogP: 5.025

Rotatable Bonds: 4

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.138

Enrichment: 0.46

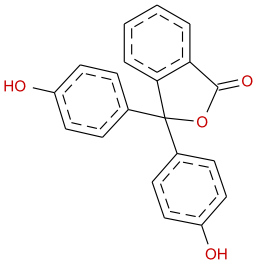
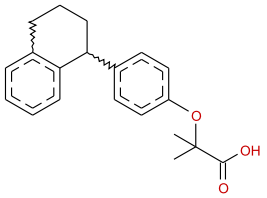
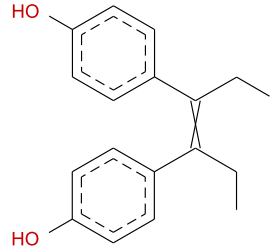
Bayesian Score: -14.8

Mahalanobis Distance: 15

Mahalanobis Distance p-value: 2.71e-006

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

Structural Similar Compounds

Name	Phenolphthalein	Nafenopin	Diethylstilbesterol
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.657	0.695	0.702
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

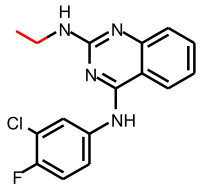
Model Applicability

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

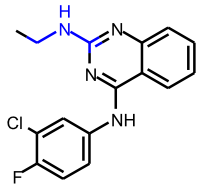
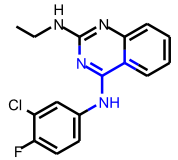
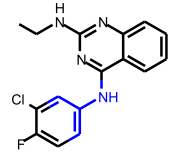
1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution

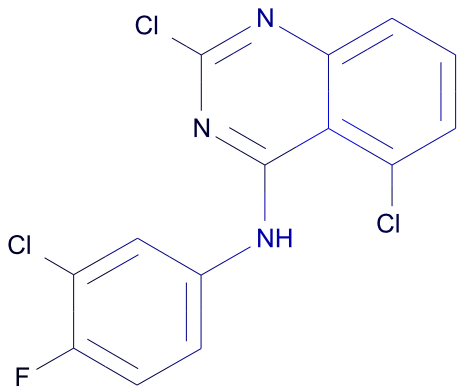
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	136597326	 [*]CC	0.0722	18 out of 49

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	1294255210	 [*]:[c](:[*])NC	-1.63	0 out of 12
FCFP_12	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	-1.11	0 out of 6
FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.998	1 out of 13

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.147

Enrichment: 0.488

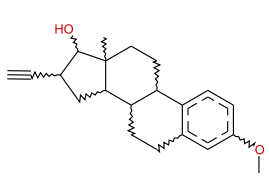
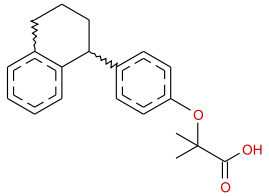
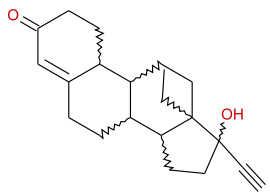
Bayesian Score: -10.6

Mahalanobis Distance: 13.2

Mahalanobis Distance p-value: 6.75e-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	Mestranol	Nafenopin	Levonorgestrel
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.647	0.654	0.658
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

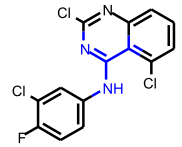
Model Applicability

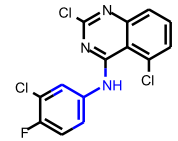
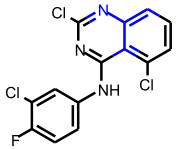
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

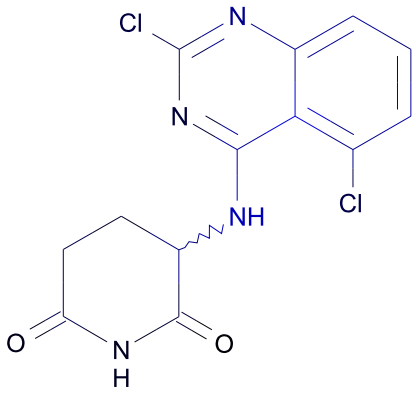
1. OPS PC7 out of range. Value: -4.1247. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]
3. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	-1.11	0 out of 6

FCFP_12	590925877	 [*]N[c](:[cH]:[*]):[c H]:[*]	-0.998	1 out of 13
FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	-0.994	0 out of 5



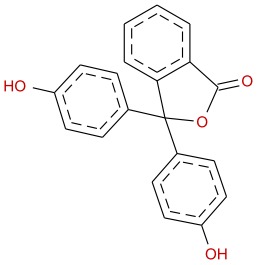
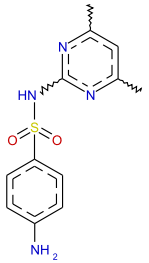
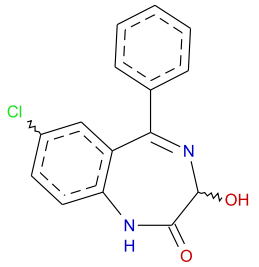
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.149
Enrichment: 0.495
Bayesian Score: -13.6
Mahalanobis Distance: 11.2
Mahalanobis Distance p-value: 0.00242

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

Structural Similar Compounds

Name	Phenolphthalein	Sulfamethazine	Oxazepam
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.652	0.657	0.678
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

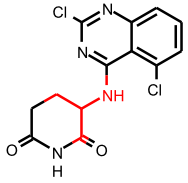
Model Applicability

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

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

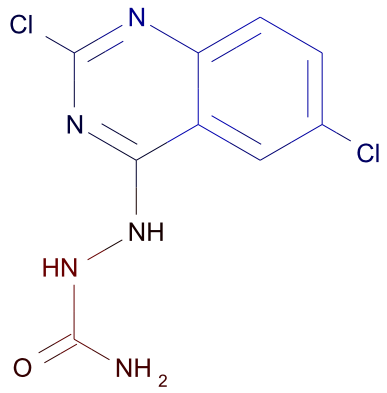
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1043250487	 [*]CC(N[*])C(=[*])[*]	0.0691	7 out of 19

Top Features for negative contribution

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09078

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.151

Enrichment: 0.5

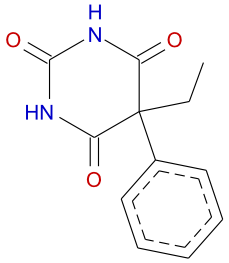
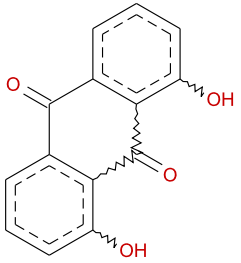
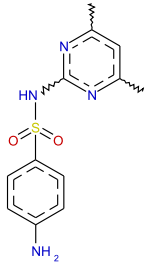
Bayesian Score: -7.23

Mahalanobis Distance: 11.7

Mahalanobis Distance p-value: 0.00105

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

Structural Similar Compounds

Name	Phenobarbital	Danthron	Sulfamethazine
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.706	0.708	0.711
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

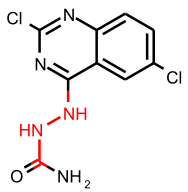
Model Applicability

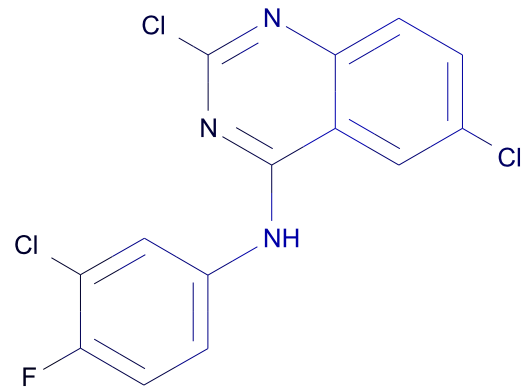
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-885461129	 [*]NNC(=[*])[*]	0.683	3 out of 3



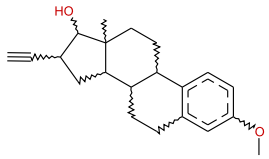
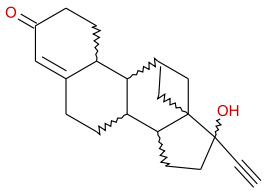
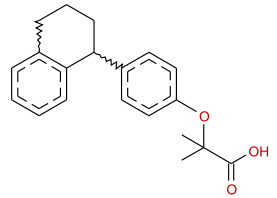
$C_{14}H_7Cl_3FN_3$
Molecular Weight: 342.58288
ALogP: 5.867
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.147
Enrichment: 0.488
Bayesian Score: -10.5
Mahalanobis Distance: 12
Mahalanobis Distance p-value: 0.000594

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

Structural Similar Compounds

Name	Mestranol	Levonorgestrel	Nafenopin
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.647	0.658	0.659
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

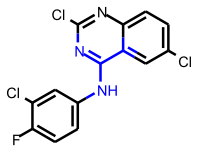
Model Applicability

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

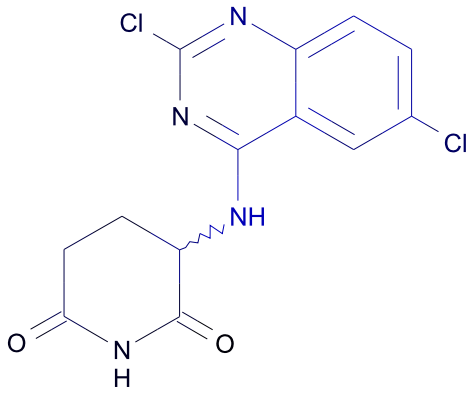
1. OPS PC7 out of range. Value: -4.4725. Training min, max, SD, explained variance: -3.9478, 3.0766, 1.527, 0.0435.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]
3. Unknown FCFP_2 feature: 1293778554: [*]:[c](:[*])N[c](:[*]):[*]

Feature Contribution

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	-1.11	0 out of 6

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.149

Enrichment: 0.496

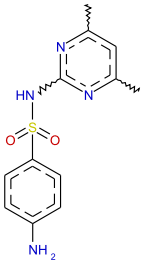
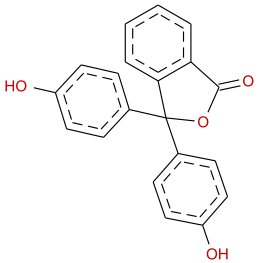
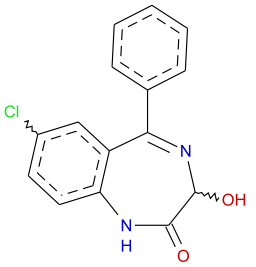
Bayesian Score: -13.6

Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.00447

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

Structural Similar Compounds

Name	Sulfamethazine	Phenolphthalein	Oxazepam
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.657	0.657	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

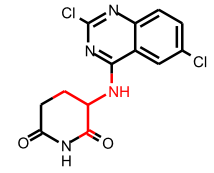
Model Applicability

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 FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

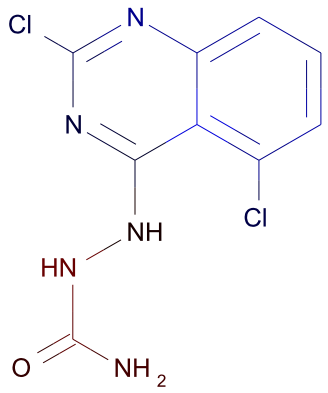
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1043250487	 [*]CC(N[*])C(=[*])[*]	0.0691	7 out of 19

Top Features for negative contribution

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09077

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.15

Enrichment: 0.5

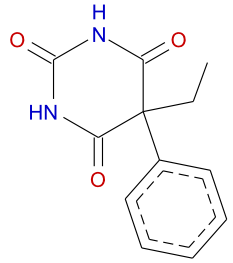
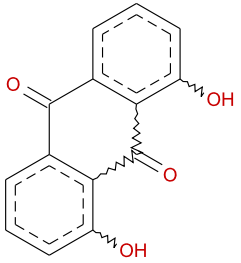
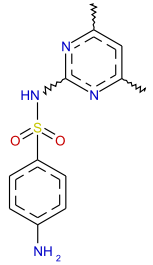
Bayesian Score: -7.27

Mahalanobis Distance: 12.5

Mahalanobis Distance p-value: 0.000231

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

Structural Similar Compounds

Name	Phenobarbital	Danthron	Sulfamethazine
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.698	0.700	0.711
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

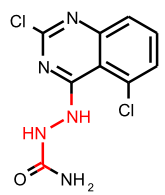
Model Applicability

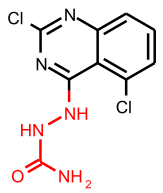
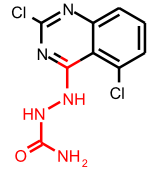
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

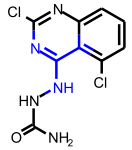
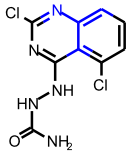
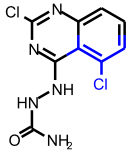
Feature Contribution

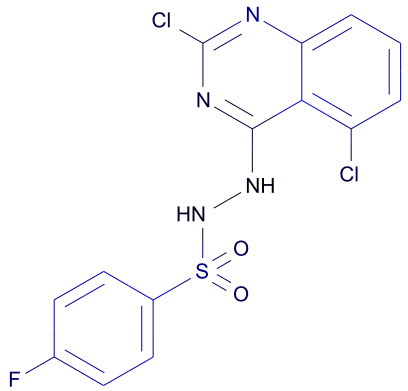
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-885461129	 [*]NNC(=[*])[*]	0.683	3 out of 3

FCFP_12	-1995759737	 [*]NNC(=O)N	0.4	1 out of 1
FCFP_12	1572236312	 [*]:[c](:[*])NNC(=O)N	0.4	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	-1.11	0 out of 6
FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]):[c](:[*]):[*]	-0.994	0 out of 5
FCFP_12	367998008	 [*][c](:[*]):[c](F):[cH]:[*]	-0.789	1 out of 10



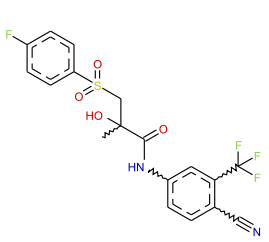
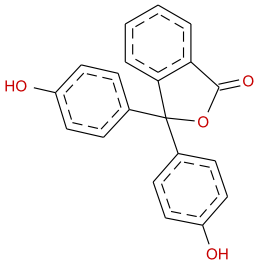
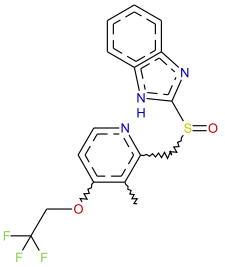
$C_{14}H_9Cl_2FN_4O_2S$
Molecular Weight: 387.21626
ALogP: 4.162
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.15
Enrichment: 0.499
Bayesian Score: -13.5
Mahalanobis Distance: 12.7
Mahalanobis Distance p-value: 0.000165

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Phenolphthalein	Lansoprazole
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.629	0.680	0.721
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

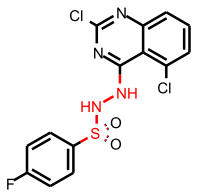
Model Applicability

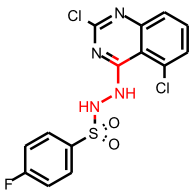
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374842118: [*]:n:[c](Cl):n:[*]

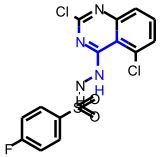
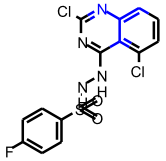
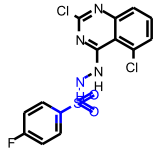
Feature Contribution

Top features for positive contribution

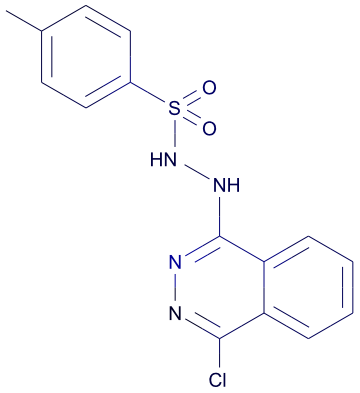
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-885461129	 [*]NNC(=[*])[*]	0.683	3 out of 3

FCFP_12	1294344583	 [*]NN[c](:[*]):[*]	0.174	1 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1151884458	 [*]N[c](:n:[*]):[c](:[*]):[*]	-1.11	0 out of 6
FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]):[c](:[*]):[*]	-0.994	0 out of 5
FCFP_12	-1096219292	 [*]NS(=O)(=O)[c](:[*]):[*]	-0.859	0 out of 4

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{15}H_{13}ClN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.149

Enrichment: 0.496

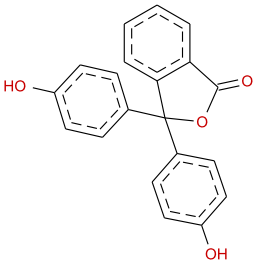
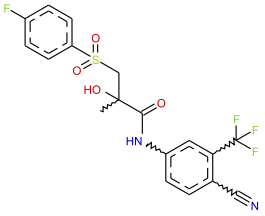
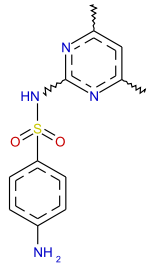
Bayesian Score: -11.3

Mahalanobis Distance: 17

Mahalanobis Distance p-value: 6.46e-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	Phenolphthalein	Bicalutamide	Sulfamethazine
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.657	0.667	0.723
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

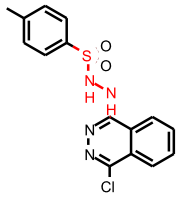
Model Applicability

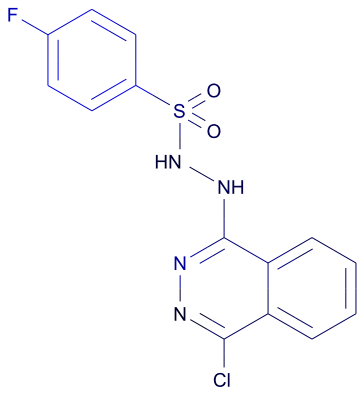
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374812327: [*]:n:[c](Cl):[c](:[*]):[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-885461129	 [*]NNC(=[*])[*]	0.683	3 out of 3



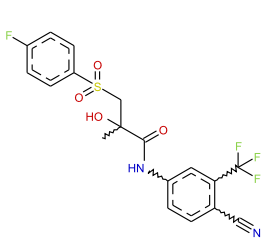
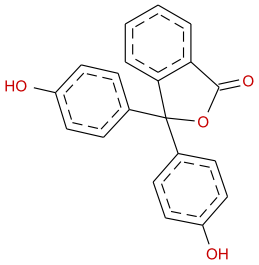
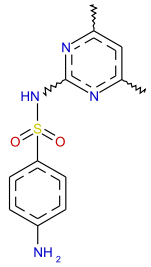
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.145
Enrichment: 0.483
Bayesian Score: -14
Mahalanobis Distance: 14.9
Mahalanobis Distance p-value: 2.99e-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	Bicalutamide	Phenolphthalein	Sulfamethazine
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.638	0.671	0.710
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

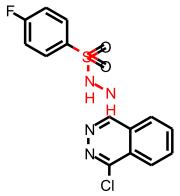
Model Applicability

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

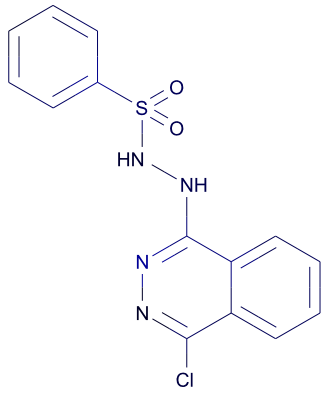
1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1374812327: [*]:n:[c](Cl):[c](:[*]):[*]

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-885461129	 [*]NNC(=[*])[*]	0.683	3 out of 3

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{14}H_{11}ClN_4O_2S$

Molecular Weight: 334.78073

ALogP: 3.043

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.148

Enrichment: 0.493

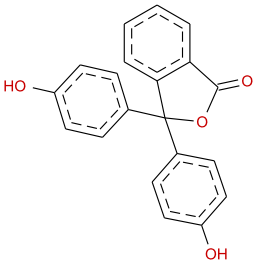
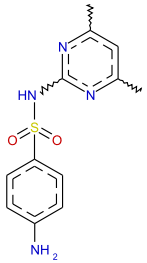
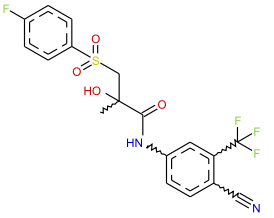
Bayesian Score: -11.1

Mahalanobis Distance: 17.6

Mahalanobis Distance p-value: 2.27e-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	Phenolphthalein	Sulfamethazine	Bicalutamide
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.667	0.673	0.681
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

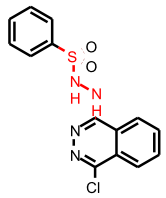
Model Applicability

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: -1374812327: [*]:n:[c](Cl):[c](:[*]):[*]

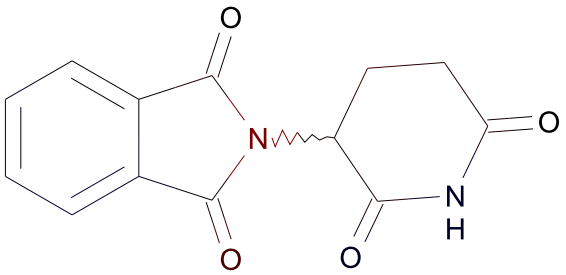
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-885461129	 [*]NNC(=[*])[*]	0.683	3 out of 3

Thalidomide

TOPKAT_Mouse_Male_FDA_Single_vs_Multiple



$C_{13}H_{10}N_2O_4$

Molecular Weight: 258.22949

ALogP: 9.7e-002

Rotatable Bonds: 1

Acceptors: 4

Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen

Probability: 0.219

Enrichment: 0.727

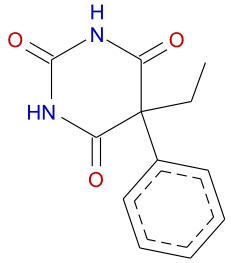
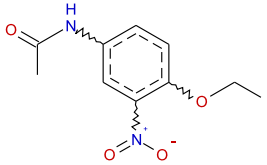
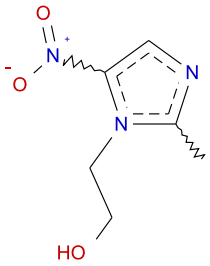
Bayesian Score: -1.62

Mahalanobis Distance: 8.2

Mahalanobis Distance p-value: 0.215

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

Structural Similar Compounds

Name	Phenobarbital	Nitroacetophenetide	Metronidazole
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.648	0.673	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

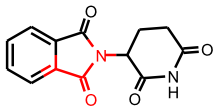
Model Applicability

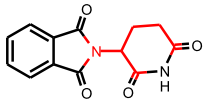
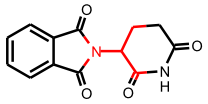
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

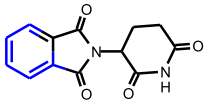
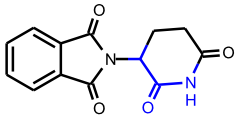
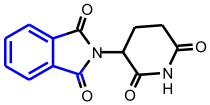
Feature Contribution

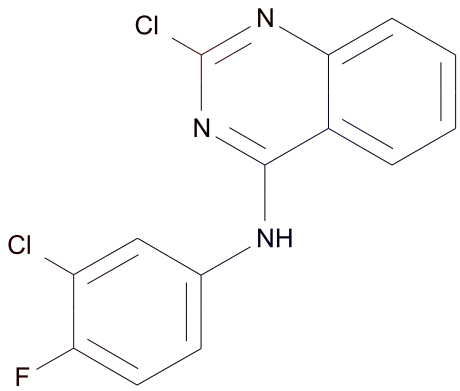
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	-1549163031	 [*]N1[*][*]:[c](:[*])C1=O	0.683	3 out of 3

FCFP_12	-989213044	 [*]N([*])C1CCC(=[*])[*]C1=[*]	0.679	5 out of 6
FCFP_12	-1043310069	 [*]CC(N([*])[*])C(=[*])[*]	0.597	7 out of 10

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.562	5 out of 28
FCFP_12	566058135	 [*]CC(=O)N[*]	-0.528	3 out of 17
FCFP_12	-1698724694	 [*]S(=[*])(=[*])[c]1: [cH]:[cH]:[cH]:[cH]: [cH]:1	-0.423	3 out of 15



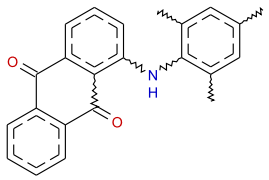
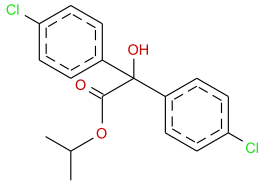
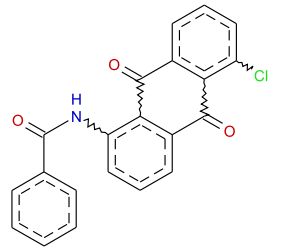
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Mild
Probability: 0.813
Enrichment: 1.18
Bayesian Score: -0.466
Mahalanobis Distance: 6.91
Mahalanobis Distance p-value: 0.999

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

Structural Similar Compounds

Name	ANTHRAQUINONE;1-(2;4;6-TRIMETHYLPHENYLAMINO)-	BENZILIC ACID; 4;4'-DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5-CHLORO-1-ANTHRAQUINONYL)-
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.552	0.574	0.608
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72

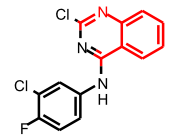
Model Applicability

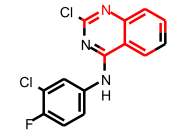
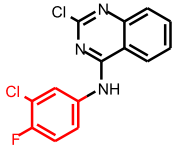
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

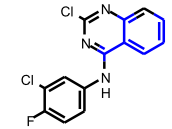
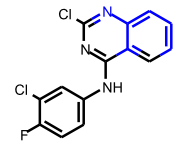
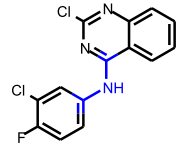
Feature Contribution

Top features for positive contribution

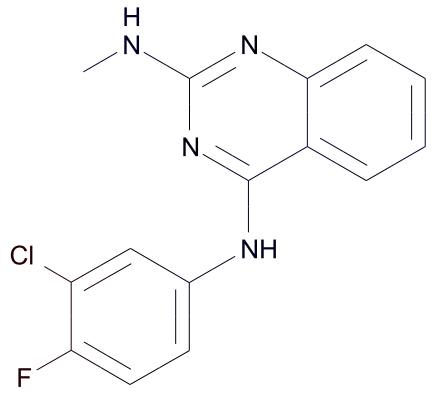
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	2047994594	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[cH]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3

FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	-2132756387	 [*][c]1:[cH]:[cH]:[c] (F):[c](Cl):[cH]:1	0.294	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1320007763	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	-0.316	19 out of 40
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.768

Enrichment: 1.11

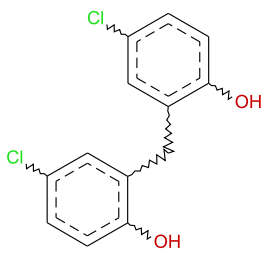
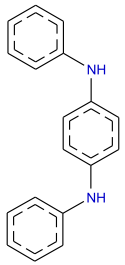
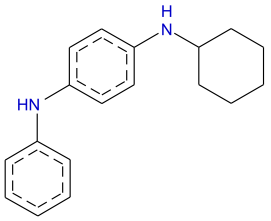
Bayesian Score: -2.04

Mahalanobis Distance: 6.81

Mahalanobis Distance p-value: 0.999

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

Structural Similar Compounds

Name	PHENOL;2;2'-METHYLENEBIS(4-CHLORO-	P-PHENYLENEDIAMINE;N;N'-DIPHENYL-	P-PHENYLENEDIAMINE;N-PHENYL-N'-CYCLOHEXYL-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.654	0.679	0.681
Reference	28ZPAK-;82;72	28ZPAK-;73;72	28ZPAK-;73;72

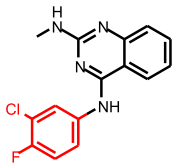
Model Applicability

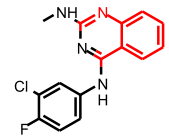
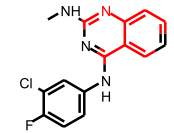
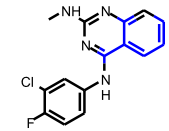
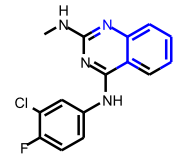
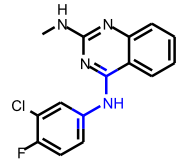
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

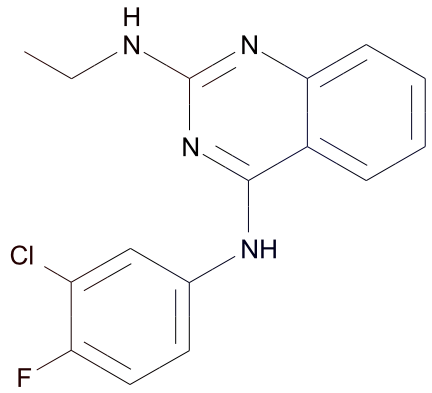
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.294	3 out of 3

FCFP_10	2047994594	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Moderate_Severe in training set**
FCFP_10	-1320007763	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	-0.316	19 out of 40
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{16}H_{14}ClFN_4$

Molecular Weight: 316.76056

ALogP: 5.025

Rotatable Bonds: 4

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.801

Enrichment: 1.16

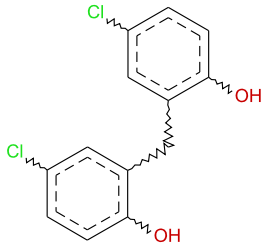
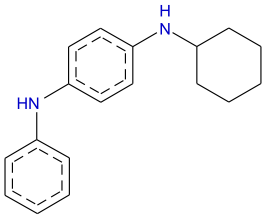
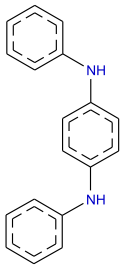
Bayesian Score: -0.984

Mahalanobis Distance: 7.95

Mahalanobis Distance p-value: 0.934

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

Structural Similar Compounds

Name	PHENOL;2;2'-METHYLENEBIS(4-CHLORO-	P-PHENYLENEDIAMINE; N-PHENYL-N'-CYCLOHEXYL-	P-PHENYLENEDIAMINE; N;N'-DIPHENYL-
Structure			
Actual Endpoint	Moderate_Severe	Mild	Mild
Predicted Endpoint	Moderate_Severe	Mild	Mild
Distance	0.674	0.681	0.682
Reference	28ZPAK-;82;72	28ZPAK-;73;72	28ZPAK-;73;72

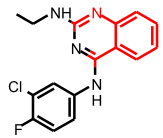
Model Applicability

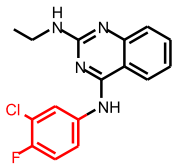
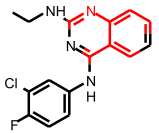
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

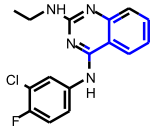
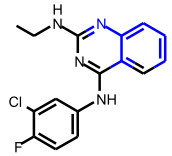
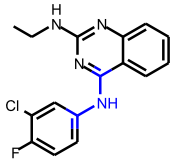
Feature Contribution

Top features for positive contribution

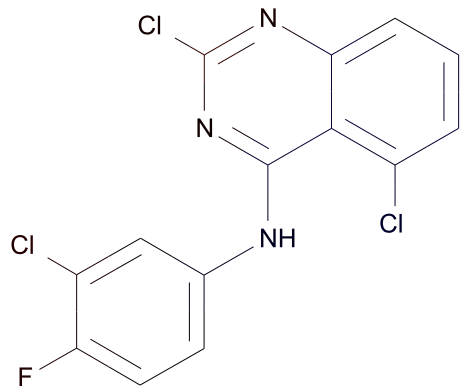
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	2047994594	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[cH]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3

FCFP_10	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.294	3 out of 3
FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1320007763	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	-0.316	19 out of 40
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Mild

Probability: 0.801

Enrichment: 1.16

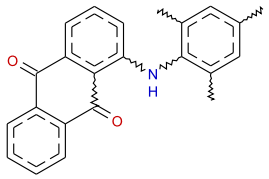
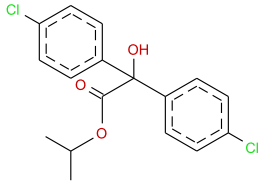
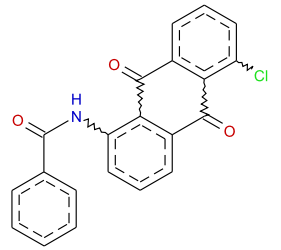
Bayesian Score: -0.954

Mahalanobis Distance: 7.04

Mahalanobis Distance p-value: 0.998

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

Structural Similar Compounds

Name	ANTHRAQUINONE;1-(2;4;6-TRIMETHYLPHENYLAMINO)-	BENZILIC ACID; 4;4'-DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5-CHLORO-1-ANTHRAQUINONYL)-
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.475	0.610	0.645
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72

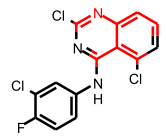
Model Applicability

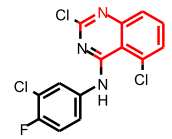
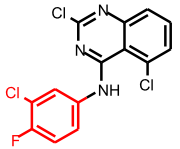
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

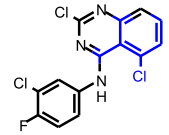
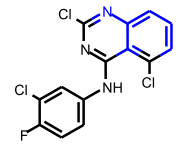
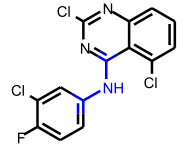
Feature Contribution

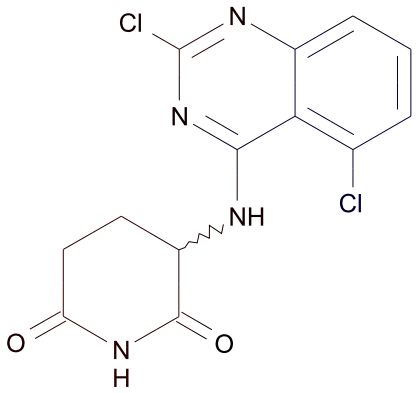
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3

FCFP_10	2047994594	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	-2132756387	 [*][c]1:[cH]:[cH]:[c] (F):[c](Cl):[cH]:1	0.294	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	 [*][c](:[*]):[c]1:[c] (Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.842	0 out of 2
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19



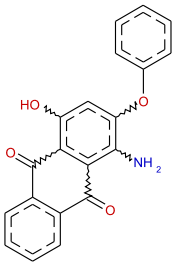
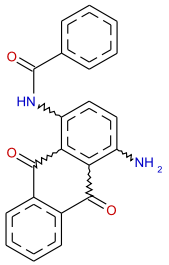
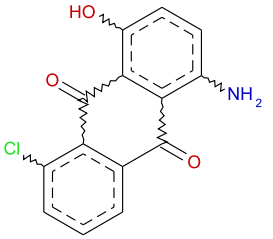
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Moderate_Severe
Probability: 0.825
Enrichment: 1.2
Bayesian Score: 0.195
Mahalanobis Distance: 8.96
Mahalanobis Distance p-value: 0.531

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

Structural Similar Compounds

Name	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	1-AMINO-4-HYDROXY-5-CHLORANTHRAQUINONE
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.562	0.570	0.597
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;83;72

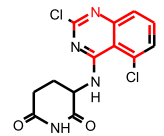
Model Applicability

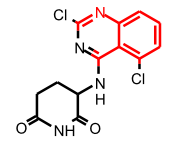
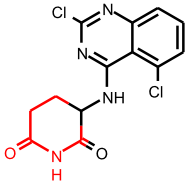
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

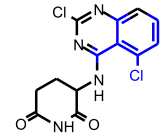
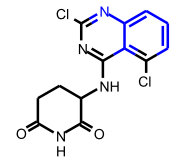
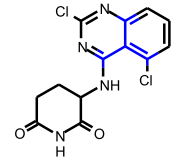
Feature Contribution

Top features for positive contribution

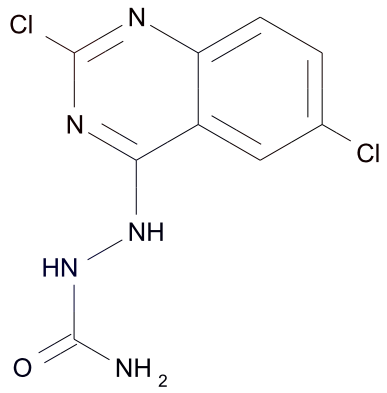
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3

FCFP_10	2047994594	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	-922480536	 [*]=C1[*]CCC(=O)N1	0.256	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	 [*][c](:[*]):[c]1:[c] (Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.842	0 out of 2
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	-0.29	21 out of 43

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09078

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Mild

Probability: 0.806

Enrichment: 1.17

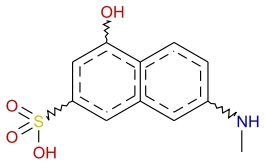
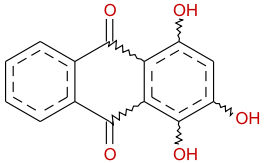
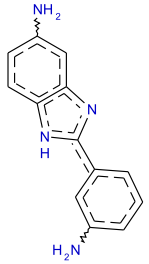
Bayesian Score: -0.782

Mahalanobis Distance: 6.62

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	2-NAPHTHALENESULFONIC ACID; 4-HYDROXY-7-(METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE	Benzimidazole; 6-amino-2-(3'-aminophenyl)-; dihydrochloride
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.583	0.594	0.613
Reference	28ZPAK 190;72	28ZPAK-;103;7	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;831;86

Model Applicability

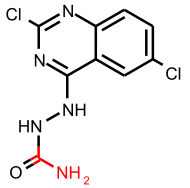
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

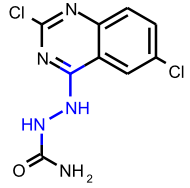
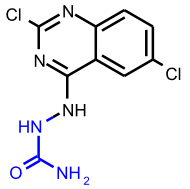
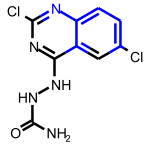
Feature Contribution

Top features for positive contribution

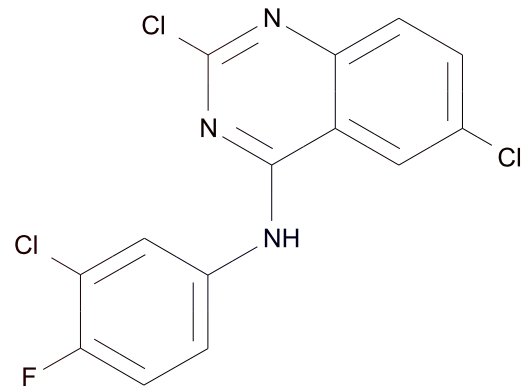
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-745491832	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	0.304	29 out of 32

FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	1070061035	 [*]C(=[*])N	0.239	284 out of 338

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1294344583	 [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	1499521844	 [*]NC(=O)N	-0.4	1 out of 3
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Moderate_Severe

Probability: 0.823

Enrichment: 1.19

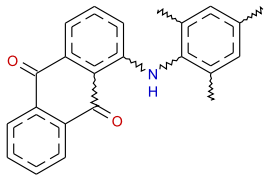
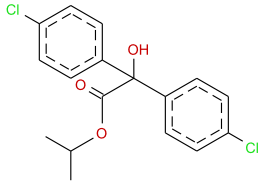
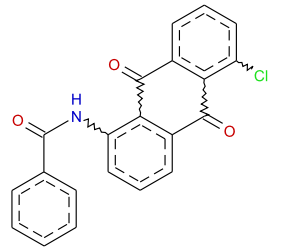
Bayesian Score: 0.0833

Mahalanobis Distance: 7.04

Mahalanobis Distance p-value: 0.998

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

Structural Similar Compounds

Name	ANTHRAQUINONE;1-(2;4;6-TRIMETHYLPHENYLAMINO)-	BENZILIC ACID; 4;4'-DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5-CHLORO-1-ANTHRAQUINONYL)-
Structure			
Actual Endpoint	Moderate_Severe	Moderate_Severe	Mild
Predicted Endpoint	Mild	Moderate_Severe	Mild
Distance	0.481	0.604	0.650
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72

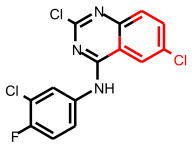
Model Applicability

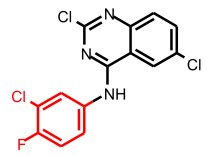
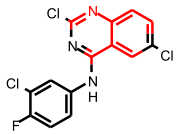
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

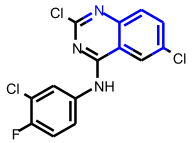
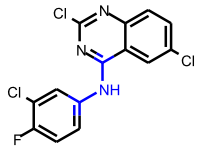
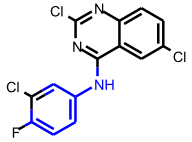
Feature Contribution

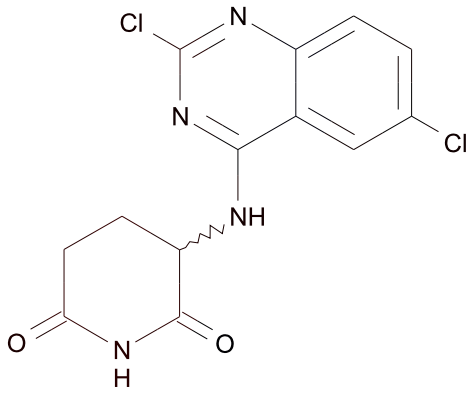
Top features for positive contribution

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

FCFP_10	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.294	3 out of 3
FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	1293778554	 [*]:[c](:[*])N[c](:[*]):[*]	-0.304	9 out of 19
FCFP_10	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.294	50 out of 102



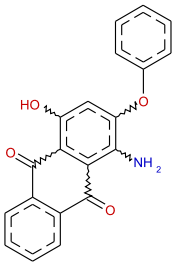
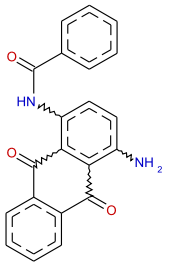
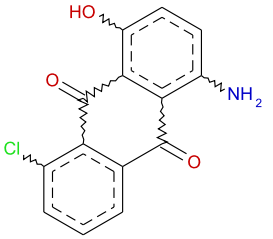
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Moderate_Severe
Probability: 0.838
Enrichment: 1.22
Bayesian Score: 1.23
Mahalanobis Distance: 8.96
Mahalanobis Distance p-value: 0.531

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

Structural Similar Compounds

Name	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	1-AMINO-4-HYDROXY-5-CHLORANTHRAQUINONE
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.567	0.576	0.602
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;83;72

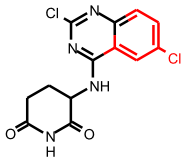
Model Applicability

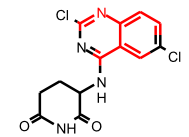
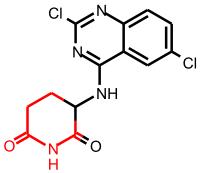
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

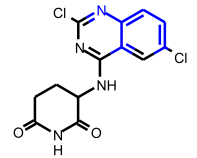
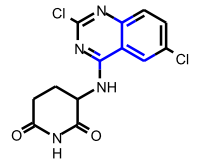
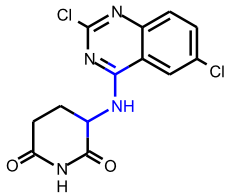
Feature Contribution

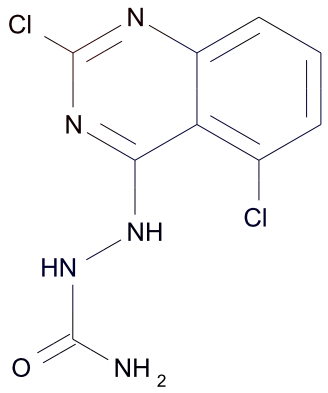
Top features for positive contribution

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

FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	-922480536	 [*]=C1[*]CCC(=O)N1	0.256	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.307	8 out of 17
FCFP_10	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	-0.29	21 out of 43
FCFP_10	1294255210	 [*]:[c](:[*])NC	-0.218	20 out of 38



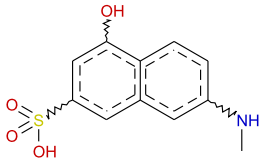
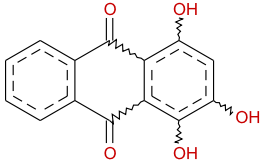
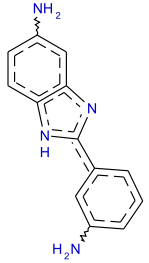
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Mild
Probability: 0.776
Enrichment: 1.13
Bayesian Score: -1.82
Mahalanobis Distance: 6.62
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	2-NAPHTHALENESULFONIC ACID; 4-HYDROXY-7-(METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE	Benzimidazole; 6-amino-2-(3'-aminophenyl)-; dihydrochloride
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.583	0.588	0.607
Reference	28ZPAK 190;72	28ZPAK-;103;7	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;831;86

Model Applicability

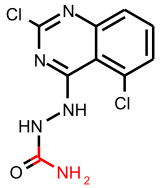
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

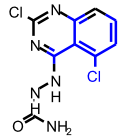
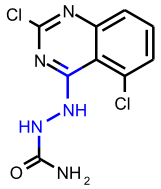
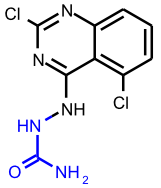
Feature Contribution

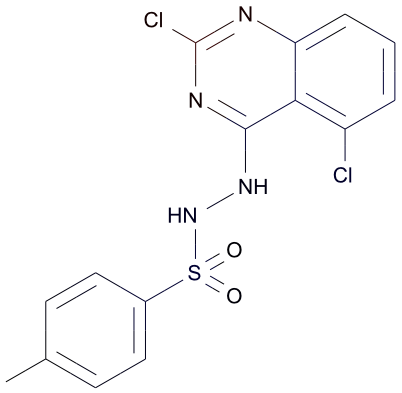
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3

FCFP_10	2047994594	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	1070061035	 [*]C(=[*])N	0.239	284 out of 338

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	 [*][c](:[*]):[c]1:[c] (Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.842	0 out of 2
FCFP_10	1294344583	 [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	1499521844	 [*]NC(=O)N	-0.4	1 out of 3



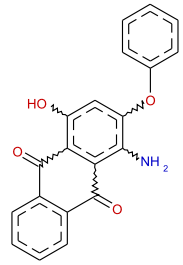
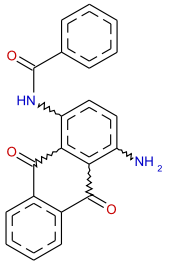
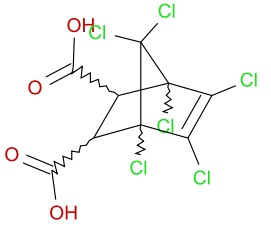
$C_{15}H_{12}Cl_2N_4O_2S$
Molecular Weight: 383.25237
ALogP: 4.443
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild
Probability: 0.772
Enrichment: 1.12
Bayesian Score: -1.92
Mahalanobis Distance: 5.56
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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.617	0.669	0.705
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72

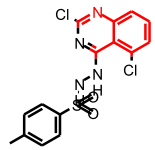
Model Applicability

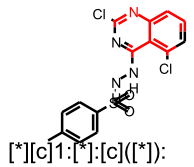
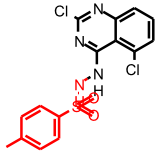
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

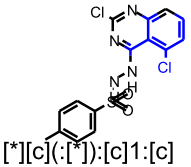
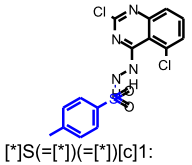
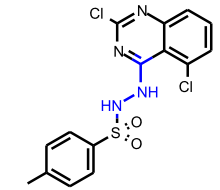
Feature Contribution

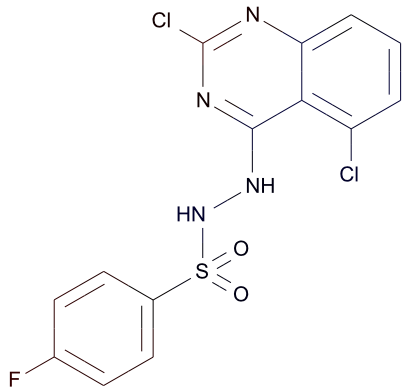
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	2047994594	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	0.294	3 out of 3

FCFP_10	-1716224640	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.294	3 out of 3
FCFP_10	1789442672	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.256	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1237732457	 [*][c](:[*]):[c]1:[c](Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.842	0 out of 2
FCFP_10	632767364	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.6	1 out of 4
FCFP_10	1294344583	 [*]NN[c](:[*]):[*]	-0.507	0 out of 1



$C_{14}H_9Cl_2FN_4O_2S$

Molecular Weight: 387.21626

ALogP: 4.162

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.802

Enrichment: 1.16

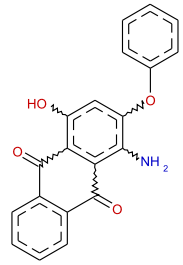
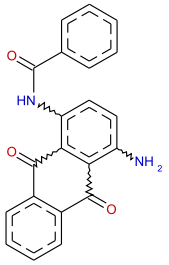
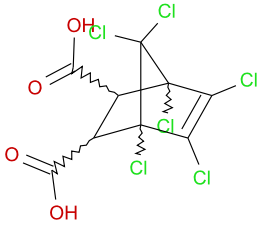
Bayesian Score: -0.921

Mahalanobis Distance: 5.48

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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.601	0.650	0.682
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72

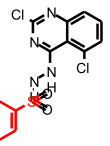
Model Applicability

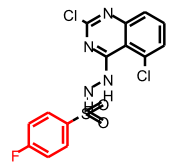
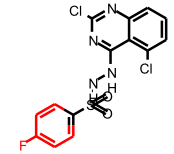
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

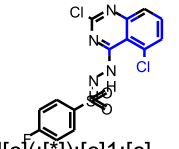
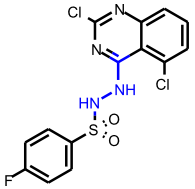
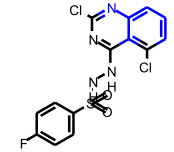
Feature Contribution

Top features for positive contribution

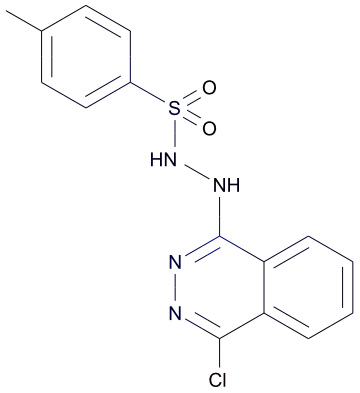
Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-149636017	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.352	7 out of 7

FCFP_10	-1508180856	 [*][c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.329	16 out of 17
FCFP_10	-745491832	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[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	1237732457	 [*][c](:[*]):[c]1:[c](Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.842	0 out of 2
FCFP_10	1294344583	 [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	713358128	 [*]:n:[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	-0.307	8 out of 17

TOPKAT_Ocular_Irritancy_Mild_vs_Moderate_Severe



$C_{15}H_{13}ClN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Mild

Probability: 0.73

Enrichment: 1.06

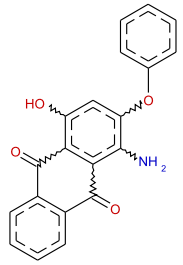
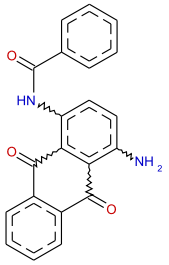
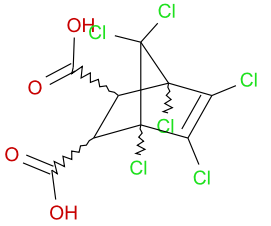
Bayesian Score: -2.95

Mahalanobis Distance: 5.53

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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.528	0.591	0.651
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72

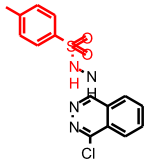
Model Applicability

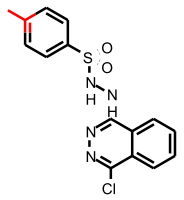
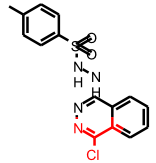
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

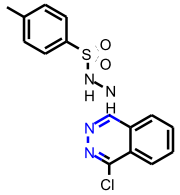
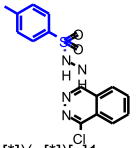
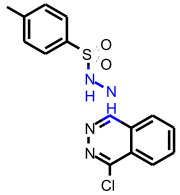
Feature Contribution

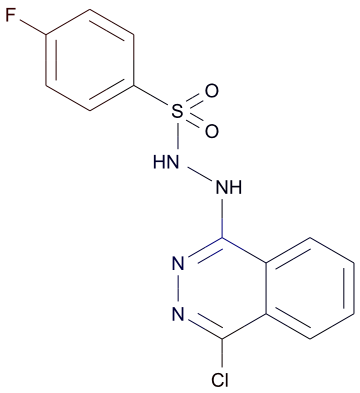
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	1789442672	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0.256	2 out of 2

FCFP_10	136120670	 [*]:[c](:[*])C	0.206	53 out of 65
FCFP_10	-1374812327	 [*]:n:[c](Cl):[c](:[*]):[*]	0.186	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	 [*][c](:[*]):n:n:[*]	-1.29	0 out of 4
FCFP_10	632767364	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.6	1 out of 4
FCFP_10	1294344583	 [*]NN[c](:[*]):[*]	-0.507	0 out of 1



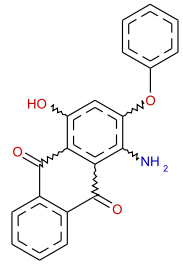
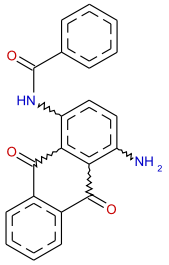
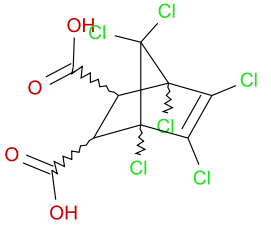
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild
Probability: 0.771
Enrichment: 1.12
Bayesian Score: -1.97
Mahalanobis Distance: 5.46
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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Mild	Mild	Moderate_Severe
Predicted Endpoint	Mild	Mild	Moderate_Severe
Distance	0.534	0.583	0.638
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72

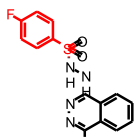
Model Applicability

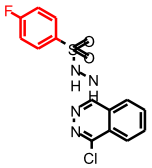
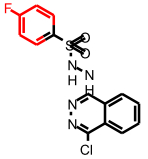
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

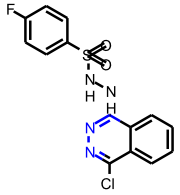
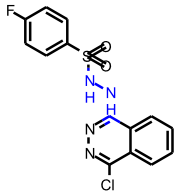
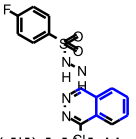
Feature Contribution

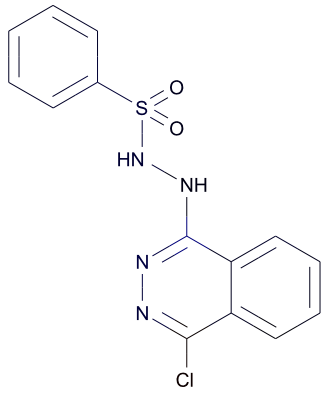
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-149636017	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.352	7 out of 7

FCFP_10	-1508180856	 [*][c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.329	16 out of 17
FCFP_10	-745491832	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[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	4427049	 [*][c](:[*]):n:n:[*]	-1.29	0 out of 4
FCFP_10	1294344583	 [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	-1320007763	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	-0.316	19 out of 40



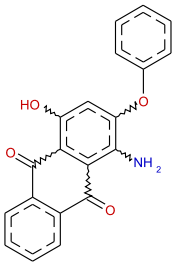
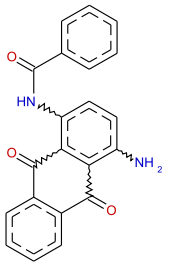
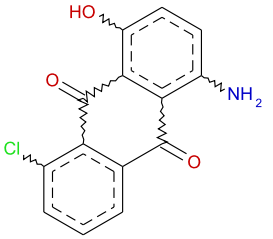
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Mild
Probability: 0.704
Enrichment: 1.02
Bayesian Score: -3.46
Mahalanobis Distance: 5.5
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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	1-AMINO-4-HYDROXY-5-CHLORANTHRAQUINONE
Structure			
Actual Endpoint	Mild	Mild	Mild
Predicted Endpoint	Mild	Mild	Mild
Distance	0.525	0.570	0.635
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;83;72

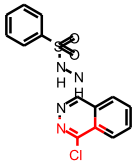
Model Applicability

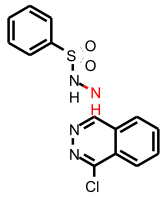
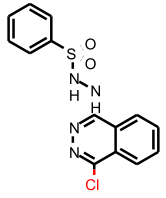
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

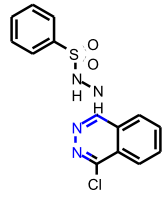
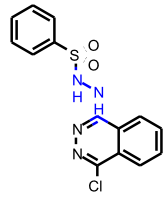
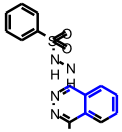
Feature Contribution

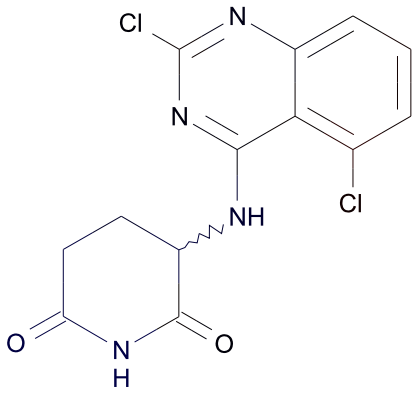
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	-1374812327	 [*]:n:[c](Cl):[c](:[*]):[*]	0.186	1 out of 1

FCFP_10	3	 [*]N[*]	0.165	383 out of 491
FCFP_10	32	 [*]Cl	0.0821	145 out of 202

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Moderate_Severe in training set
FCFP_10	4427049	 [*][c](:[*]):n:n:[*]	-1.29	0 out of 4
FCFP_10	1294344583	 [*]NN[c](:[*]):[*]	-0.507	0 out of 1
FCFP_10	-1320007763	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	-0.316	19 out of 40



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Moderate

Probability: 0.581

Enrichment: 0.938

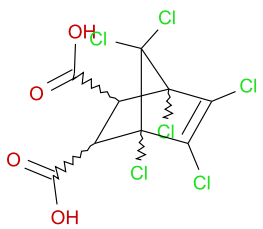
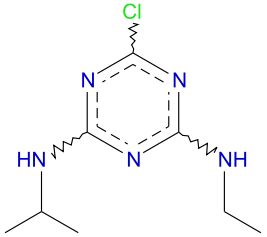
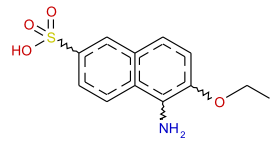
Bayesian Score: -2.63

Mahalanobis Distance: 10.3

Mahalanobis Distance p-value: 0.0189

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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-	s-TRIAZINE; 2-CHLORO-4-ETHYLAMINO-6-ISOPROPYLAMINO-	2-NAPHTHALENESULFONIC ACID;5-AMINO-6-ETHOXY-
Structure			
Actual Endpoint	Severe	Severe	Moderate
Predicted Endpoint	Severe	Moderate	Moderate
Distance	0.624	0.644	0.660
Reference	28ZPAK-;92;72	CIGET* -;-;77	28ZPAK-;191;72

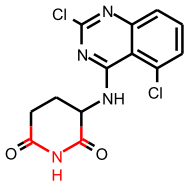
Model Applicability

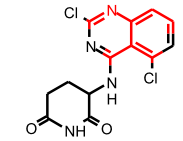
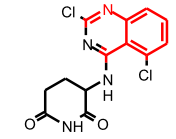
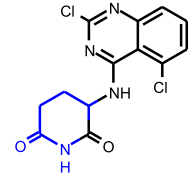
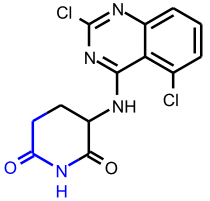
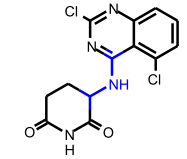
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

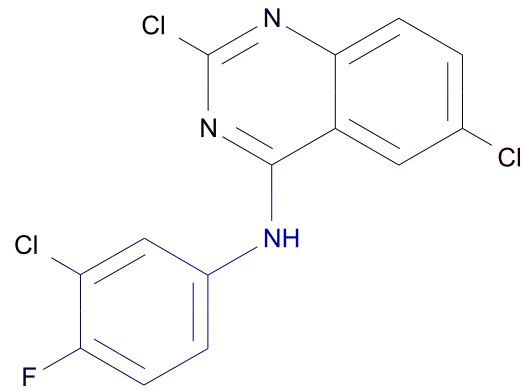
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1631785938	 [*]C(=[*])NC(=[*])[*]	0.218	1 out of 1

SCFP_12	622342378	 [*][c]1:[*]:[c]([*]):[c]2:[c]([*]):[*]:[cH]:[cH]:[c]:2:n:1	0.213	4 out of 5
SCFP_12	403834996	 [*][c]1:[*]:[c]([*]):[c]2:[c]([*]):[cH]:[cH]:[cH]:[c]:2:n:1	0.213	4 out of 5
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Severe in training set**
SCFP_12	-1072897324	 [*]C1[*]NC(=O)CC1	-0.8	3 out of 13
SCFP_12	1256995004	 [*]CC(=O)N[*]	-0.483	12 out of 33
SCFP_12	18117904	 [*]C([*])N[c](:[*]):[*]	-0.481	3 out of 9

TOPKAT_Ocular_Irritancy_Moderate_vs_Severe



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

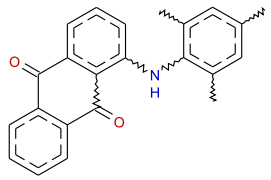
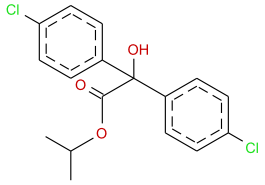
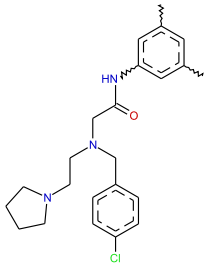
ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Structural Similar Compounds

Name	ANTHRAQUINONE;1-(2;4;6-TRIMETHYLPHENYLAMINO)-	BENZILIC ACID; 4;4'-DICHLORO-; ISOPROPYL ESTER	2';6'-Acetoxylidide; 2-(p-chlorobenzyl(2-(pyrrolidinyl)ethyl)amino)-;
Structure			
Actual Endpoint	Moderate	Severe	Severe
Predicted Endpoint	Moderate	Severe	Severe
Distance	0.484	0.615	0.704
Reference	28ZPAK-;242;72	CIGET* -;-;77	Arzneimittel-Forschung 9;167;59

Model Prediction

Prediction: Moderate

Probability: 0.479

Enrichment: 0.772

Bayesian Score: -4.46

Mahalanobis Distance: 5.84

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.

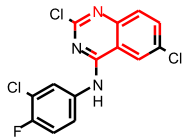
Model Applicability

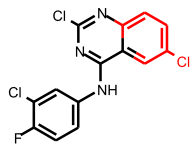
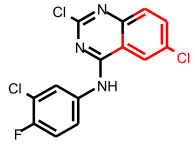
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

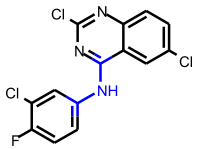
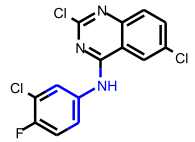
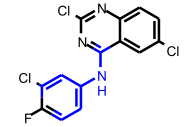
Feature Contribution

Top features for positive contribution

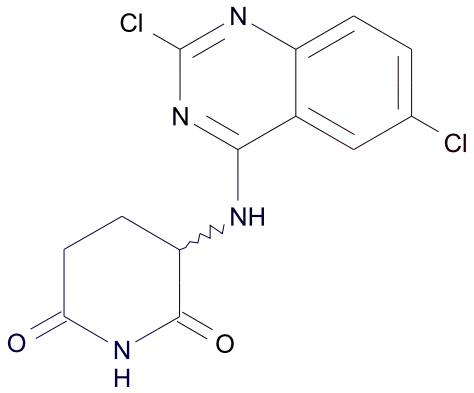
Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	622342378	 [*][c]1:[*]:[c]([*]):[c]2:[c]([*]):[*]:[cH]:[cH]:[c]:2:n:1	0.213	4 out of 5

SCFP_12	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	0.211	22 out of 29
SCFP_12	-116109291	 [*]:[c]1:[*]:[cH]:[cH]:[c](Cl):[cH]:1	0.174	19 out of 26

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	951581613	 [*]:[c](:[*])N[c](:[*]):[*]	-0.769	2 out of 9
SCFP_12	1334669481	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.685	28 out of 93
SCFP_12	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.561	2 out of 7

TOPKAT_Ocular_Irritancy_Moderate_vs_Severe



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Moderate

Probability: 0.579

Enrichment: 0.934

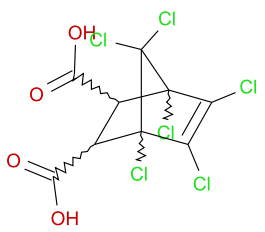
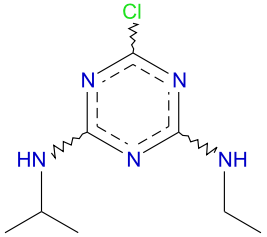
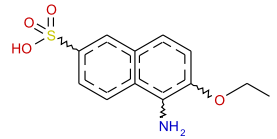
Bayesian Score: -2.67

Mahalanobis Distance: 10.3

Mahalanobis Distance p-value: 0.0189

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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-	s-TRIAZINE; 2-CHLORO-4-ETHYLAMINO-6-ISOPROPYLAMINO-	2-NAPHTHALENESULFONIC ACID;5-AMINO-6-ETHOXY-
Structure			
Actual Endpoint	Severe	Severe	Moderate
Predicted Endpoint	Severe	Moderate	Moderate
Distance	0.624	0.645	0.660
Reference	28ZPAK-;92;72	CIGET* -;-;77	28ZPAK-;191;72

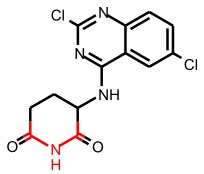
Model Applicability

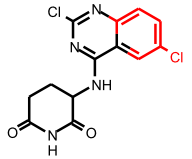
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

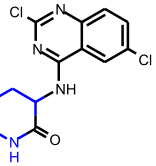
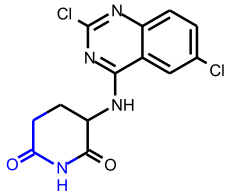
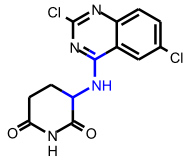
Feature Contribution

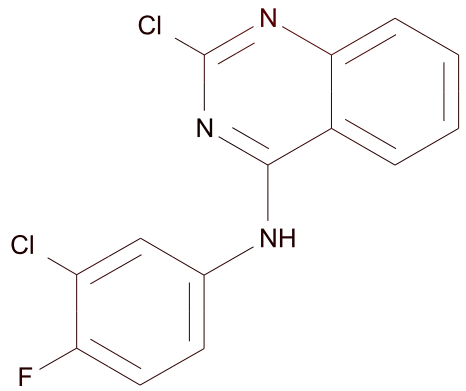
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	1631785938	 [*]C(=[*])NC(=[*])[*]	0.218	1 out of 1

SCFP_12	622342378	 [*][c]1:[*]:[c]([*]):[c]2:[c]([*]):[*]:[cH]:[cH]:[c]:2:n:1	0.213	4 out of 5
SCFP_12	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	0.211	22 out of 29

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Severe in training set
SCFP_12	-1072897324	 [*]C1[*]NC(=O)CC1	-0.8	3 out of 13
SCFP_12	1256995004	 [*]CC(=O)N[*]	-0.483	12 out of 33
SCFP_12	18117904	 [*]C([*])N[c](:[*]):[*]	-0.481	3 out of 9



$C_{14}H_8Cl_2FN_3$

Molecular Weight: 308.13782

ALogP: 5.202

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

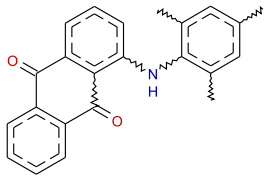
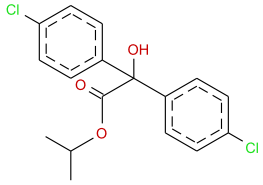
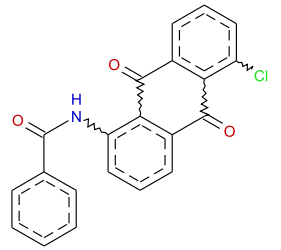
Bayesian Score: 3.13

Mahalanobis Distance: 6.09

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	ANTHRAQUINONE;1-(2;4;6-TRIMETHYLPHENYLAMINO)-	BENZILIC ACID; 4;4'-DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5-CHLORO-1-ANTHRAQUINONYL)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.545	0.563	0.599
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72

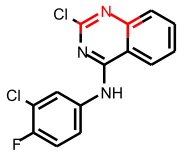
Model Applicability

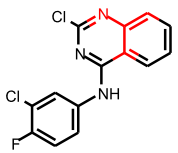
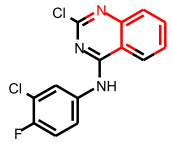
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

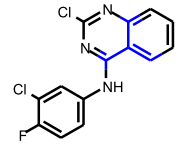
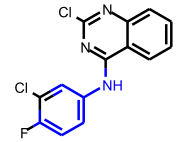
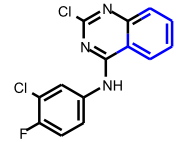
Feature Contribution

Top features for positive contribution

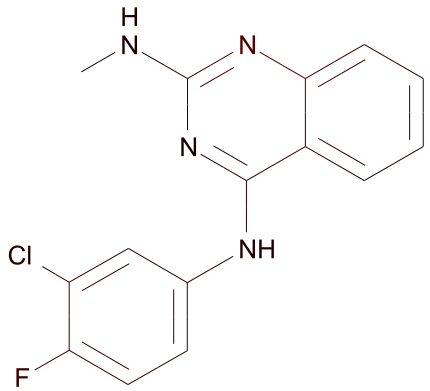
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

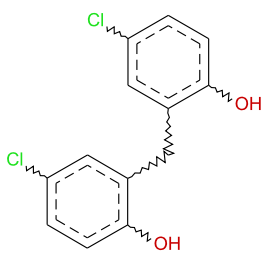
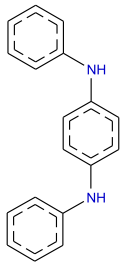
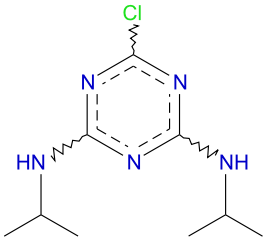
Bayesian Score: 3.49

Mahalanobis Distance: 4.96

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	PHENOL;2;2'-METHYLENEBIS(4-CHLORO-	P-PHENYLENEDIAMINE; N;N'-DIPHENYL-	s-TRIAZINE; 2-CHLORO-4;6-BIS(ISOPROPYLAMINO)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.647	0.674	0.675
Reference	28ZPAK-;82;72	28ZPAK-;73;72	CIGET 77

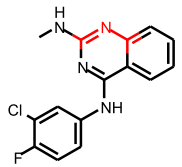
Model Applicability

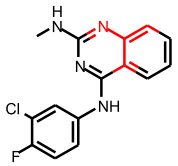
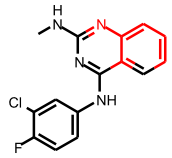
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

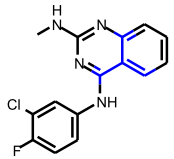
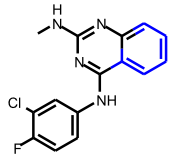
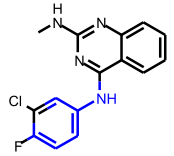
Feature Contribution

Top features for positive contribution

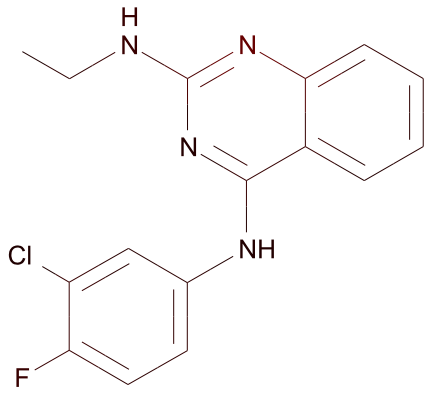
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{16}H_{14}ClFN_4$

Molecular Weight: 316.76056

ALogP: 5.025

Rotatable Bonds: 4

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

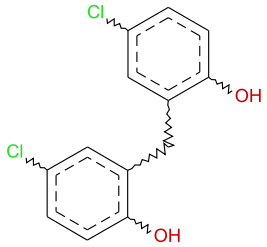
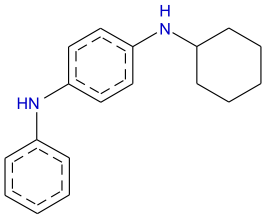
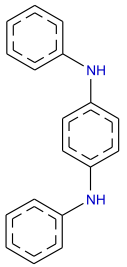
Bayesian Score: 3.53

Mahalanobis Distance: 5.94

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	PHENOL;2;2'-METHYLENEBIS(4-CHLORO-	P-PHENYLENEDIAMINE;N-PHENYL-N'-CYCLOHEXYL-	P-PHENYLENEDIAMINE;N;N'-DIPHENYL-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.665	0.677	0.677
Reference	28ZPAK-;82;72	28ZPAK-;73;72	28ZPAK-;73;72

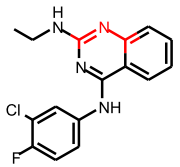
Model Applicability

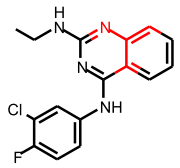
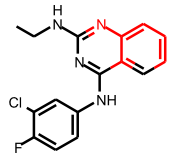
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

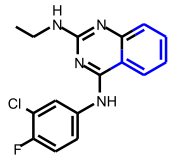
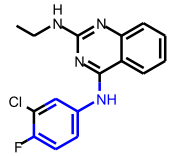
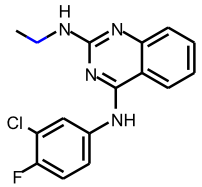
Feature Contribution

Top features for positive contribution

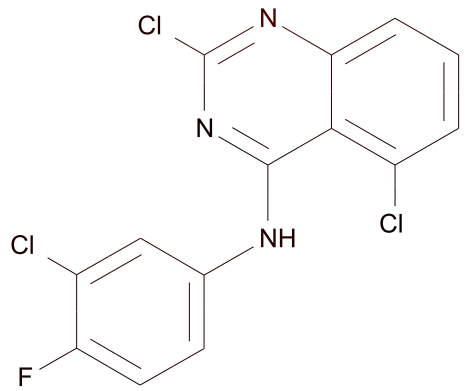
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121
FCFP_12	0	 [*]C	0	1184 out of 1397

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

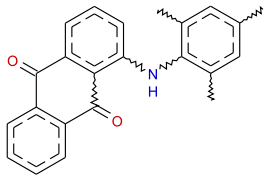
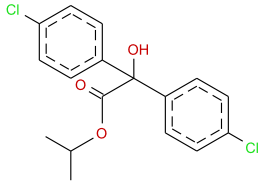
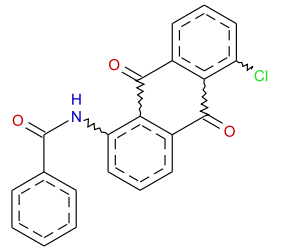
Bayesian Score: 2.64

Mahalanobis Distance: 6.25

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	ANTHRAQUINONE;1-(2;4;6-TRIMETHYLPHENYLAMINO)-	BENZILIC ACID; 4;4'-DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5-CHLORO-1-ANTHRAQUINONYL)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.476	0.593	0.628
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72

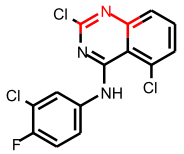
Model Applicability

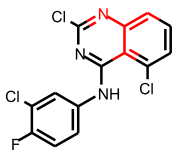
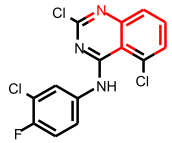
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

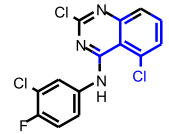
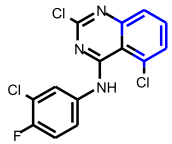
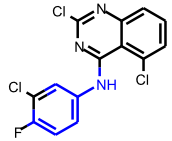
Feature Contribution

Top features for positive contribution

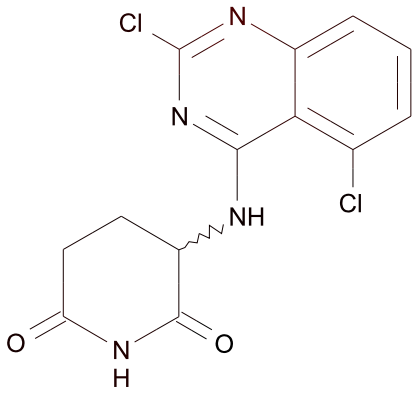
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	 [*][c](:[*]):[c]1:[c] (Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.132	2 out of 3
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

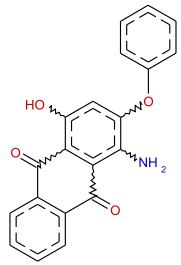
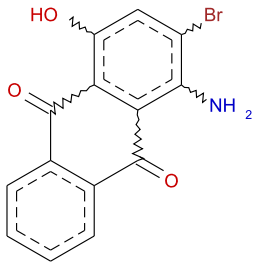
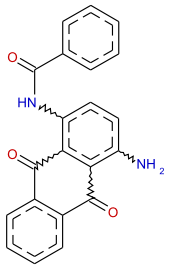
Bayesian Score: 2.85

Mahalanobis Distance: 6.88

Mahalanobis Distance p-value: 0.999

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

Structural Similar Compounds

Name	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	Anthraquinone; 1-amino-2-bromo-4-hydroxy-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE
Structure			
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.552	0.562	0.567
Reference	28ZPAK 239;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 535;86	28ZPAK-;124;72

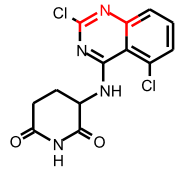
Model Applicability

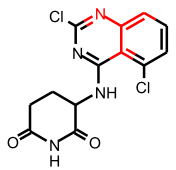
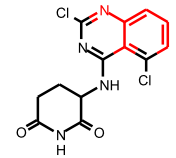
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

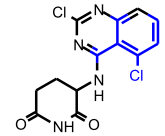
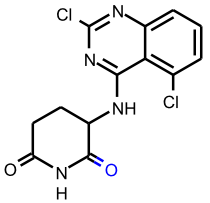
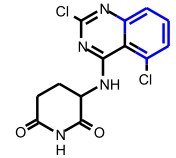
Feature Contribution

Top features for positive contribution

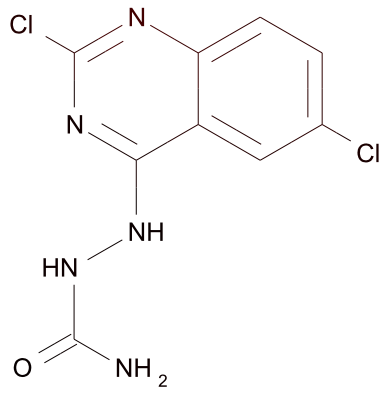
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	 [*][c](:[*]):[c]1:[c] (Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.132	2 out of 3
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09078

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

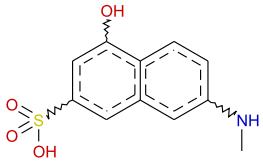
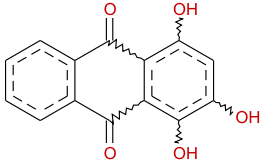
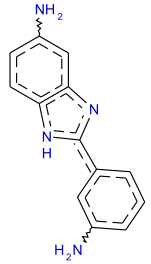
Bayesian Score: 2.46

Mahalanobis Distance: 4.47

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	2-NAPHTHALENESULFONIC ACID; 4-HYDROXY-7-(METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE	Benzimidazole; 6-amino-2-(3'-aminophenyl)-; dihydrochloride
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.575	0.588	0.605
Reference	28ZPAK 190;72	28ZPAK-;103;7	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;831;86

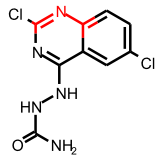
Model Applicability

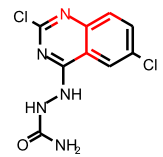
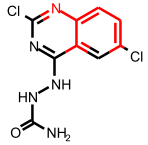
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

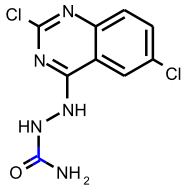
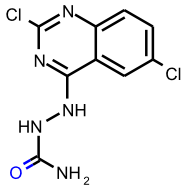
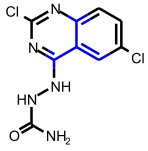
Feature Contribution

Top features for positive contribution

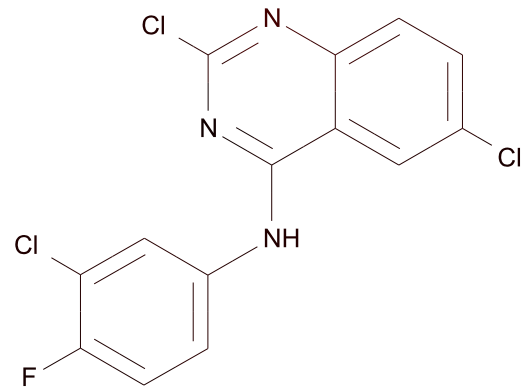
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	0	 [*]C	0	1184 out of 1397
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

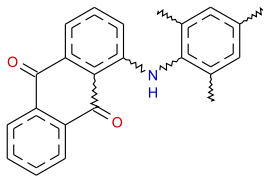
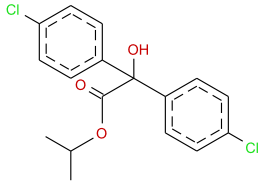
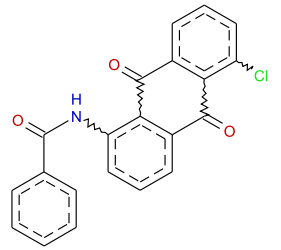
Bayesian Score: 2.8

Mahalanobis Distance: 6.25

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	ANTHRAQUINONE;1-(2;4;6-TRIMETHYLPHENYLAMINO)-	BENZILIC ACID; 4;4'-DICHLORO-; ISOPROPYL ESTER	BENZAMIDE; N-(5-CHLORO-1-ANTHRAQUINONYL)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.481	0.587	0.633
Reference	28ZPAK-;242;72	CIGET* -;-;77	28ZPAK 89;72

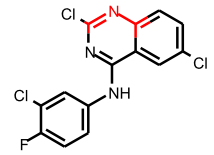
Model Applicability

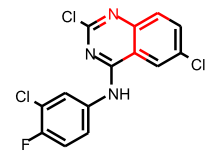
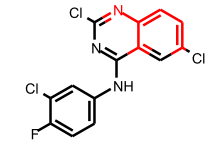
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

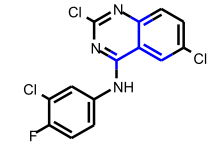
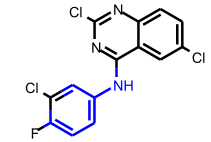
Feature Contribution

Top features for positive contribution

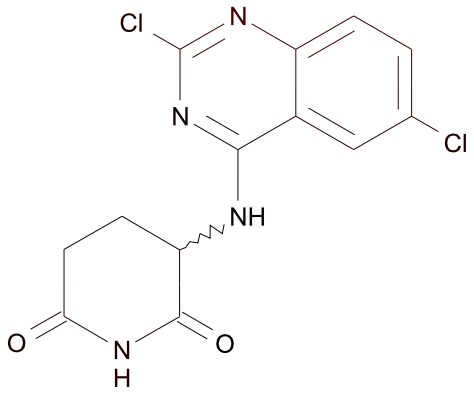
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c] ([*]):[cH]:[cH]:1	0	102 out of 121

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

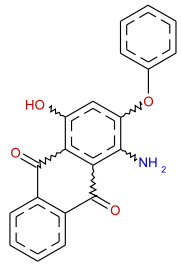
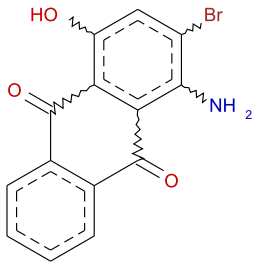
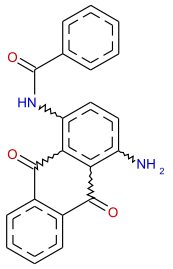
Bayesian Score: 3

Mahalanobis Distance: 6.88

Mahalanobis Distance p-value: 0.999

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

Structural Similar Compounds

Name	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	Anthraquinone; 1-amino-2-bromo-4-hydroxy-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE
Structure			
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Irritant
Distance	0.556	0.568	0.572
Reference	28ZPAK 239;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 535;86	28ZPAK-;124;72

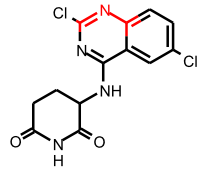
Model Applicability

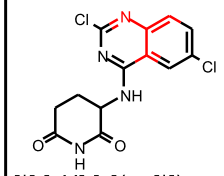
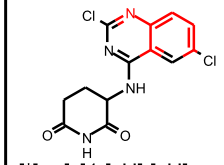
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]

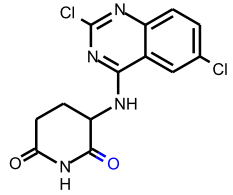
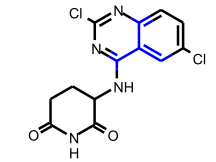
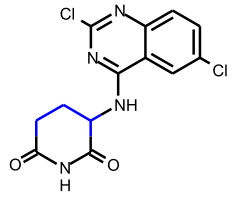
Feature Contribution

Top features for positive contribution

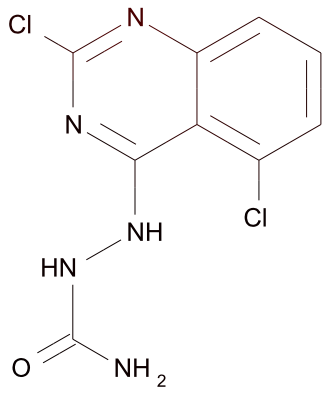
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52
FCFP_12	-1272798659	 [*]CCC([*])[*]	0	517 out of 643

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09077

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

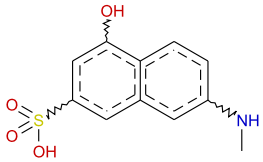
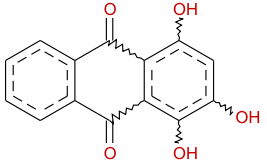
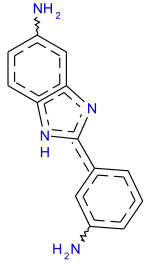
Bayesian Score: 2.3

Mahalanobis Distance: 4.47

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	2-NAPHTHALENESULFONIC ACID; 4-HYDROXY-7-(METHYLAMINO)-	1;2;4-TRIHYDROXY ANTHRAQUINONE	Benzimidazole; 6-amino-2-(3'-aminophenyl)-; dihydrochloride
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.574	0.582	0.599
Reference	28ZPAK 190;72	28ZPAK-;103;7	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. -;831;86

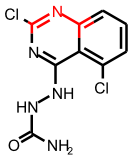
Model Applicability

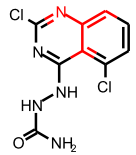
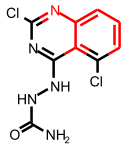
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

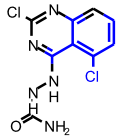
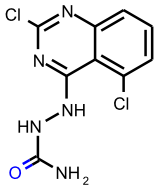
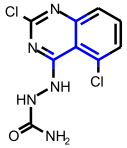
Feature Contribution

Top features for positive contribution

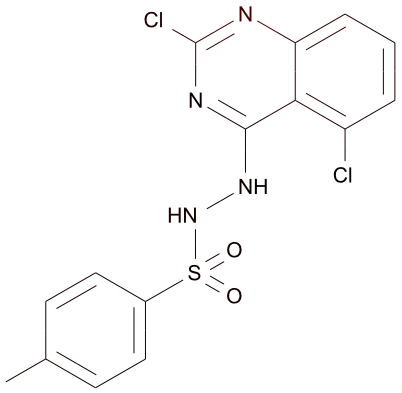
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]):[c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	 [*][c](:[*]):[c]1:[c](Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.132	2 out of 3
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{15}H_{12}Cl_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

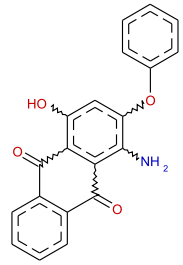
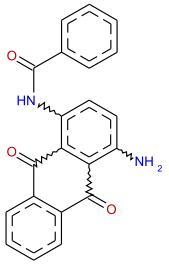
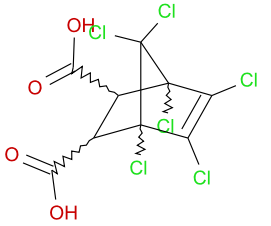
Bayesian Score: 2.72

Mahalanobis Distance: 3.95

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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.605	0.651	0.681
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72

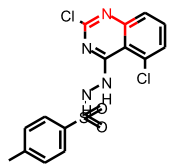
Model Applicability

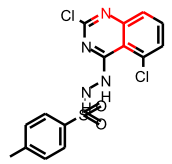
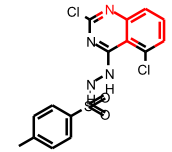
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

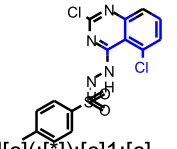
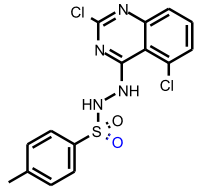
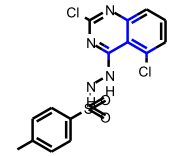
Feature Contribution

Top features for positive contribution

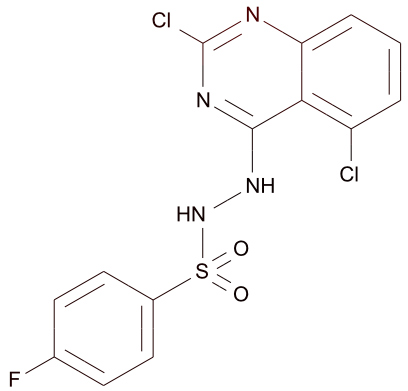
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	 [*][c](:[*]):[c]1:[c] (Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.132	2 out of 3
FCFP_12	1	 [*]=O	0	872 out of 1051
FCFP_12	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0	43 out of 52

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{14}H_9Cl_2FN_4O_2S$

Molecular Weight: 387.21626

ALogP: 4.162

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

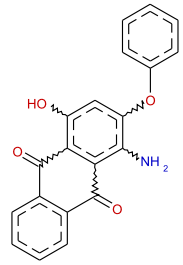
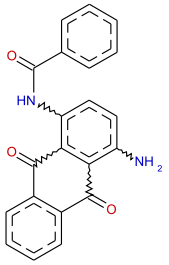
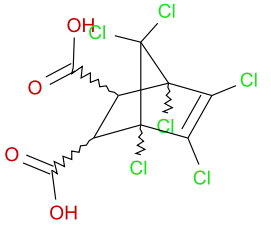
Bayesian Score: 2.81

Mahalanobis Distance: 3.89

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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	5-NORBORNENE-2;3-DICARBOXYLIC ACID; 1;4;5;6;7;7-HEXACHLORO-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.592	0.635	0.662
Reference	28ZPAK 239;72	28ZPAK-;124;72	28ZPAK-;92;72

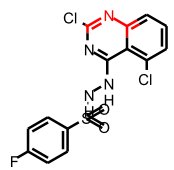
Model Applicability

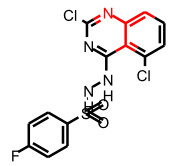
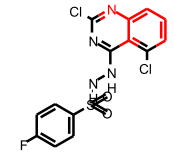
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

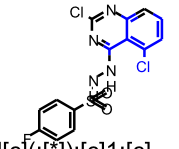
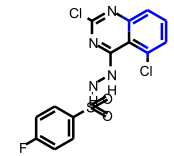
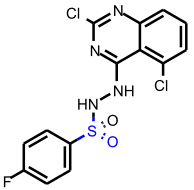
Feature Contribution

Top features for positive contribution

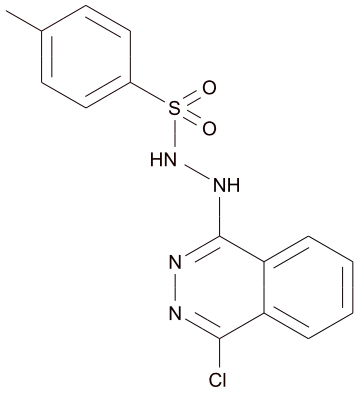
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1747237384	 [*][c](:[*]):n:[c](:[*]):[*]	0.208	44 out of 44

FCFP_12	178336375	 [*]:[cH]:[c](:n:[*]): [c](:[*]):[*]	0.202	19 out of 19
FCFP_12	713358128	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	0.2	17 out of 17

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1237732457	 [*][c](:[*]):[c]1:[c] (Cl):[cH]:[cH]:[*]:[c]:1:[*]	-0.132	2 out of 3
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	1872154524	 [*]C(=O)[*]	0	563 out of 690

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{15}H_{13}ClN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

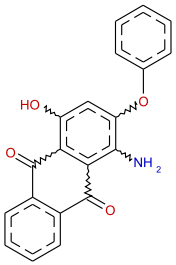
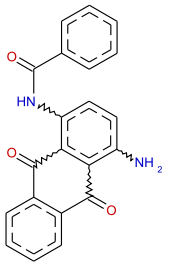
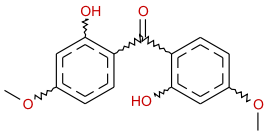
Bayesian Score: 1.69

Mahalanobis Distance: 3.74

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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	2;2';-Dihydroxy-4;4'-dimethoxybenzophenone
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.524	0.584	0.609
Reference	28ZPAK 239;72	28ZPAK-;124;72	J. Am. Coll. Toxicol. 2(5):35;1983

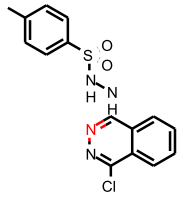
Model Applicability

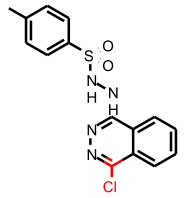
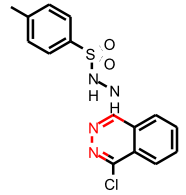
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

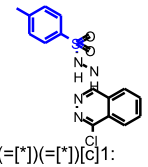
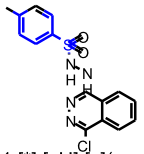
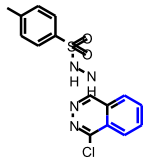
Feature Contribution

Top features for positive contribution

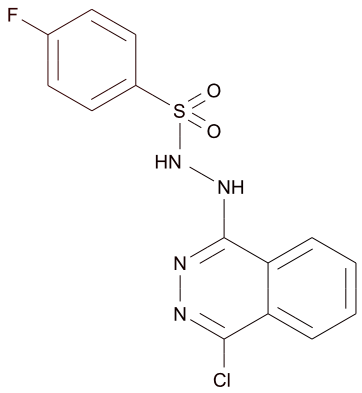
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	17	 [*]:n:[*]	0.189	48 out of 49

FCFP_12	71476542	 [*]:[c](:[*])Cl	0.175	81 out of 84
FCFP_12	4427049	 [*][c](:[*]):n:n:[*]	0.167	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	632767364	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	0	4 out of 5
FCFP_12	-453677277	 [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*])(=[*])[*]	0	264 out of 323
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291

TOPKAT_Ocular_Irritancy_None_vs_Irritant



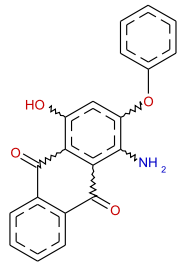
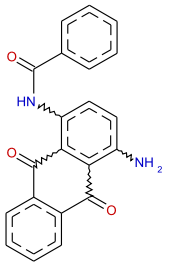
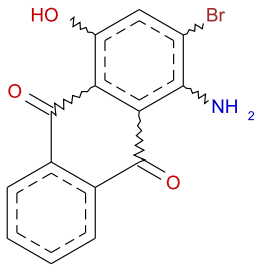
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Irritant
Probability: 1
Enrichment: 1.18
Bayesian Score: 2.11
Mahalanobis Distance: 3.69
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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	Anthraquinone; 1-amino-2-bromo-4-hydroxy-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.530	0.579	0.602
Reference	28ZPAK 239;72	28ZPAK-;124;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 535;86

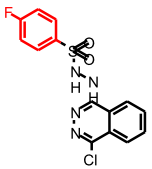
Model Applicability

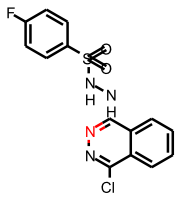
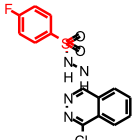
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

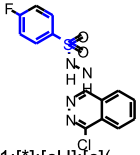
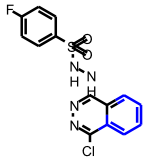
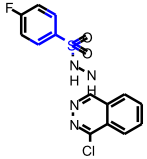
Feature Contribution

Top features for positive contribution

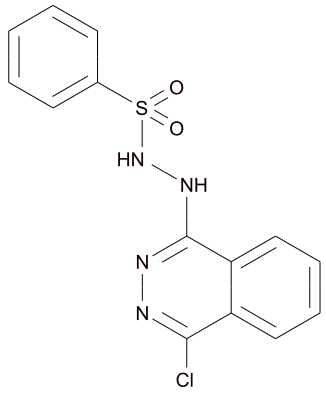
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1508180856	 [*][c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.2	17 out of 17

FCFP_12	17	 [*]:n:[*]	0.189	48 out of 49
FCFP_12	-149636017	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.184	7 out of 7

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-453677277	 [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*])(=[*])[*]	0	264 out of 323
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291
FCFP_12	203677720	 [*]S(=[*])(=[*])[c](:[cH]:[*]):[cH]:[*]	0	319 out of 382

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{14}H_{11}ClN_4O_2S$

Molecular Weight: 334.78073

ALogP: 3.043

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Irritant

Probability: 1

Enrichment: 1.18

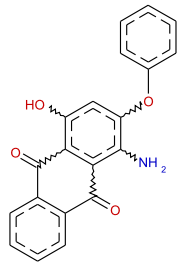
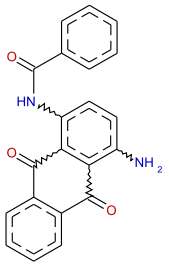
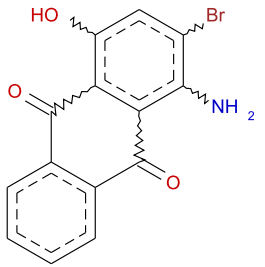
Bayesian Score: 1.17

Mahalanobis Distance: 3.64

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	ANTHRAQUINONE; 1-AMINO-4-HYDROXY-2-PHENOXY-	1-AMINO-4-BENZOYLAMINO-ANTHRAQUINONE	Anthraquinone; 1-amino-2-bromo-4-hydroxy-
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Irritant	Non-Irritant
Distance	0.521	0.567	0.582
Reference	28ZPAK 239;72	28ZPAK-;124;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 535;86

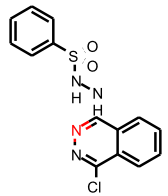
Model Applicability

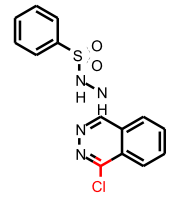
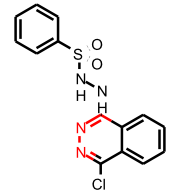
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown FCFP_2 feature: -1151884458: [*]N[c](:n:[*]):[c](:[*]):[*]
3. Unknown FCFP_2 feature: -885461129: [*]NNC(=[*])[*]

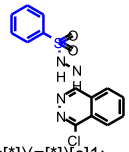
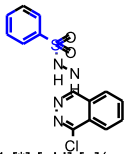
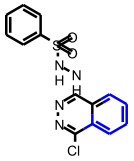
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	17	 [*]:n:[*]	0.189	48 out of 49

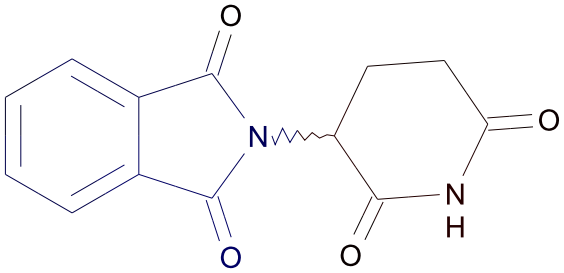
FCFP_12	71476542	 [*]:[c](:[*])Cl	0.175	81 out of 84
FCFP_12	4427049	 [*][c](:[*]):n:n:[*]	0.167	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1698724694	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.0964	107 out of 146
FCFP_12	-453677277	 [*][c]1:[*]:[cH]:[c](:[cH]:[cH]:1)S(=[*])(=[*])[*]	0	264 out of 323
FCFP_12	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	0	237 out of 291

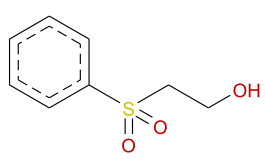
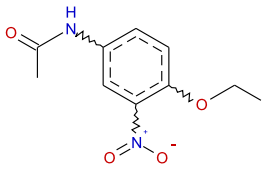
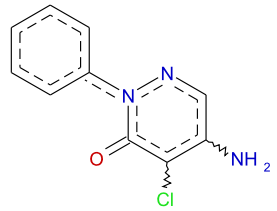
Thalidomide

TOPKAT_Ocular_Irritancy_None_vs_Irritant



$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

Structural Similar Compounds

Name	ETHANOL;2-(PHENYLSULFONYL)-	P-ACETOPHENETIDIDE;3-NITRO-	3(2H)-Pyridazinone; 5-amino-4-chloro-2-phenyl-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Irritant	Irritant	Irritant
Distance	0.570	0.582	0.611
Reference	28ZPAK-;200;72	28ZPAK-;115;72	Prehled Prumyslove Toxikologie; Organicke Latky; Marhold; J. pp 868;86

Model Prediction

Prediction: Non-Irritant
Probability: 0.971
Enrichment: 1.14
Bayesian Score: -1.01
Mahalanobis Distance: 6.67
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.

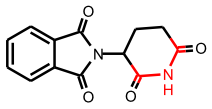
Model Applicability

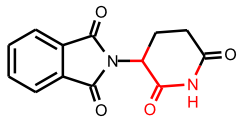
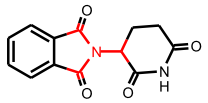
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

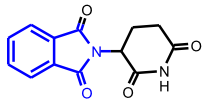
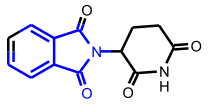
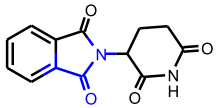
Feature Contribution

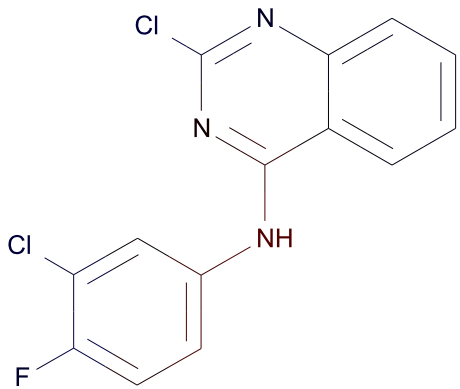
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-885550502	 [*]C(=[*])NC(=[*])[*]	0.18	64 out of 66

FCFP_12	566058135	 [*]CC(=O)N[*]	0.163	23 out of 24
FCFP_12	-1553874037	 [*]C([*])N1C(=[*])[*] :[*]C1=[*]	0.163	42 out of 44

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1393189956	 [*]N1C(=[*])[c]2:[cH] :[cH]:[cH]:[cH]:[c]: 2C1=O	-0.782	5 out of 15
FCFP_12	2139882011	 [*]N1C(=[*])[c]2:[cH] :[*]:[cH]:[cH]:[c]:2 C1=O	-0.782	5 out of 15
FCFP_12	-1549163031	 [*]N1[*][*]:[c](:[*]) C1=O	-0.623	16 out of 38



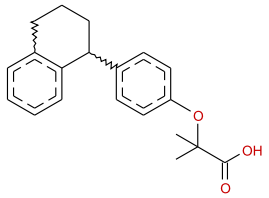
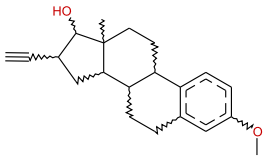
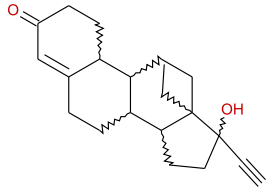
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.251
Enrichment: 0.78
Bayesian Score: -2.7
Mahalanobis Distance: 9.27
Mahalanobis Distance p-value: 0.714

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

Structural Similar Compounds

Name	Nafenopin	Mestranol	Levonorgestrel
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.607	0.610	0.611
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

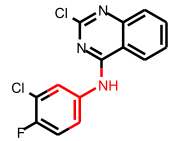
Model Applicability

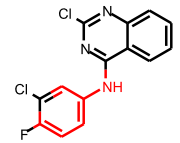
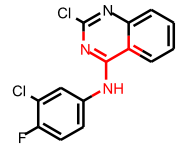
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

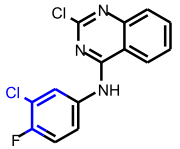
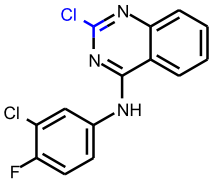
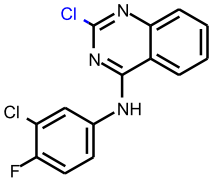
Feature Contribution

Top features for positive contribution

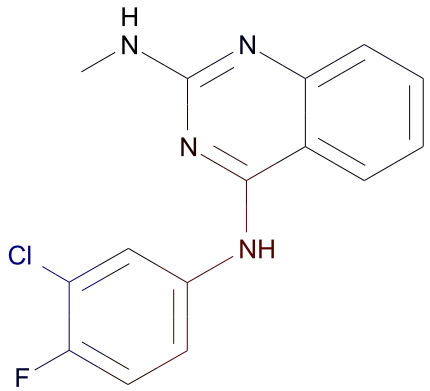
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10

ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1

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	-817402818	 [*]Cl	-0.368	17 out of 80

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.258

Enrichment: 0.802

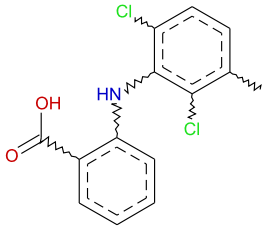
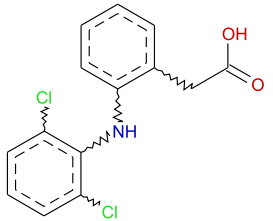
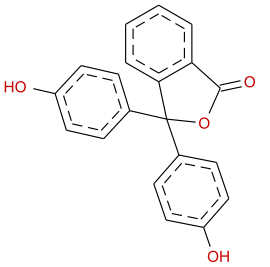
Bayesian Score: -2.25

Mahalanobis Distance: 9.5

Mahalanobis Distance p-value: 0.607

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

Structural Similar Compounds

Name	Meclofenamate	Diclofenac	Phenolphthalein
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.540	0.566	0.594
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

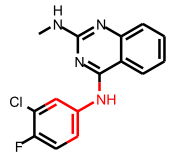
Model Applicability

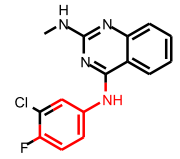
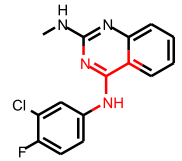
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1240852525: [*]:[c](:[*])NC

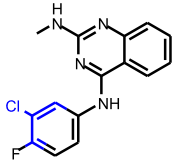
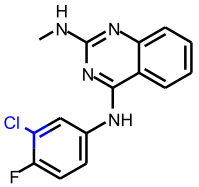
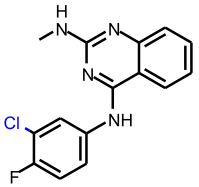
Feature Contribution

Top features for positive contribution

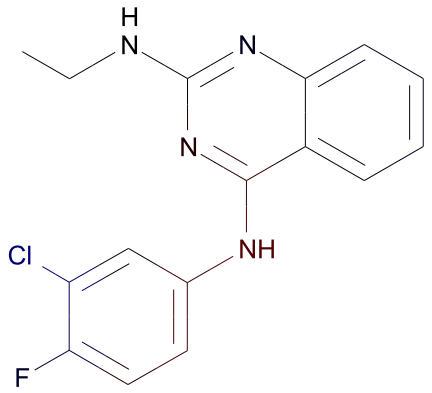
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10

ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1

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	-817402818	 [*]Cl	-0.368	17 out of 80

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{16}H_{14}ClFN_4$

Molecular Weight: 316.76056

ALogP: 5.025

Rotatable Bonds: 4

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.251

Enrichment: 0.779

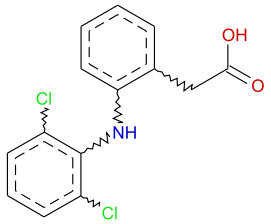
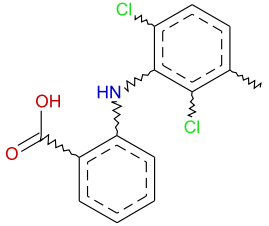
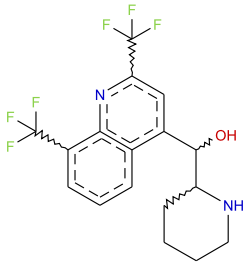
Bayesian Score: -2.72

Mahalanobis Distance: 10.1

Mahalanobis Distance p-value: 0.341

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

Structural Similar Compounds

Name	Diclofenac	Meclofenamate	Mefloquine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.568	0.569	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

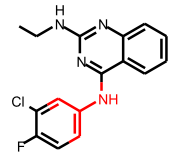
Model Applicability

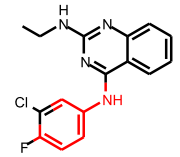
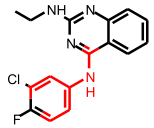
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

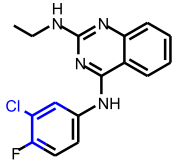
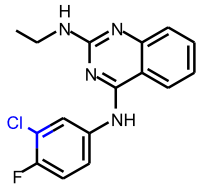
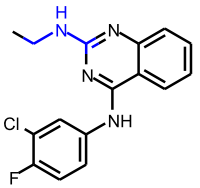
Feature Contribution

Top features for positive contribution

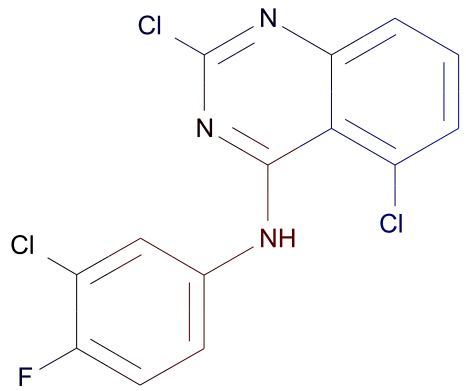
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10

ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	1434334340	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	0.421	1 out of 1

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	491100606	 [*]CN[c](:[*]):[*]	-0.485	0 out of 2

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.228

Enrichment: 0.708

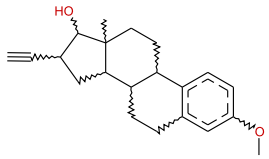
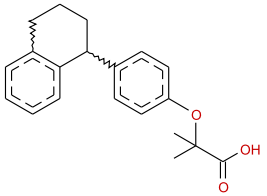
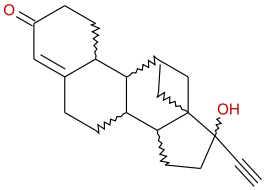
Bayesian Score: -4.41

Mahalanobis Distance: 8.07

Mahalanobis Distance p-value: 0.984

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

Structural Similar Compounds

Name	Mestranol	Nafenopin	Levonorgestrel
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.639	0.649	0.652
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

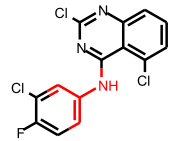
Model Applicability

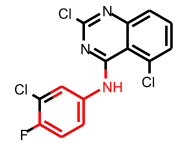
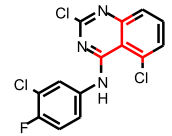
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

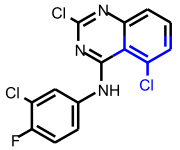
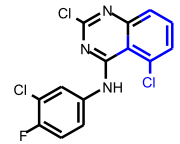
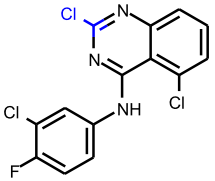
Feature Contribution

Top features for positive contribution

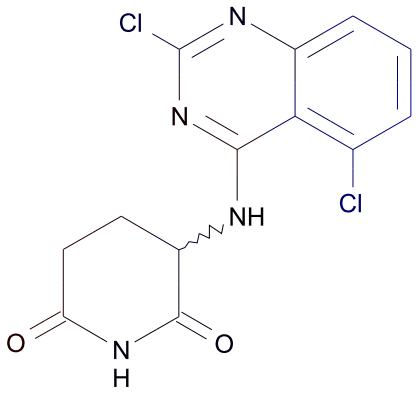
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10

ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.437	2 out of 3

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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227

Enrichment: 0.705

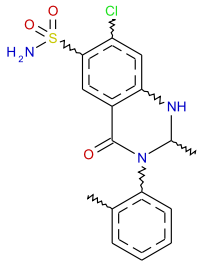
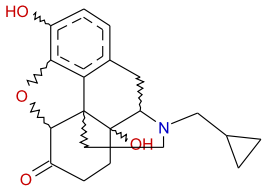
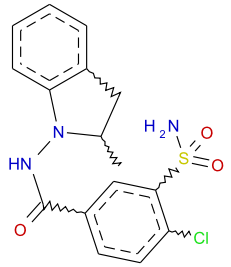
Bayesian Score: -4.5

Mahalanobis Distance: 10.8

Mahalanobis Distance p-value: 0.0952

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

Structural Similar Compounds

Name	Metolazone	Naltrexone	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.574	0.589	0.590
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

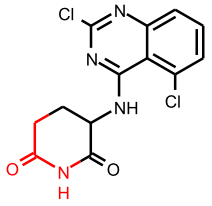
Model Applicability

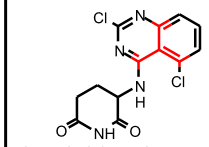
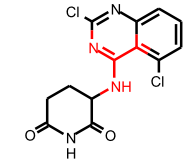
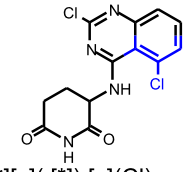
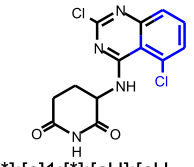
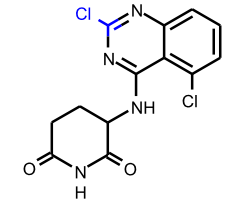
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC30 out of range. Value: 3.7537. Training min, max, SD, explained variance: -2.9582, 2.682, 0.9684, 0.0098.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

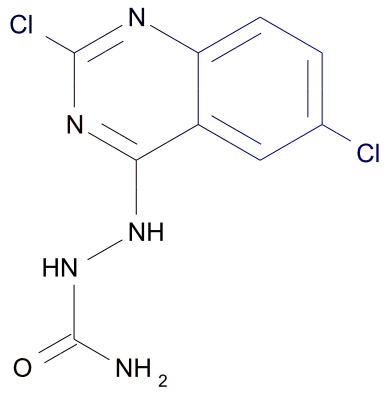
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-742538367	 [*]CC(=O)N[*]	0.445	3 out of 5

ECFP_12	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.437	2 out of 3
ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1
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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09078

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227

Enrichment: 0.704

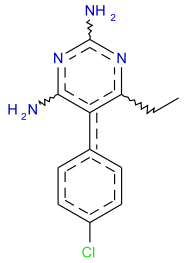
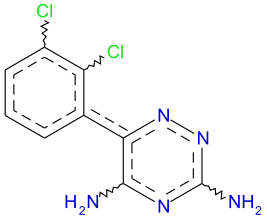
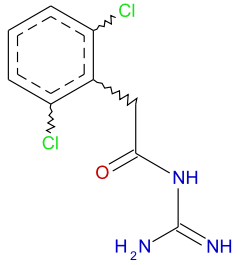
Bayesian Score: -4.51

Mahalanobis Distance: 9.86

Mahalanobis Distance p-value: 0.434

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

Structural Similar Compounds

Name	Pyrimethamine	Lamotrigine	Guanfacine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.627	0.635	0.639
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

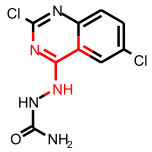
Model Applicability

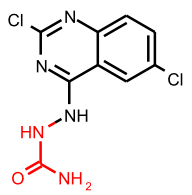
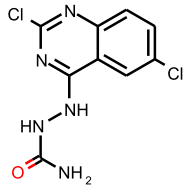
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]

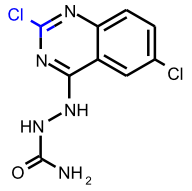
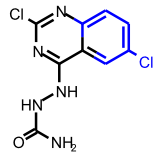
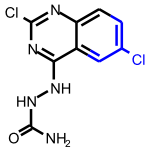
Feature Contribution

Top features for positive contribution

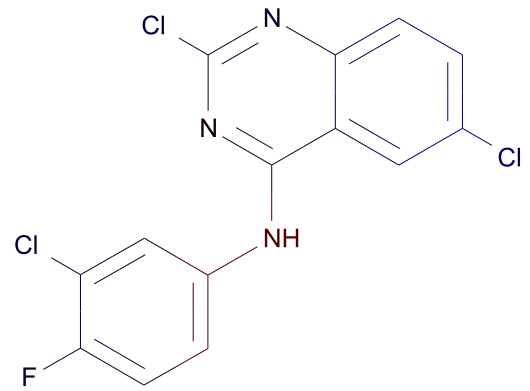
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1

ECFP_12	432952415	 [*]NC(=O)N	0.208	1 out of 2
ECFP_12	-1074141656	 [*]=O	0.103	86 out of 248

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	-176494269	 [*]:[cH]:[c](Cl):[cH]:[*]	-0.714	5 out of 36

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.217

Enrichment: 0.673

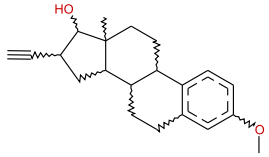
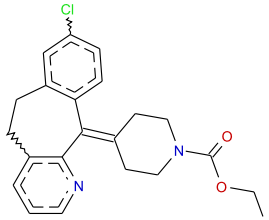
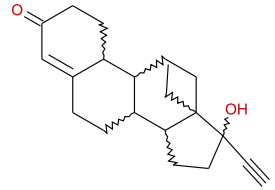
Bayesian Score: -5.4

Mahalanobis Distance: 9.72

Mahalanobis Distance p-value: 0.503

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

Structural Similar Compounds

Name	Mestranol	Loratidine	Levonorgestrel
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.639	0.647	0.652
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

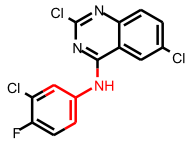
Model Applicability

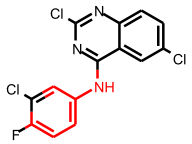
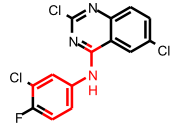
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

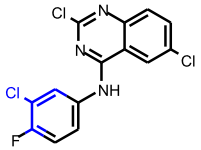
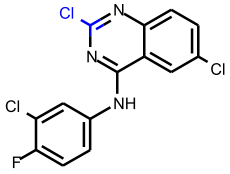
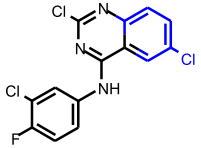
Feature Contribution

Top features for positive contribution

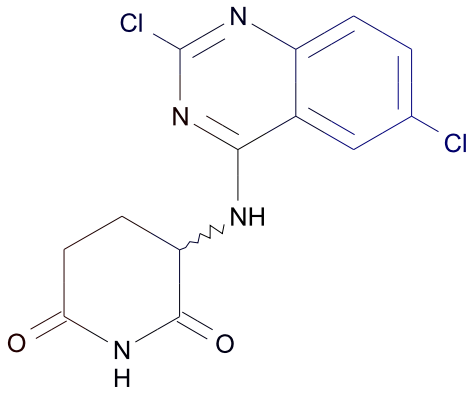
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-177077903	 [*]N[c](:[cH]:[*]):[cH]:[*]	0.529	6 out of 10

ECFP_12	888054369	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.454	5 out of 9
ECFP_12	1434334340	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	0.421	1 out of 1

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	1854732111	 [*]:[c]1:[*]:[cH]:[c](Cl):[cH]:[cH]:1	-0.816	4 out of 33

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.227

Enrichment: 0.707

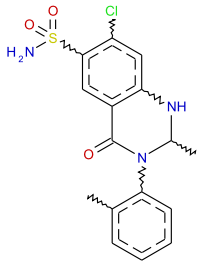
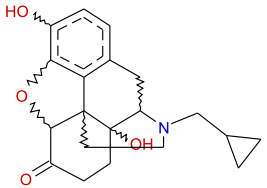
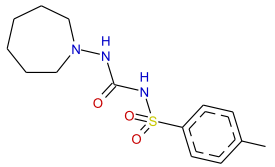
Bayesian Score: -4.46

Mahalanobis Distance: 11.5

Mahalanobis Distance p-value: 0.0199

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

Structural Similar Compounds

Name	Metolazone	Naltrexone	Tolazamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.578	0.589	0.594
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

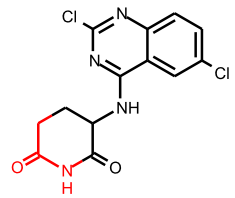
Model Applicability

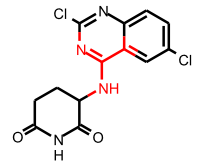
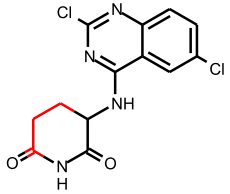
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

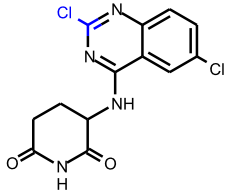
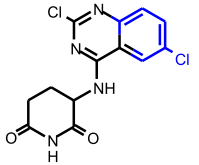
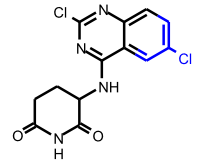
Feature Contribution

Top features for positive contribution

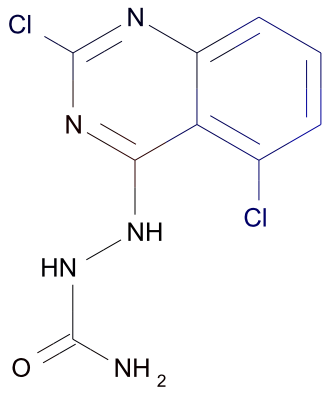
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-742538367	 [*]CC(=O)N[*]	0.445	3 out of 5

ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1
ECFP_12	51876938	 [*]CCC(=[*])[*]	0.232	18 out of 45

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	-176494269	 [*]:[cH]:[c](Cl):[cH]:[*]	-0.714	5 out of 36

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09077

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.226

Enrichment: 0.703

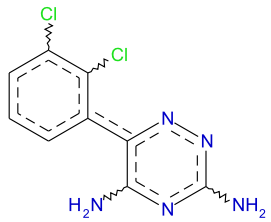
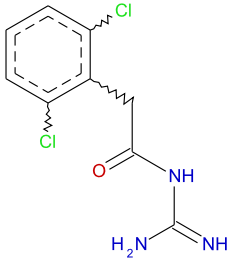
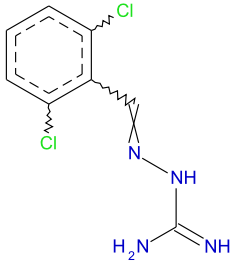
Bayesian Score: -4.56

Mahalanobis Distance: 8.39

Mahalanobis Distance p-value: 0.957

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

Structural Similar Compounds

Name	Lamotrigine	Guanfacine	Guanabenz
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.608	0.609	0.628
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

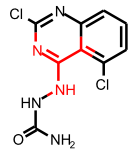
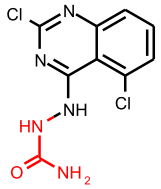
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]

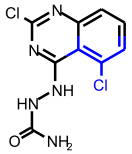
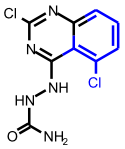
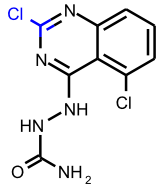
Feature Contribution

Top features for positive contribution

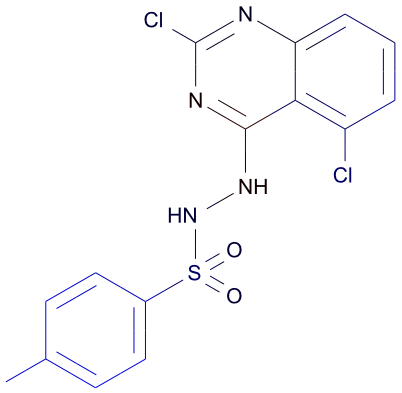
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.437	2 out of 3

ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1
ECFP_12	432952415	 [*]NC(=O)N	0.208	1 out of 2

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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{15}H_{12}Cl_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.189

Enrichment: 0.588

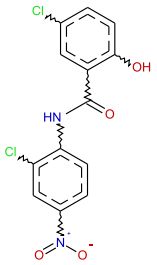
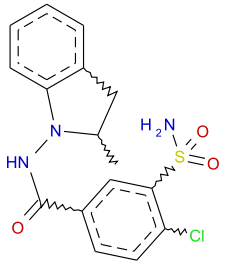
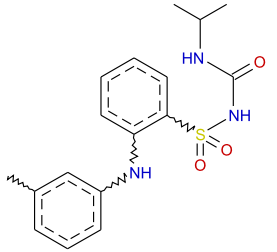
Bayesian Score: -8.62

Mahalanobis Distance: 8.88

Mahalanobis Distance p-value: 0.858

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

Structural Similar Compounds

Name	Niclosamide	Indapamide	Torsemide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.647	0.663	0.670
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

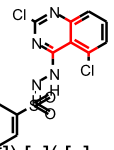
Model Applicability

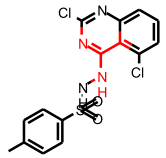
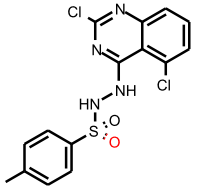
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
4. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

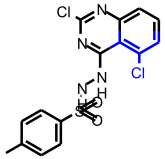
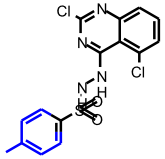
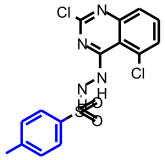
Feature Contribution

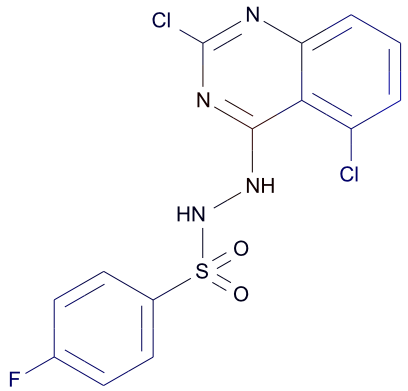
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.437	2 out of 3

ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1
ECFP_12	-1074141656	 [*]=O	0.103	86 out of 248

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	-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



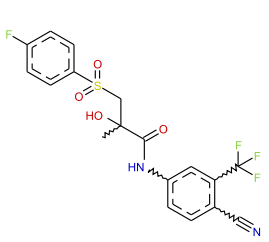
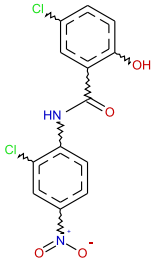
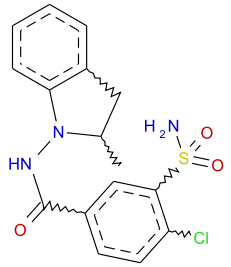
$C_{14}H_9Cl_2FN_4O_2S$
Molecular Weight: 387.21626
ALogP: 4.162
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.202
Enrichment: 0.627
Bayesian Score: -6.95
Mahalanobis Distance: 9.01
Mahalanobis Distance p-value: 0.816

Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Probability: The esimated probability that the sample is in the positive category. This assumes that the Bayesian score follows a normal distribution and is different from the prediction using a cutoff.
Enrichment: An estimate of enrichment, that is, the increased likelihood (versus random) of this sample being in the category.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the 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	Niclosamide	Indapamide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.632	0.637	0.645
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

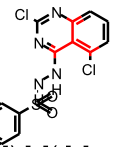
Model Applicability

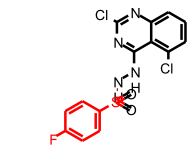
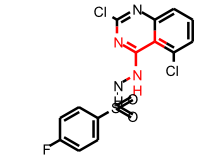
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
4. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

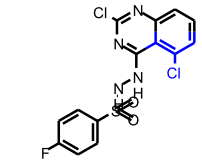
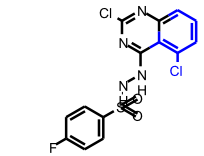
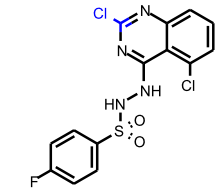
Feature Contribution

Top features for positive contribution

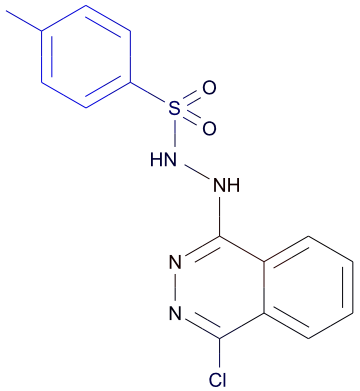
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1661653144	 [*][c](:[*]):[c](:[c]([*]):[*]):[c](:[*]):[*]	0.437	2 out of 3

ECFP_12	74189795	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.421	1 out of 1
ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1

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

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



$C_{15}H_{13}ClN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.202

Enrichment: 0.628

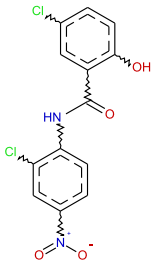
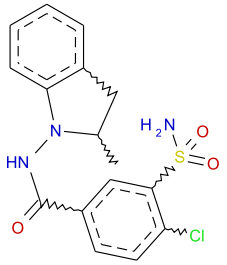
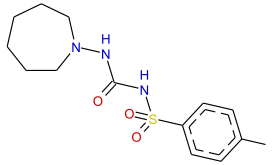
Bayesian Score: -6.93

Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.3

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

Structural Similar Compounds

Name	Niclosamide	Indapamide	Tolazamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.592	0.594	0.605
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

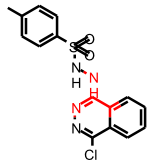
Model Applicability

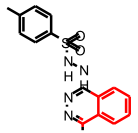
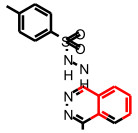
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

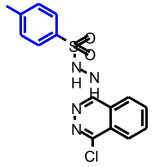
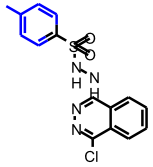
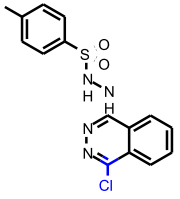
Feature Contribution

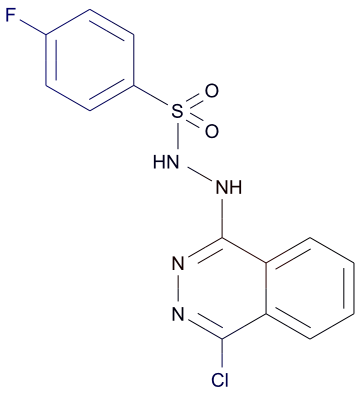
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1

ECFP_12	1306977740	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.271	4 out of 9
ECFP_12	1639858918	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.119	4 out of 11

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-1926229349	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-1.06	0 out of 6
ECFP_12	-533780882	 C[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-1.06	0 out of 6
ECFP_12	99947387	 [*]:[c](:[*])Cl	-0.817	8 out of 62



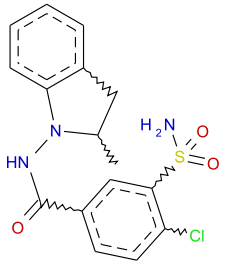
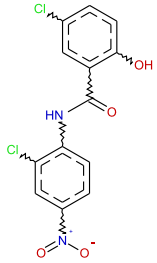
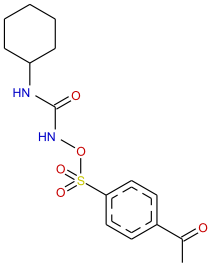
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.224
Enrichment: 0.695
Bayesian Score: -4.77
Mahalanobis Distance: 9.96
Mahalanobis Distance p-value: 0.385

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

Structural Similar Compounds

Name	Indapamide	Niclosamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.583	0.595	0.598
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

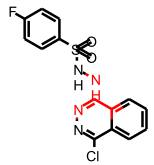
Model Applicability

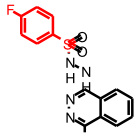
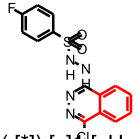
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

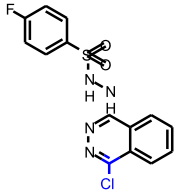
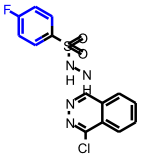
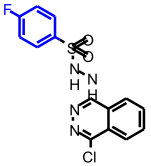
Feature Contribution

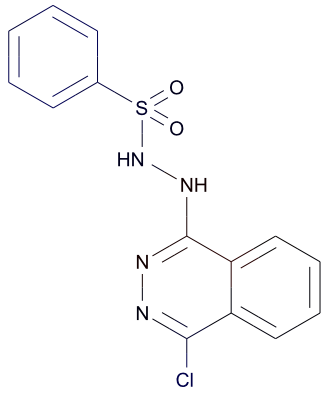
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1

ECFP_12	74189795	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.421	1 out of 1
ECFP_12	1306977740	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.271	4 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	-296909061	 F[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.56	1 out of 8
ECFP_12	-1659169698	 [*][c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	-0.56	1 out of 8



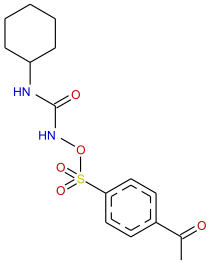
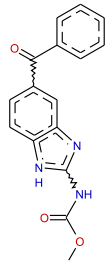
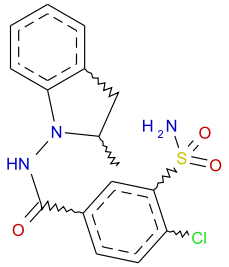
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.223
Enrichment: 0.693
Bayesian Score: -4.83
Mahalanobis Distance: 10.6
Mahalanobis Distance p-value: 0.137

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

Structural Similar Compounds

Name	Acetohexamide	Mebendazole	Indapamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.579	0.581	0.584
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

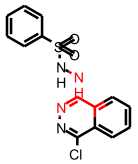
Model Applicability

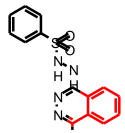
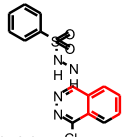
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1238415266: [*]NN[c](:[*]):[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

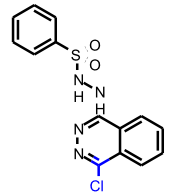
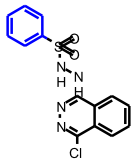
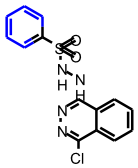
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	1049768340	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.421	1 out of 1

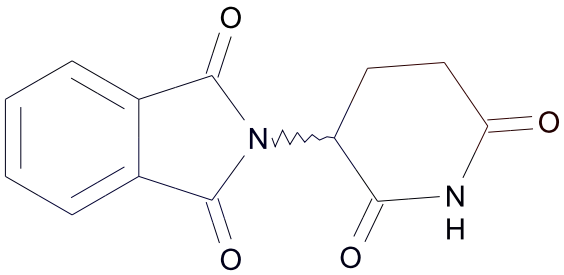
ECFP_12	1306977740	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.271	4 out of 9
ECFP_12	1639858918	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.119	4 out of 11

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	-281505363	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64
ECFP_12	1571214559	 [*]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.56	11 out of 64

Thalidomide

TOPKAT_Rat_Female_FDA_None_vs_Carcinogen



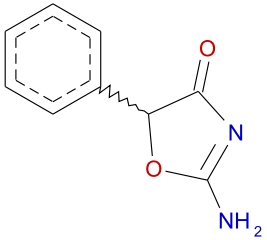
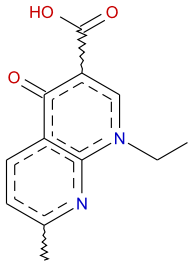
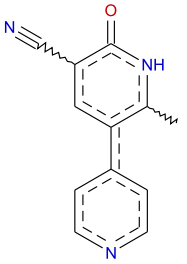
$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.283
Enrichment: 0.879
Bayesian Score: -0.795
Mahalanobis Distance: 8.89
Mahalanobis Distance p-value: 0.853

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

Structural Similar Compounds

Name	Pemoline	Nalidixic acid	Milrinone
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.629	0.638	0.647
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

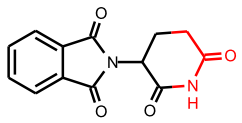
Model Applicability

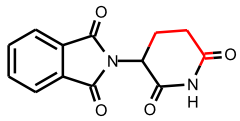
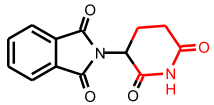
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

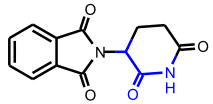
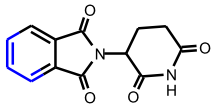
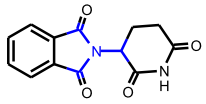
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-742538367	 [*]CC(=O)N[*]	0.445	3 out of 5

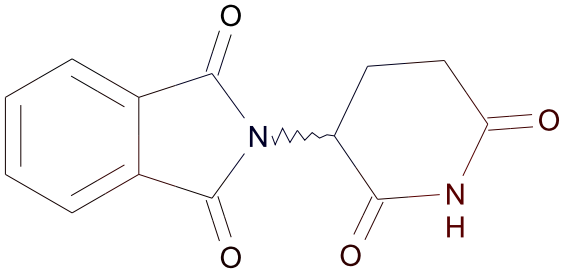
ECFP_12	51876938	 [*]CCC(=[*])[*]	0.232	18 out of 45
ECFP_12	-2105515128	 [*]=C1[*]CCC(=O)N1	0.208	1 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
ECFP_12	-867777309	 [*]NC(=O)C([*])[*]	-0.661	0 out of 3
ECFP_12	1997021792	 [*]:[cH]:[cH]:[cH]:[*]	-0.296	36 out of 156
ECFP_12	-957084426	 [*]C([*])N1C(=[*])[*]:[*]C1=[*]	-0.272	0 out of 1

Thalidomide

TOPKAT_Rat_Female_FDA_Single_vs_Multiple



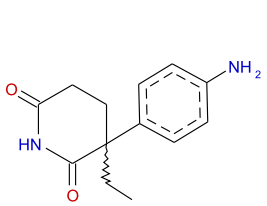
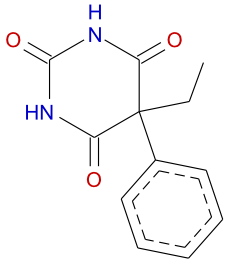
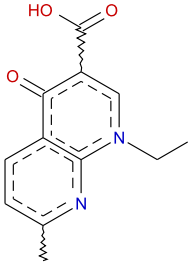
$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

Model Prediction

Prediction: Multiple-Carcinogen
Probability: 0.524
Enrichment: 1.4
Bayesian Score: 0.842
Mahalanobis Distance: 12.4
Mahalanobis Distance p-value: 0.000809

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

Structural Similar Compounds

Name	Aminoglutethimide	Phenobarbital	Nalidixic acid
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Single-Carcinogen
Distance	0.579	0.599	0.599
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

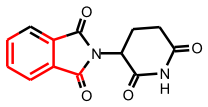
Model Applicability

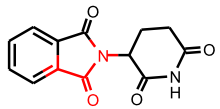
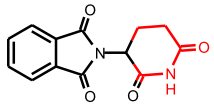
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

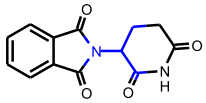
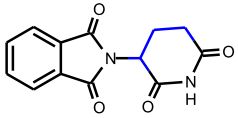
Feature Contribution

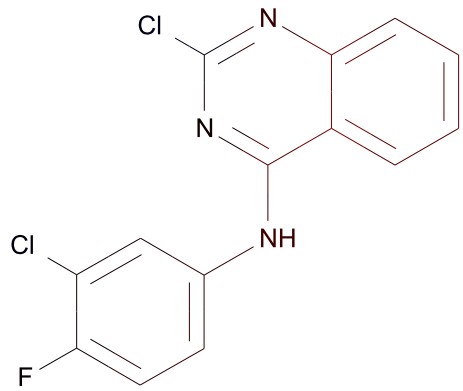
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	387787917	 [*]=C1[*][*][c]2:[*]:[cH]:[cH]:[cH]:[c]1:2	0.449	6 out of 11

SCFP_4	1257084377	 [*]N1[*][*]:[c](:[*])C1=O	0.44	3 out of 5
SCFP_4	345243876	 [*]=C1[*]CCC(=O)N1	0.419	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_4	-1946889102	 [*]CC(N([*])[*])C(=[*])[*]	-0.816	0 out of 4
SCFP_4	399659969	 [*]C([*])N1C(=[*])[*]:[*]C1=[*]	-0.666	0 out of 3
SCFP_4	-1272798659	 [*]CCC([*])[*]	-0.421	10 out of 50



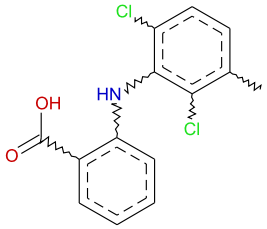
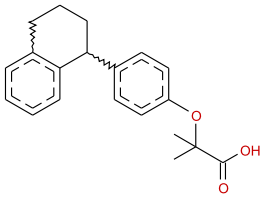
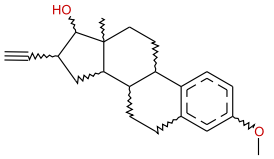
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Carcinogen
Probability: 0.454
Enrichment: 1.36
Bayesian Score: 3.08
Mahalanobis Distance: 13.9
Mahalanobis Distance p-value: 5.62e-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	Meclofenamate	Nafenopin	Mestranol
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Carcinogen
Distance	0.605	0.609	0.613
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

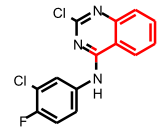
Model Applicability

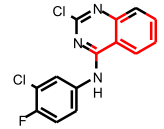
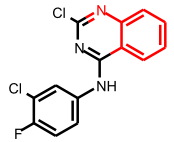
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC15 out of range. Value: 5.9999. Training min, max, SD, explained variance: -4.0246, 5.5336, 1.388, 0.0180.

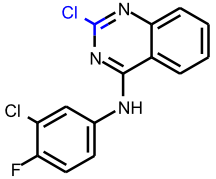
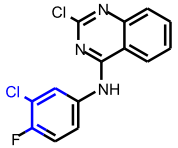
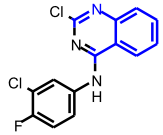
Feature Contribution

Top features for positive contribution

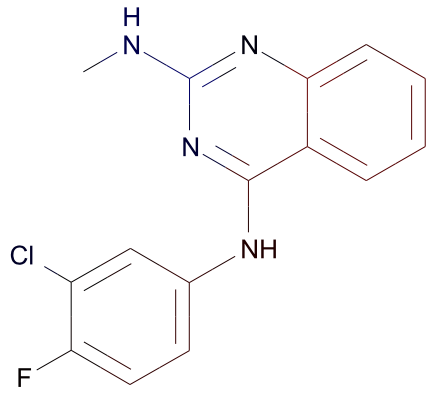
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.643	7 out of 10

SCFP_6	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.526	11 out of 19
SCFP_6	1655199790	 [*]:n:[c]1:[cH]:[cH]: [cH]:[cH]:[c]:1:[*]	0.52	5 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	-601571304	 [*][c](:[*]):[c](Cl): [cH]:[*]	-0.3	14 out of 60
SCFP_6	403834996	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[cH]:[cH]: [cH]:[c]:2:n:1	-0.264	1 out of 5

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.417

Enrichment: 1.25

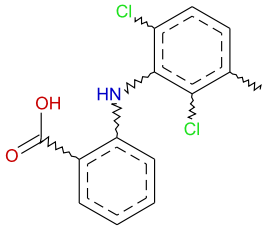
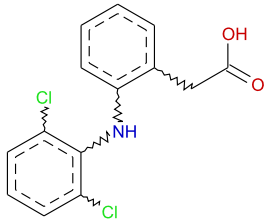
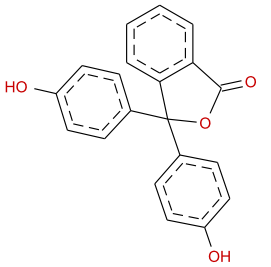
Bayesian Score: 1.98

Mahalanobis Distance: 13.9

Mahalanobis Distance p-value: 5.82e-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	Meclofenamate	Diclofenac	Phenolphthalein
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.525	0.544	0.592
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

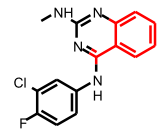
Model Applicability

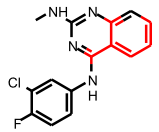
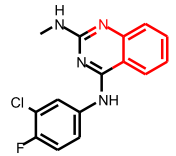
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC15 out of range. Value: 5.5915. Training min, max, SD, explained variance: -4.0246, 5.5336, 1.388, 0.0180.

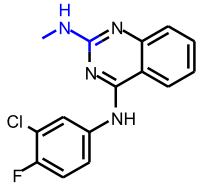
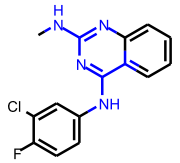
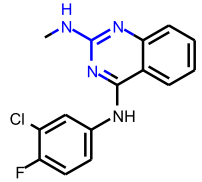
Feature Contribution

Top features for positive contribution

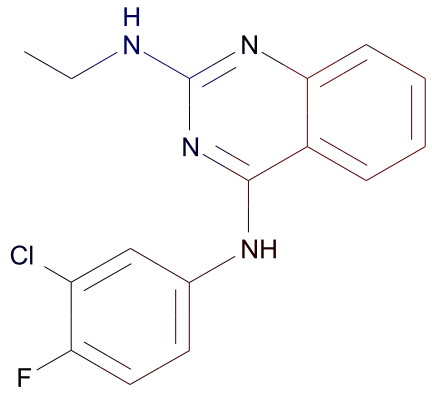
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.643	7 out of 10

SCFP_6	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.526	11 out of 19
SCFP_6	1655199790	 [*]:n:[c]1:[cH]:[cH]: [cH]:[cH]:[c]:1:[*]	0.52	5 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	18117904	 [*]:[c](:[*])NC	-0.578	1 out of 8
SCFP_6	-1545804258	 [*]N[c]1:n:[*]:[c](:[*]):[c](N[*]):n:1	-0.578	1 out of 8
SCFP_6	649648475	 [*]N[c](:n:[*]):n:[*]	-0.48	2 out of 12

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{16}H_{14}ClFN_4$

Molecular Weight: 316.76056

ALogP: 5.025

Rotatable Bonds: 4

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Carcinogen

Probability: 0.399

Enrichment: 1.19

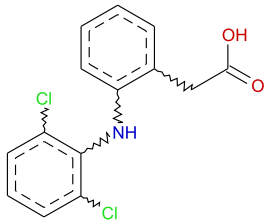
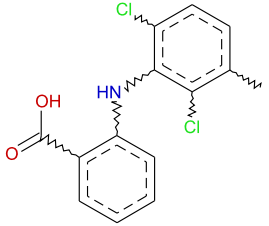
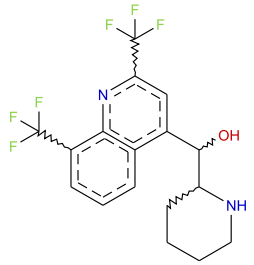
Bayesian Score: 1.42

Mahalanobis Distance: 16

Mahalanobis Distance p-value: 4.53e-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	Diclofenac	Meclofenamate	Mefloquine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.550	0.554	0.610
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

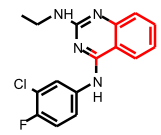
Model Applicability

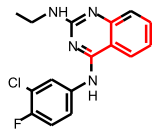
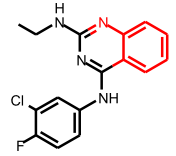
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

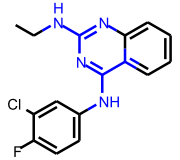
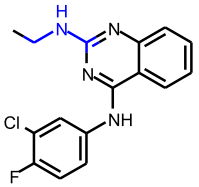
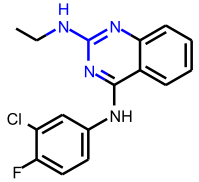
Feature Contribution

Top features for positive contribution

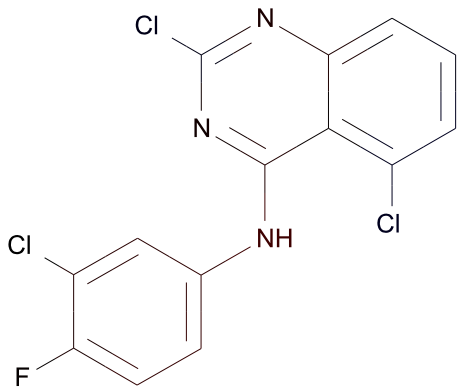
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.643	7 out of 10

SCFP_6	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]: 1:[*]	0.526	11 out of 19
SCFP_6	1655199790	 [*]:n:[c]1:[cH]:[cH]: [cH]:[cH]:[c]:1:[*]	0.52	5 out of 8

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1545804258	 [*]N[c]1:n:[*]:[c](:[*]):[c](N[*]):n:1	-0.578	1 out of 8
SCFP_6	18117904	 [*]:[c](:[*])NC	-0.578	1 out of 8
SCFP_6	649648475	 [*]N[c](:n:[*]):n:[*]	-0.48	2 out of 12

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.364

Enrichment: 1.09

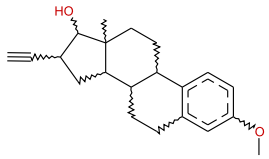
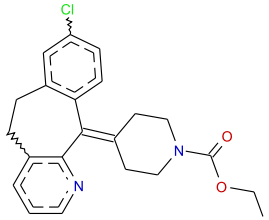
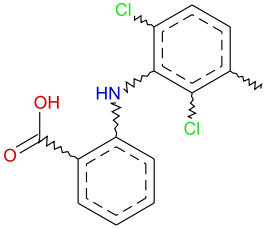
Bayesian Score: 0.273

Mahalanobis Distance: 13.5

Mahalanobis Distance p-value: 0.000258

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

Structural Similar Compounds

Name	Mestranol	Loratidine	Meclofenamate
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.642	0.642	0.642
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

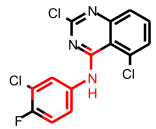
Model Applicability

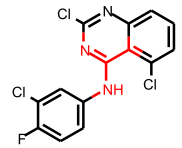
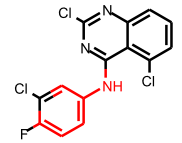
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

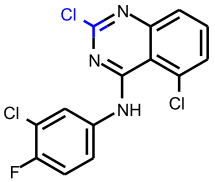
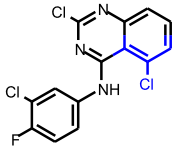
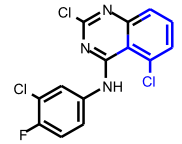
Feature Contribution

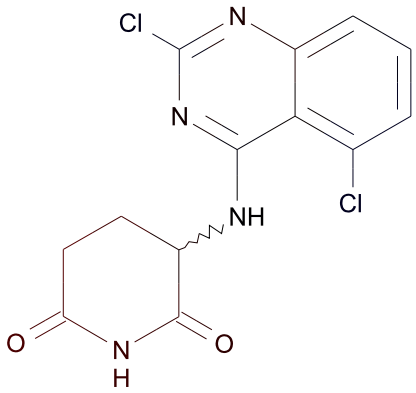
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	0.415	1 out of 1

SCFP_6	1334878018	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.333	8 out of 17
SCFP_6	-1380909229	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.287	17 out of 39

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.289	12 out of 51



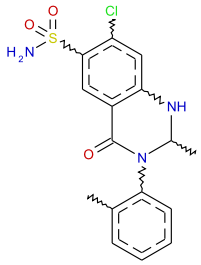
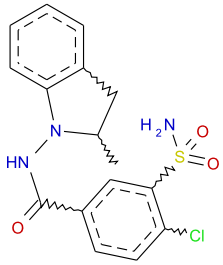
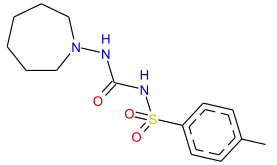
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.366
Enrichment: 1.1
Bayesian Score: 0.331
Mahalanobis Distance: 14.1
Mahalanobis Distance p-value: 3.26e-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	Metolazone	Indapamide	Tolazamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.548	0.566	0.573
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

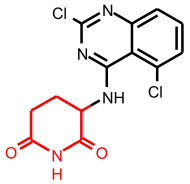
Model Applicability

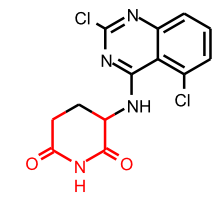
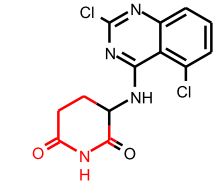
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

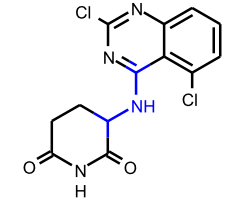
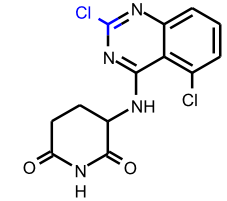
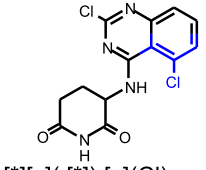
Feature Contribution

Top features for positive contribution

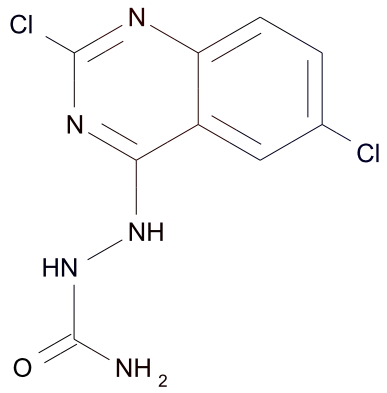
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1533214013	 [*]C1CCC(=O)NC1=O	0.415	1 out of 1

SCFP_6	395945879	 [*]C1[*]CC(=O)NC1=O	0.415	1 out of 1
SCFP_6	345243876	 [*]=C1[*]CCC(=O)N1	0.415	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	18117904	 [*]:[c](:[*])NC	-0.578	1 out of 8
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.3	14 out of 60

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



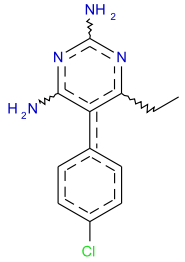
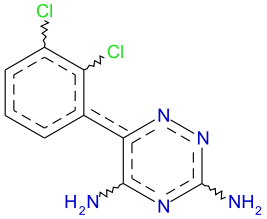
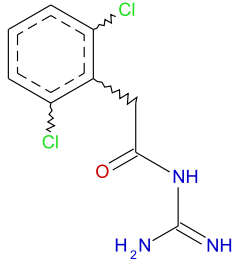
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09078
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.348
Enrichment: 1.04
Bayesian Score: -0.291
Mahalanobis Distance: 14.3
Mahalanobis Distance p-value: 1.37e-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	Pyrimethamine	Lamotrigine	Guanfacine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.603	0.608	0.622
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

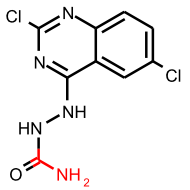
Model Applicability

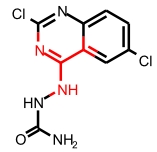
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

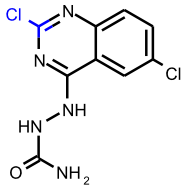
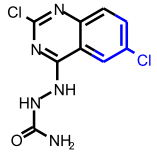
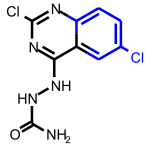
Feature Contribution

Top features for positive contribution

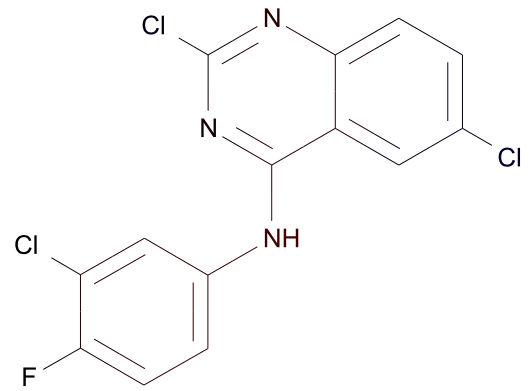
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1357949052	 [*]C(=[*])N	0.494	8 out of 14

SCFP_6	1334878018	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.333	8 out of 17
SCFP_6	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.276	13 out of 30

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.289	12 out of 51

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Carcinogen

Probability: 0.365

Enrichment: 1.09

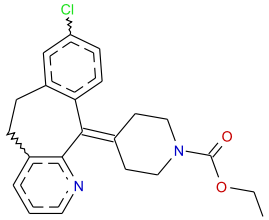
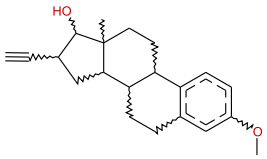
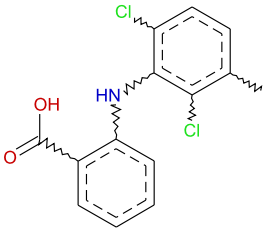
Bayesian Score: 0.284

Mahalanobis Distance: 13.6

Mahalanobis Distance p-value: 0.000236

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

Structural Similar Compounds

Name	Loratidine	Mestranol	Meclofenamate
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.637	0.642	0.649
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

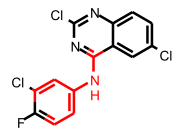
Model Applicability

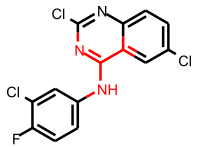
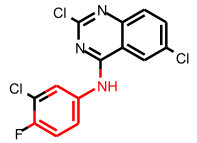
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

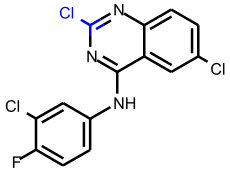
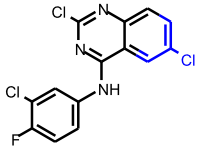
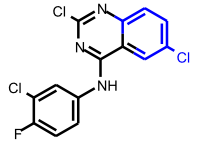
Feature Contribution

Top features for positive contribution

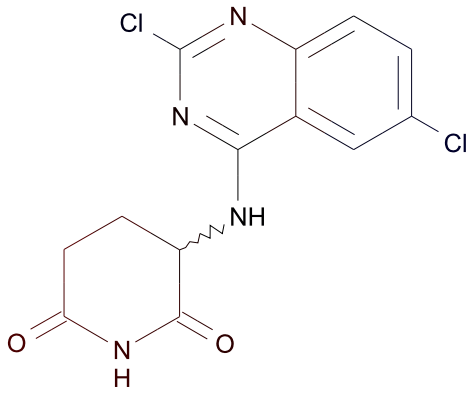
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-300914917	 [*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.415	1 out of 1

SCFP_6	1334878018	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.333	8 out of 17
SCFP_6	-1380909229	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	0.287	17 out of 39

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.289	12 out of 51

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



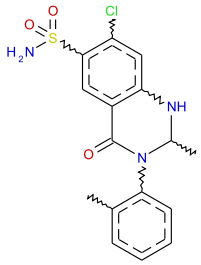
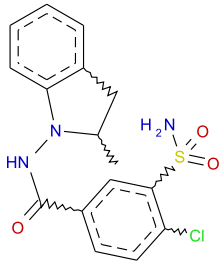
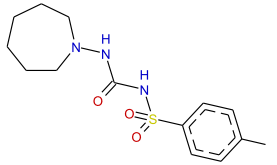
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.366
Enrichment: 1.1
Bayesian Score: 0.343
Mahalanobis Distance: 14.4
Mahalanobis Distance p-value: 9.01e-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	Metolazone	Indapamide	Tolazamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.554	0.572	0.573
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

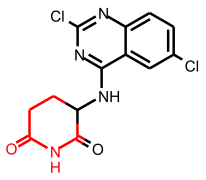
Model Applicability

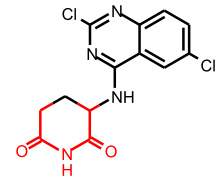
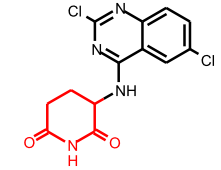
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

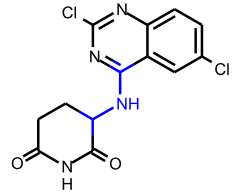
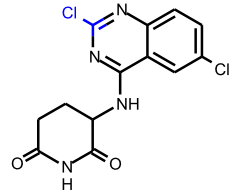
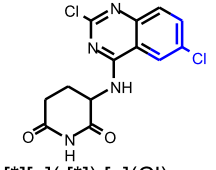
Feature Contribution

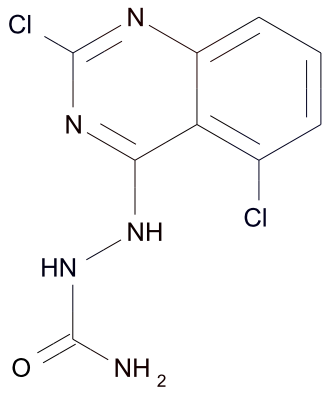
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	345243876	 [*]=C1[*]CCC(=O)N1	0.415	1 out of 1

SCFP_6	395945879	 [*]C1[*]CC(=O)NC1=O	0.415	1 out of 1
SCFP_6	-1533214013	 [*]C1CCC(=O)NC1=O	0.415	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	18117904	 [*]:[c](:[*])NC	-0.578	1 out of 8
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.3	14 out of 60



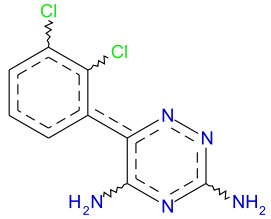
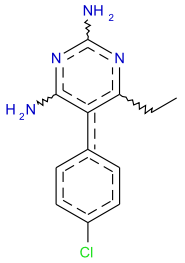
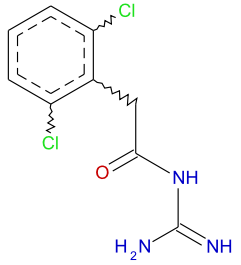
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.348
Enrichment: 1.04
Bayesian Score: -0.303
Mahalanobis Distance: 13.8
Mahalanobis Distance p-value: 9.34e-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	Lamotrigine	Pyrimethamine	Guanfacine
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.600	0.611	0.613
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

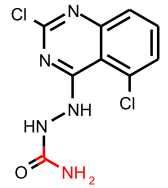
Model Applicability

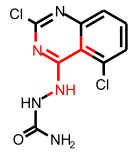
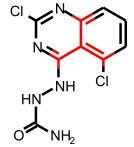
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

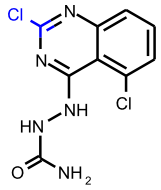
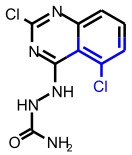
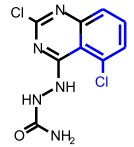
Feature Contribution

Top features for positive contribution

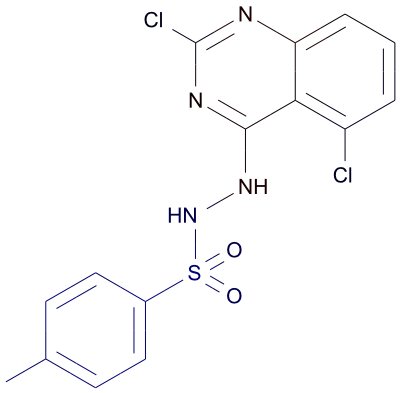
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1357949052	 [*]C(=[*])N	0.494	8 out of 14

SCFP_6	1334878018	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.333	8 out of 17
SCFP_6	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.276	13 out of 30

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.289	12 out of 51

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



$C_{15}H_{12}Cl_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.284

Enrichment: 0.849

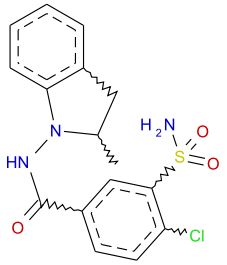
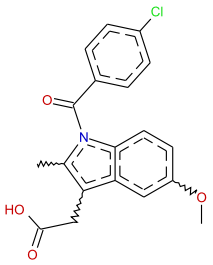
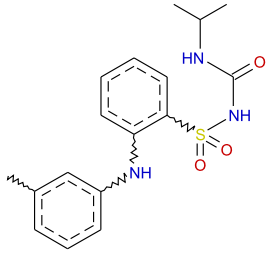
Bayesian Score: -2.82

Mahalanobis Distance: 14.5

Mahalanobis Distance p-value: 5.33e-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	Indapamide	Indomethacin	Torsemide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Carcinogen
Distance	0.617	0.628	0.630
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

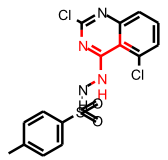
Model Applicability

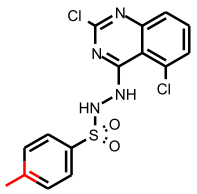
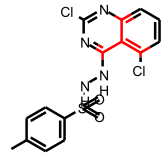
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

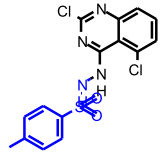
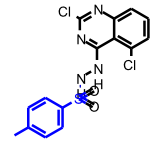
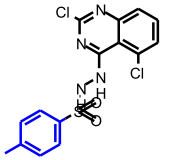
Feature Contribution

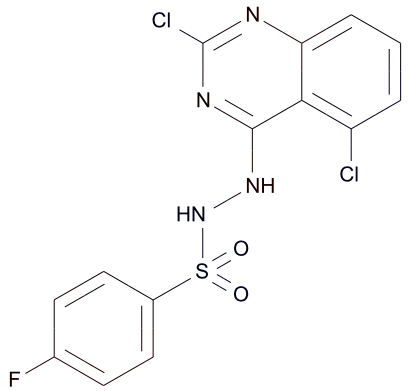
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1334878018	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.333	8 out of 17

SCFP_6	136686699	 [*]:[c](:[*])C	0.287	17 out of 39
SCFP_6	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.276	13 out of 30

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1892882306	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.957	0 out of 5
SCFP_6	2054891299	 [*]S(=[*])(=[*])[c]1: [cH]:[cH]:[c](C):[cH]:[cH]:1	-0.496	0 out of 2
SCFP_6	795925860	 [*][c]1:[cH]:[cH]:[c] (C):[cH]:[cH]:1	-0.38	1 out of 6



$C_{14}H_9Cl_2FN_4O_2S$

Molecular Weight: 387.21626

ALogP: 4.162

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Carcinogen

Probability: 0.347

Enrichment: 1.04

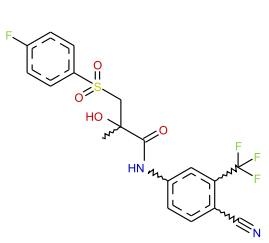
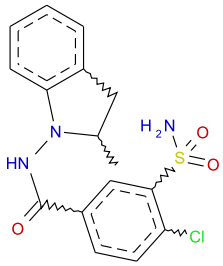
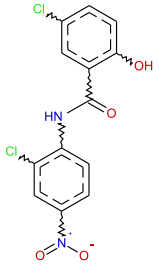
Bayesian Score: -0.338

Mahalanobis Distance: 15.5

Mahalanobis Distance p-value: 7.06e-008

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

Structural Similar Compounds

Name	Bicalutamide	Indapamide	Niclosamide
Structure			
Actual Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.612	0.613	0.621
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

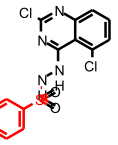
Model Applicability

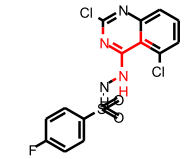
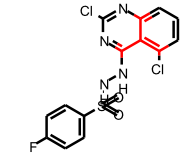
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC1 out of range. Value: -5.807. Training min, max, SD, explained variance: -5.694, 8.348, 2.949, 0.0811.

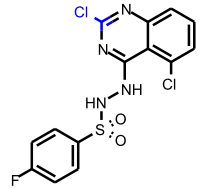
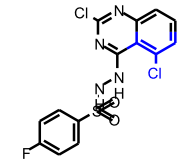
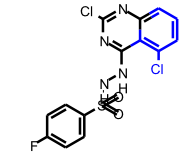
Feature Contribution

Top features for positive contribution

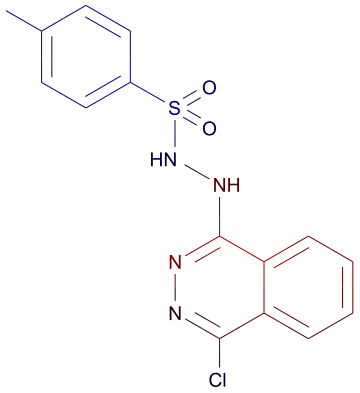
Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	2144224020	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.415	1 out of 1

SCFP_6	1334878018	 [*]N[c](:n:[*]):[c](:[*]):[*]	0.333	8 out of 17
SCFP_6	112346096	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.276	13 out of 30

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.3	14 out of 60
SCFP_6	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.289	12 out of 51

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



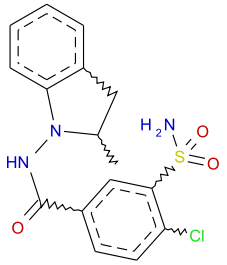
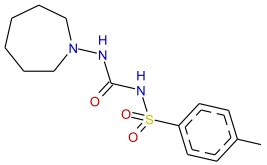
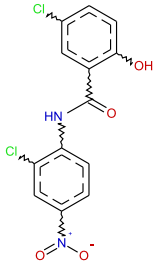
$C_{15}H_{13}ClN_4O_2S$
Molecular Weight: 348.80731
ALogP: 3.529
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.349
Enrichment: 1.04
Bayesian Score: -0.258
Mahalanobis Distance: 25.5
Mahalanobis Distance p-value: 8.64e-034

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

Structural Similar Compounds

Name	Indapamide	Tolazamide	Niclosamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.559	0.569	0.581
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

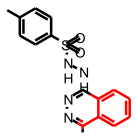
Model Applicability

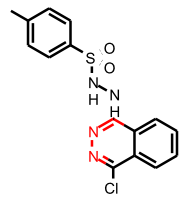
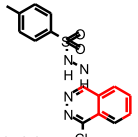
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

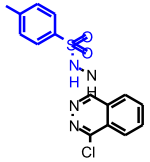
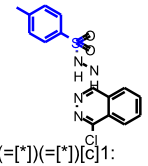
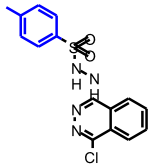
Feature Contribution

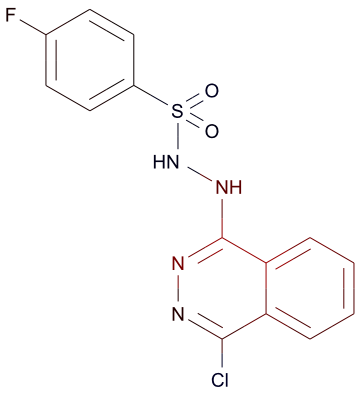
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.643	7 out of 10

SCFP_6	149212520	 [*][c](:[*]):n:n:[*]	0.543	9 out of 15
SCFP_6	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.526	11 out of 19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1892882306	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.957	0 out of 5
SCFP_6	2054891299	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.496	0 out of 2
SCFP_6	795925860	 [*][c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.38	1 out of 6



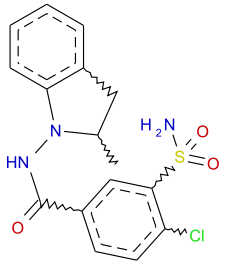
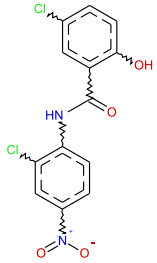
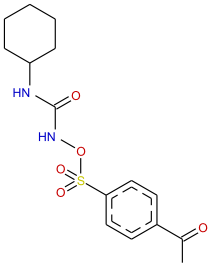
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.437
Enrichment: 1.31
Bayesian Score: 2.56
Mahalanobis Distance: 25.8
Mahalanobis Distance p-value: 1.04e-034

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

Structural Similar Compounds

Name	Indapamide	Niclosamide	Acetohexamide
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.561	0.585	0.587
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

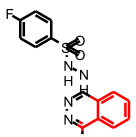
Model Applicability

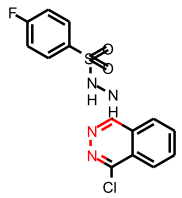
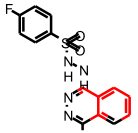
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

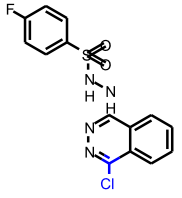
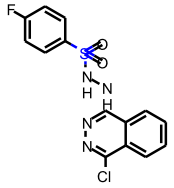
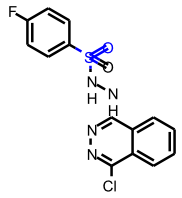
Feature Contribution

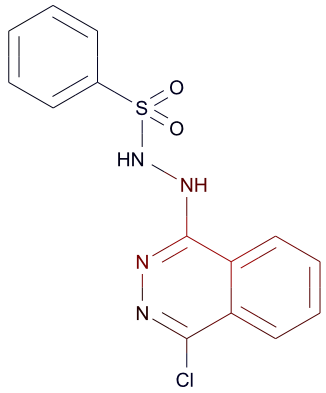
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1651620003	 [*][c](:[*]):[c]:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.643	7 out of 10

SCFP_6	149212520	 [*][c](:[*]):n:n:[*]	0.543	9 out of 15
SCFP_6	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.526	11 out of 19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	21	 [*]S(=[*])(=[*])[*]	-0.283	7 out of 30
SCFP_6	1311429347	 [*]S(=[*])(=O)[*]	-0.283	7 out of 30



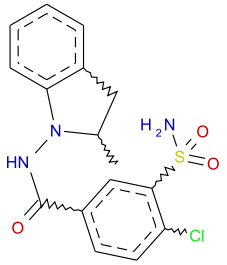
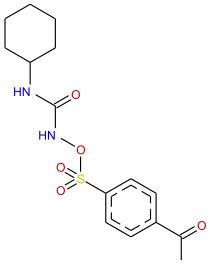
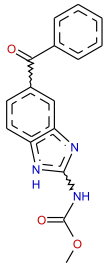
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Carcinogen
Probability: 0.386
Enrichment: 1.16
Bayesian Score: 1
Mahalanobis Distance: 25.7
Mahalanobis Distance p-value: 1.97e-034

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

Structural Similar Compounds

Name	Indapamide	Acetohexamide	Mebendazole
Structure			
Actual Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Non-Carcinogen	Non-Carcinogen
Distance	0.562	0.563	0.576
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

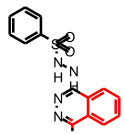
Model Applicability

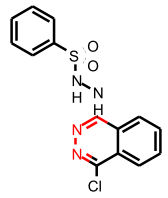
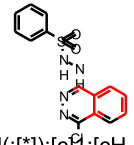
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

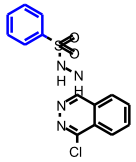
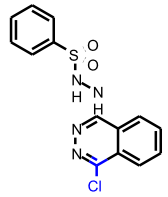
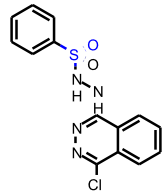
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1651620003	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[cH]:[c]:1:[*]	0.643	7 out of 10

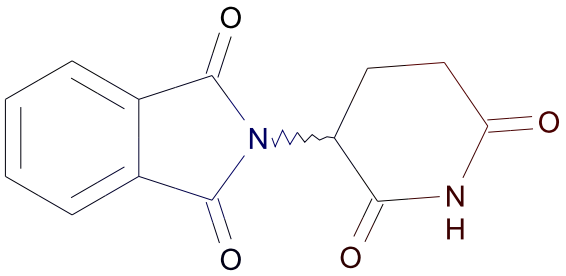
SCFP_6	149212520	 [*][c](:[*]):n:n:[*]	0.543	9 out of 15
SCFP_6	-1379673609	 [*][c](:[*]):[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	0.526	11 out of 19

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.504	12 out of 64
SCFP_6	-52074512	 [*]:[c](:[*])Cl	-0.315	14 out of 61
SCFP_6	1311429347	 [*]S(=[*])(=O)[*]	-0.283	7 out of 30

Thalidomide

TOPKAT_Rat_Male_FDA_None_vs_Carcinogen



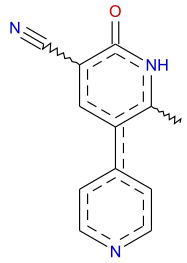
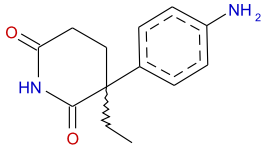
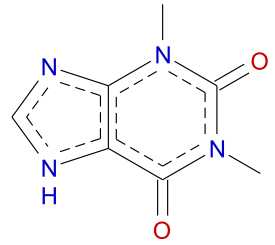
$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

Model Prediction

Prediction: Non-Carcinogen
Probability: 0.332
Enrichment: 0.995
Bayesian Score: -0.857
Mahalanobis Distance: 13
Mahalanobis Distance p-value: 0.00164

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

Structural Similar Compounds

Name	Milrinone	Aminoglutethimide	Theophylline
Structure			
Actual Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Predicted Endpoint	Non-Carcinogen	Carcinogen	Non-Carcinogen
Distance	0.582	0.597	0.614
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

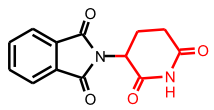
Model Applicability

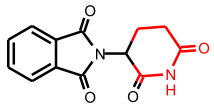
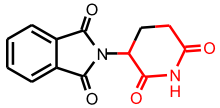
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

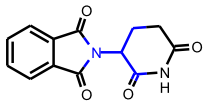
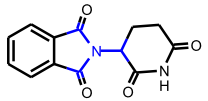
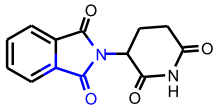
Feature Contribution

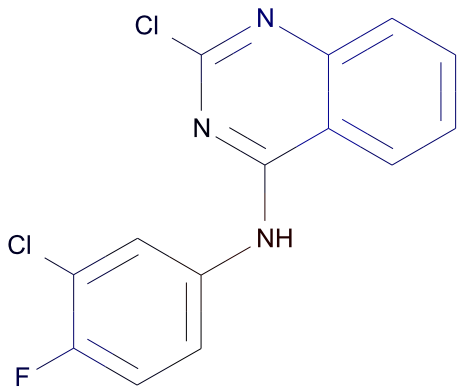
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1533214013	 [*]C1CCC(=O)NC1=O	0.415	1 out of 1

SCFP_6	345243876	 [*]=C1[*]CCC(=O)N1	0.415	1 out of 1
SCFP_6	395945879	 [*]C1[*]CC(=O)NC1=O	0.415	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Carcinogen in training set
SCFP_6	-1946889102	 [*]CC(N([*])[*])C(=[*])[*]	-0.885	1 out of 12
SCFP_6	399659969	 [*]C([*])N1C(=[*])[*]:[*]C1=[*]	-0.578	1 out of 8
SCFP_6	1257084377	 [*]N1[*][*]:[c](:[*]) C1=O	-0.436	4 out of 21



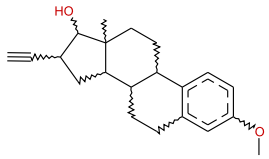
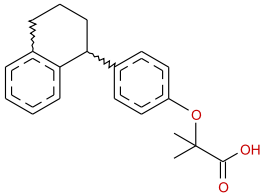
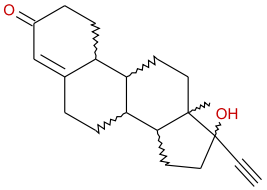
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.436
Enrichment: 1.05
Bayesian Score: -5.26
Mahalanobis Distance: 17.2
Mahalanobis Distance p-value: 9.6e-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	Mestranol	Nafenopin	Norethindrone
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.620	0.621	0.686
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

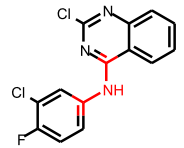
Model Applicability

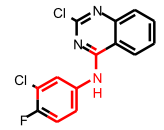
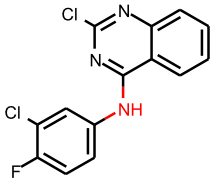
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

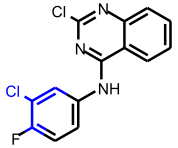
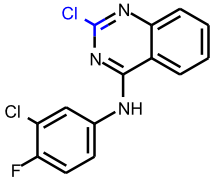
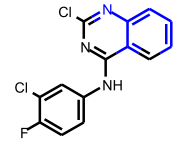
Feature Contribution

Top features for positive contribution

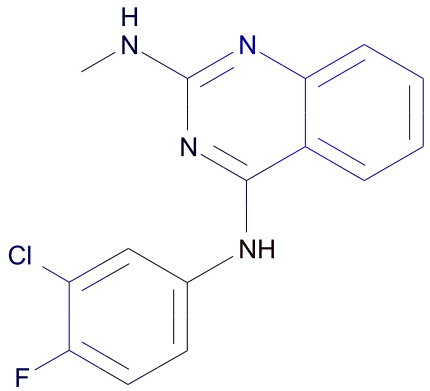
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	951581613	 [*]:[c](:[*])N[c](:[*])[*]	0.383	1 out of 1

SCFP_8	-300914917	 [*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.383	1 out of 1
SCFP_8	10	 [*]N[*]	0.226	18 out of 39

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-601571304	 [*][c](:[*]):[c](Cl): [cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14
SCFP_8	-1381862798	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.572	1 out of 7

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.412

Enrichment: 0.995

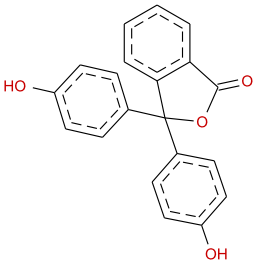
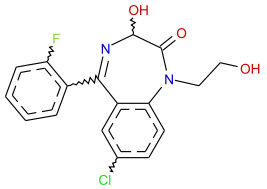
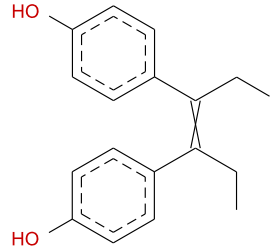
Bayesian Score: -5.85

Mahalanobis Distance: 18.4

Mahalanobis Distance p-value: 8.86e-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	Phenolphthalein	Doxefazepam	Diethylstilbesterol
Structure			
Actual Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.599	0.679	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

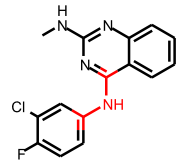
Model Applicability

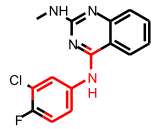
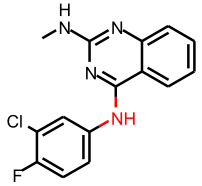
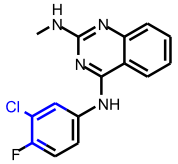
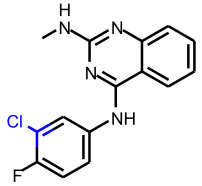
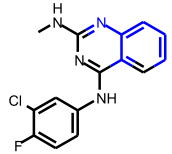
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

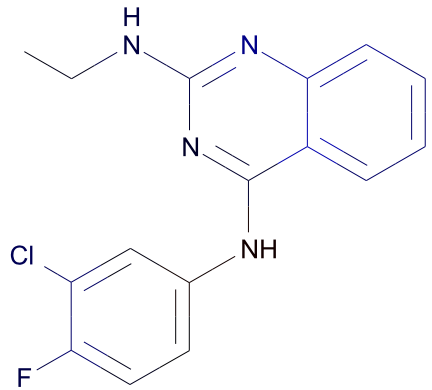
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	951581613	 [*]:[c](:[*])N[c](:[*])[*]	0.383	1 out of 1

SCFP_8	-300914917	 [*][c]1:[*]:[cH]:[cH] :[c](N[c](:[*]):[*]) :[cH]:1	0.383	1 out of 1
SCFP_8	10	 [*]N[*]	0.226	18 out of 39
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Multiple-Carcinogen in training set**
SCFP_8	-601571304	 [*][c](:[*]):[c](Cl): [cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14
SCFP_8	-1381862798	 [*]:n:[c]1:[cH]:[cH]: [cH]:[*]:[c]:1:[*]	-0.572	1 out of 7

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



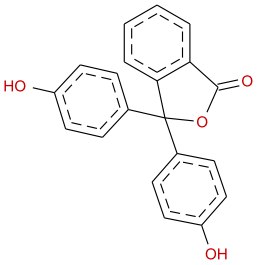
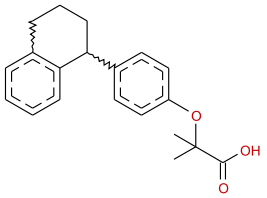
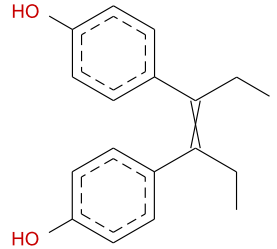
$C_{16}H_{14}ClFN_4$
Molecular Weight: 316.76056
ALogP: 5.025
Rotatable Bonds: 4
Acceptors: 4
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.391
Enrichment: 0.945
Bayesian Score: -6.32
Mahalanobis Distance: 19.2
Mahalanobis Distance p-value: 1.65e-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	Phenolphthalein	Nafenopin	Diethylstilbesterol
Structure			
Actual Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.665	0.686	0.686
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

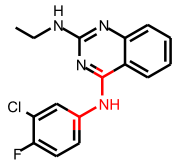
Model Applicability

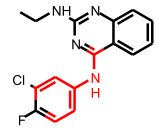
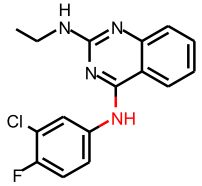
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

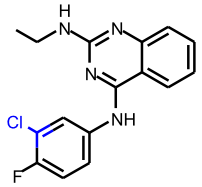
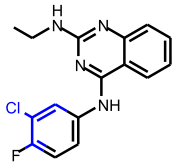
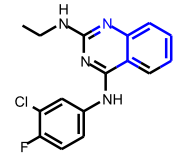
Feature Contribution

Top features for positive contribution

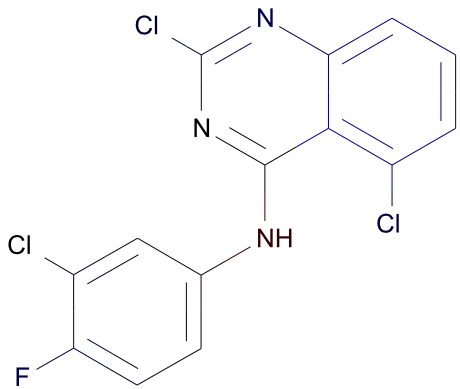
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	951581613	 [*]:[c](:[*])N[c](:[*]):[*]	0.383	1 out of 1

SCFP_8	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	0.383	1 out of 1
SCFP_8	10	 [*]N[*]	0.226	18 out of 39

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14
SCFP_8	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.707	2 out of 14
SCFP_8	-1381862798	 [*]:n:[c]1:[cH]:[cH]:[cH]:[*]:[c]:1:[*]	-0.572	1 out of 7

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Single-Carcinogen

Probability: 0.417

Enrichment: 1.01

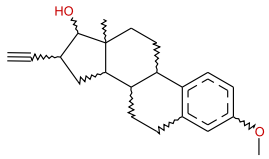
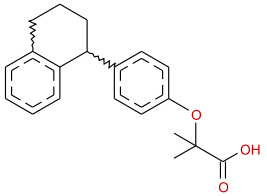
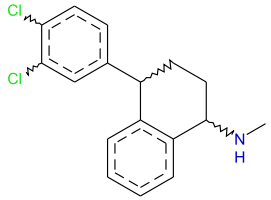
Bayesian Score: -5.74

Mahalanobis Distance: 18.5

Mahalanobis Distance p-value: 7.37e-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	Mestranol	Nafenopin	Sertraline
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.653	0.662	0.706
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

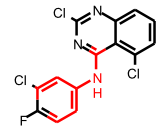
Model Applicability

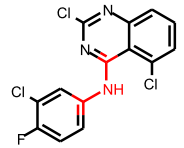
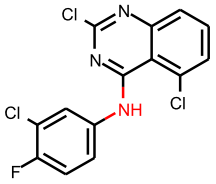
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

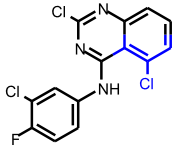
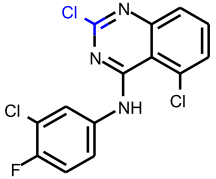
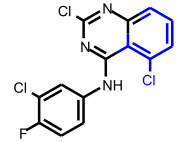
Feature Contribution

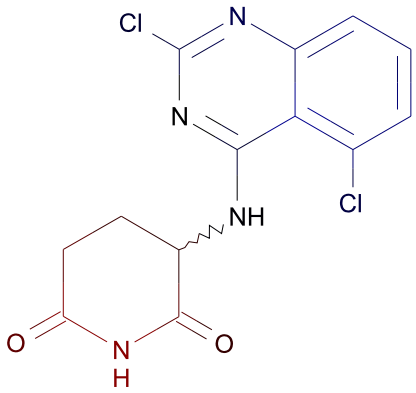
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	0.383	1 out of 1

SCFP_8	951581613	 [*]:[c](:[*])N[c](:[*]):[*]	0.383	1 out of 1
SCFP_8	10	 [*]N[*]	0.226	18 out of 39

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14
SCFP_8	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.58	2 out of 12



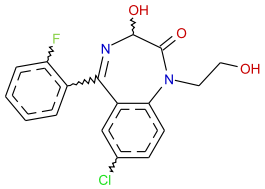
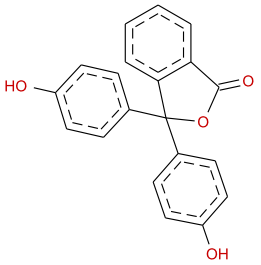
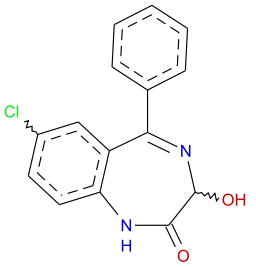
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.535
Enrichment: 1.29
Bayesian Score: -2
Mahalanobis Distance: 17.4
Mahalanobis Distance p-value: 6.3e-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	Doxefazepam	Phenolphthalein	Oxazepam
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.597	0.663	0.666
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

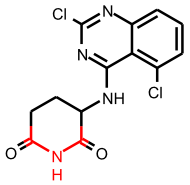
Model Applicability

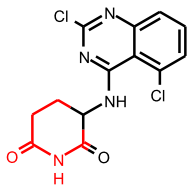
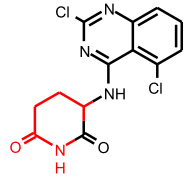
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

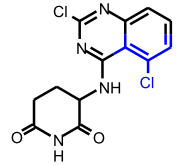
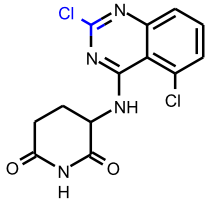
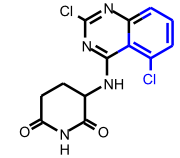
Feature Contribution

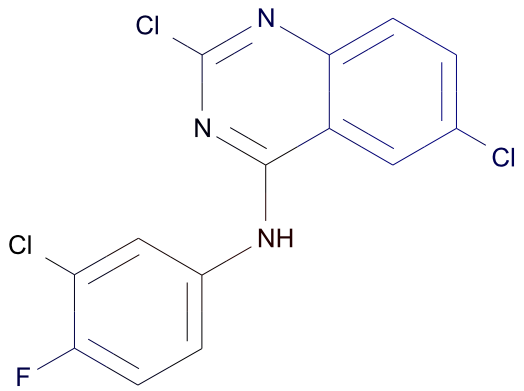
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	1631785938	 [*]C(=[*])NC(=[*])[*]	0.574	4 out of 5

SCFP_8	345243876	 [*]=C1[*]CCC(=O)N1	0.383	1 out of 1
SCFP_8	-1072897324	 [*]C1[*]NC(=O)CC1	0.383	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14
SCFP_8	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.58	2 out of 12



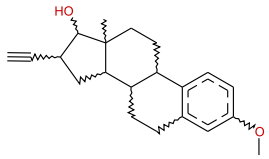
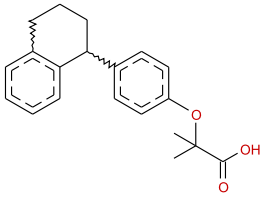
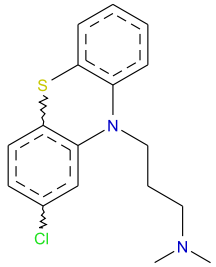
$C_{14}H_7Cl_3FN_3$
Molecular Weight: 342.58288
ALogP: 5.867
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.413
Enrichment: 0.999
Bayesian Score: -5.82
Mahalanobis Distance: 18.5
Mahalanobis Distance p-value: 7.12e-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	Mestranol	Nafenopin	Chlorpromazine
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Multiple-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.653	0.667	0.707
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

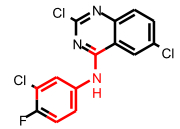
Model Applicability

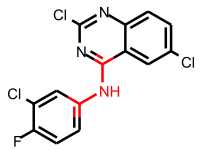
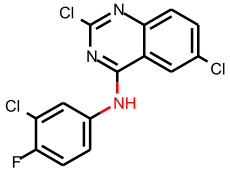
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

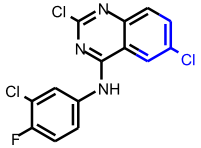
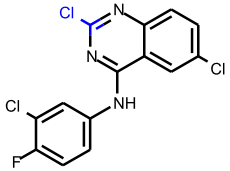
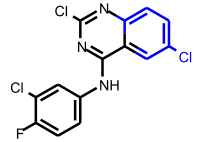
Feature Contribution

Top features for positive contribution

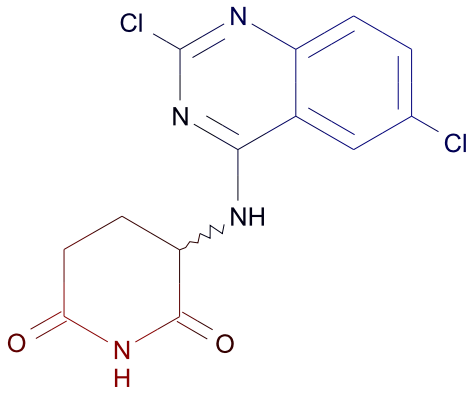
Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-300914917	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	0.383	1 out of 1

SCFP_8	951581613	 [*]:[c](:[*])N[c](:[*]):[*]	0.383	1 out of 1
SCFP_8	10	 [*]N[*]	0.226	18 out of 39

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.707	2 out of 14
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14
SCFP_8	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.58	2 out of 12

TOPKAT_Rat_Male_FDA_Single_vs_Multiple



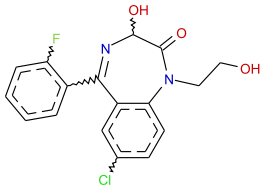
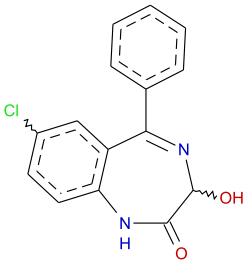
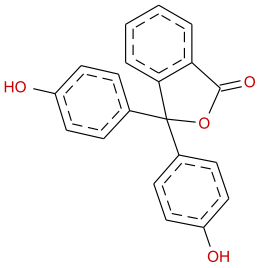
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.533
Enrichment: 1.29
Bayesian Score: -2.08
Mahalanobis Distance: 17
Mahalanobis Distance p-value: 1.43e-007

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

Structural Similar Compounds

Name	Doxefazepam	Oxazepam	Phenolphthalein
Structure			
Actual Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Single-Carcinogen	Multiple-Carcinogen
Distance	0.597	0.666	0.668
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

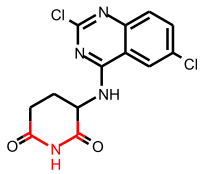
Model Applicability

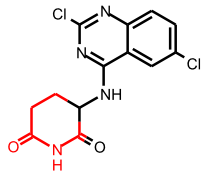
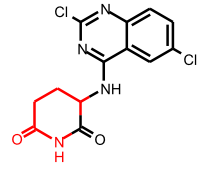
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

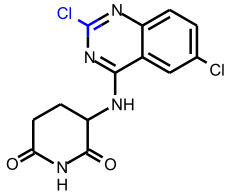
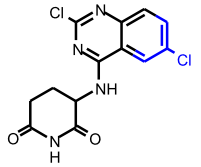
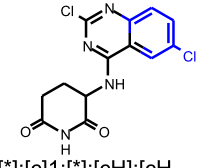
Feature Contribution

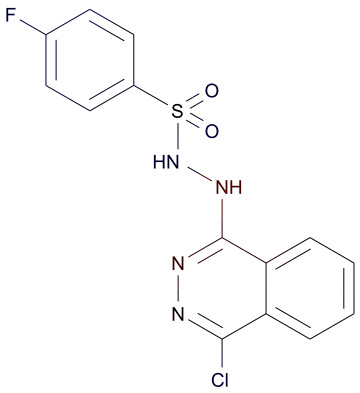
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	1631785938	 [*]C(=[*])NC(=[*])[*]	0.574	4 out of 5

SCFP_8	345243876	 [*]=C1[*]CCC(=O)N1	0.383	1 out of 1
SCFP_8	-1072897324	 [*]C1[*]NC(=O)CC1	0.383	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14
SCFP_8	-601571304	 [*][c](:[*]):[c](Cl):[cH]:[*]	-0.707	2 out of 14
SCFP_8	-1378360678	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[c]:1Cl	-0.58	2 out of 12



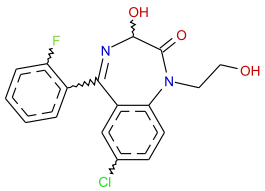
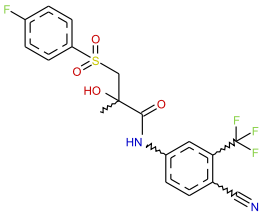
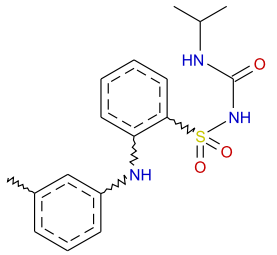
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.522
Enrichment: 1.26
Bayesian Score: -2.59
Mahalanobis Distance: 14.7
Mahalanobis Distance p-value: 1.95e-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	Doxefazepam	Bicalutamide	Torsemide
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Multiple-Carcinogen
Distance	0.615	0.635	0.669
Reference	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997	US FDA (Centre for Drug Eval.& Res./Off. Testing & Res.) Sept. 1997

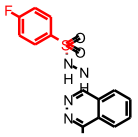
Model Applicability

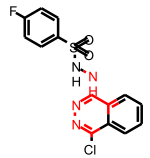
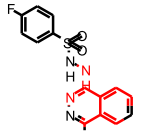
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

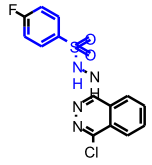
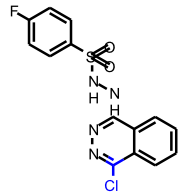
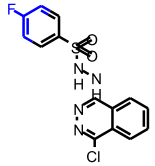
Feature Contribution

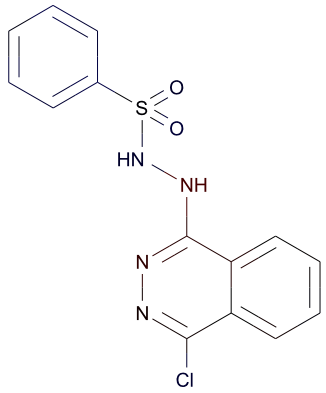
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	2144224020	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.383	1 out of 1

SCFP_8	-705181820	 [*]N[c]1:n:n:[c]([*]):[*]:[c]:1:[*]	0.383	1 out of 1
SCFP_8	-1038366601	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.383	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.737	0 out of 3
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14
SCFP_8	-730654023	 [*][c](:[*]):[c](F):[cH]:[*]	-0.463	1 out of 6



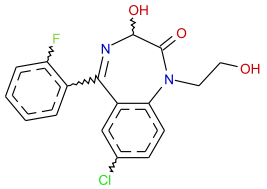
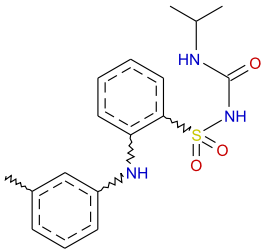
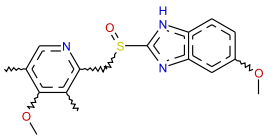
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Single-Carcinogen
Probability: 0.522
Enrichment: 1.26
Bayesian Score: -2.6
Mahalanobis Distance: 15.3
Mahalanobis Distance p-value: 5.78e-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	Doxefazepam	Torsemide	Omeprazole
Structure			
Actual Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Predicted Endpoint	Single-Carcinogen	Multiple-Carcinogen	Single-Carcinogen
Distance	0.636	0.666	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

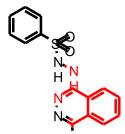
Model Applicability

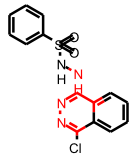
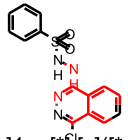
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

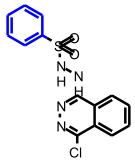
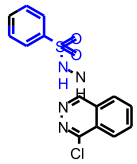
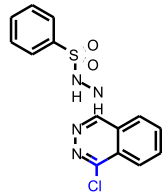
Feature Contribution

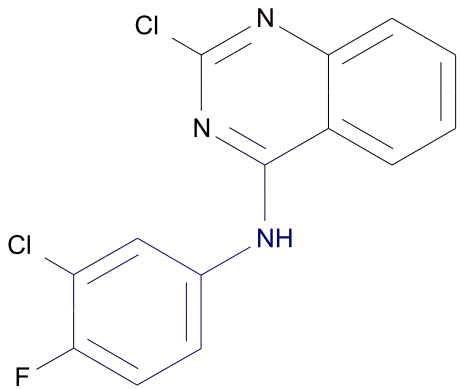
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	2097816882	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[cH]:[cH]:[cH]:[c]:1:2	0.383	1 out of 1

SCFP_8	-705181820	 [*]N[c]1:n:n:[c]([*]):[*]:[c]:1:[*]	0.383	1 out of 1
SCFP_8	-1038366601	 [*]N[c]1:n:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:1:2	0.383	1 out of 1

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Multiple-Carcinogen in training set
SCFP_8	1653911926	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.985	1 out of 12
SCFP_8	-1247518081	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[*]:[cH]:[cH]:1	-0.737	0 out of 3
SCFP_8	-52074512	 [*]:[c](:[*])Cl	-0.707	2 out of 14



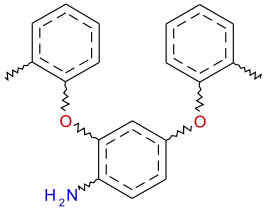
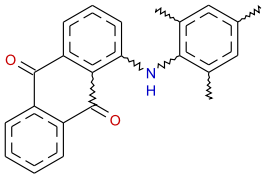
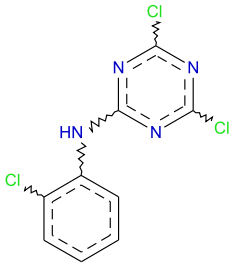
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 0.916
Enrichment: 0.995
Bayesian Score: -2.55
Mahalanobis Distance: 7.24
Mahalanobis Distance p-value: 0.989

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

Structural Similar Compounds

Name	Aniline, 2,4-bis(o-methylphenoxy)-	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	s-Triazine, 2,4-dichloro-6-(o-chloroanilino)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.566	0.568	0.584
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,242,1	34ZIAG* -,235,69

Model Applicability

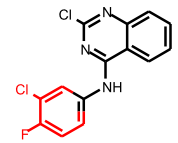
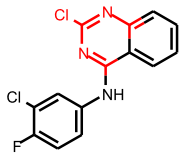
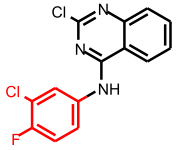
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

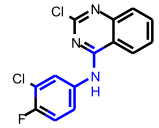
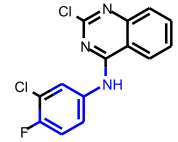
Feature Contribution

Top features for positive contribution

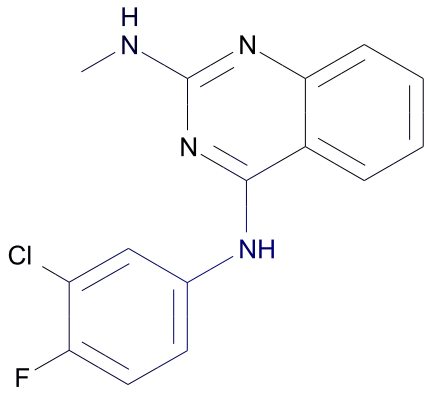
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1185376954	 F[c]1:[cH]:[cH]:[*]:[cH]:[c]:1Cl	0.0756	6 out of 6
FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.0703	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.708	4 out of 10
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.444	46 out of 79

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.547

Enrichment: 0.594

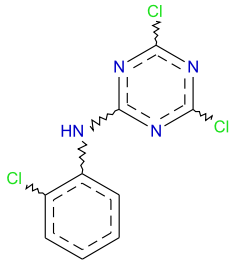
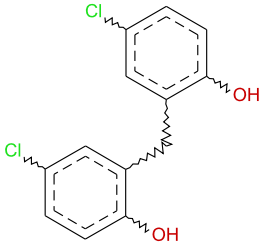
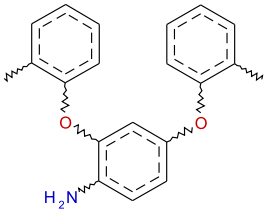
Bayesian Score: -4.42

Mahalanobis Distance: 7.41

Mahalanobis Distance p-value: 0.98

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

Structural Similar Compounds

Name	s-Triazine, 2,4-dichloro-6-(o-chloroanilino)-	Phenol, 2,2'-methylenebis(4-chloro-	Aniline, 2,4-bis(o-methylphenoxy)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.639	0.676	0.707
Reference	34ZIAG* -,235,69	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,533,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986

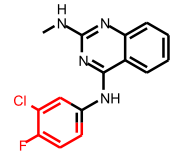
Model Applicability

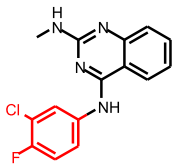
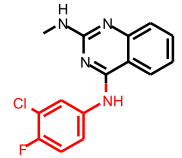
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

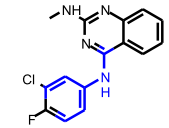
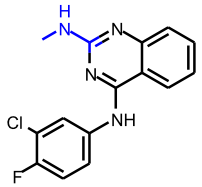
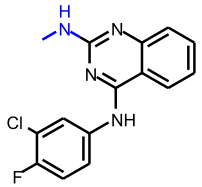
Feature Contribution

Top features for positive contribution

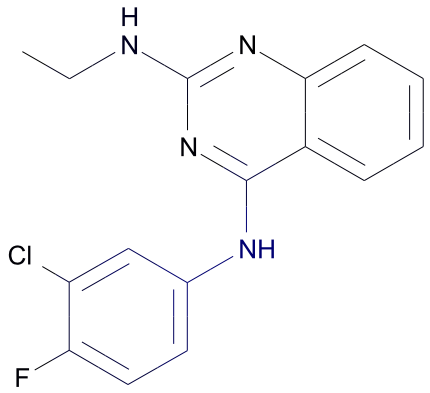
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1185376954	 F[c]1:[cH]:[cH]:[*]:[cH]:[c]:1Cl	0.0756	6 out of 6

FCFP_12	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.0703	4 out of 4
FCFP_12	-2050771350	 [*]N[c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.708	4 out of 10
FCFP_12	1294255210	 [*]:[c](:[*])NC	-0.486	12 out of 22
FCFP_12	136686699	 [*]NC	-0.484	3 out of 6

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{16}H_{14}ClFN_4$

Molecular Weight: 316.76056

ALogP: 5.025

Rotatable Bonds: 4

Acceptors: 4

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.732

Enrichment: 0.794

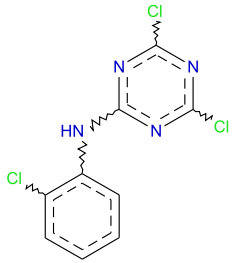
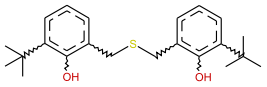
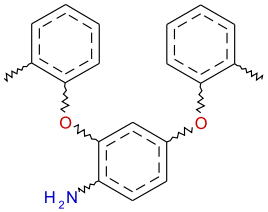
Bayesian Score: -3.82

Mahalanobis Distance: 8.93

Mahalanobis Distance p-value: 0.456

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

Structural Similar Compounds

Name	s-Triazine, 2,4-dichloro-6-(o-chloroanilino)-	Sulfide, bis(4-t-butyl-m-cresyl)-	Aniline, 2,4-bis(o-methylphenoxy)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.674	0.689	0.699
Reference	34ZIAG* -,235,69	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	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986

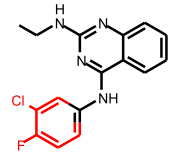
Model Applicability

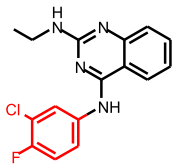
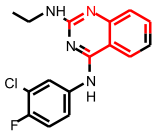
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

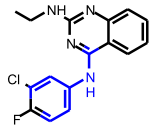
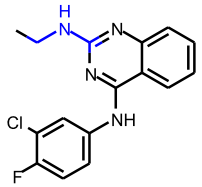
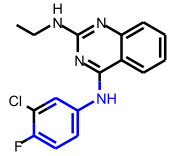
Feature Contribution

Top features for positive contribution

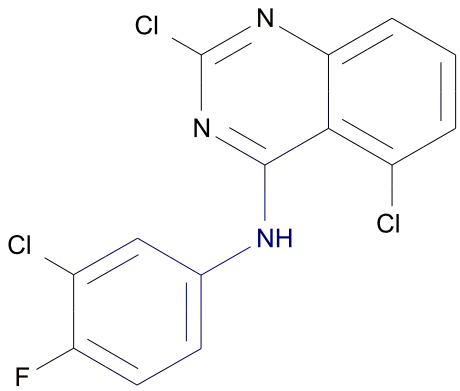
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1185376954	 F[c]1:[cH]:[cH]:[*]:[cH]:[c]:1Cl	0.0756	6 out of 6

FCFP_12	-2132756387	 [*][c]1:[cH]:[cH]:[c](F):[c](Cl):[cH]:1	0.0703	4 out of 4
FCFP_12	-1716224640	 [*][c]1:[*]:[c]([*]):[c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.708	4 out of 10
FCFP_12	1294255210	 [*]:[c](:[*])NC	-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



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: Non-Irritant

Probability: 0.926

Enrichment: 1.01

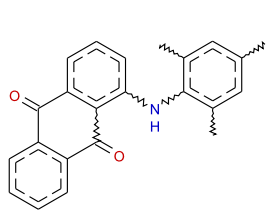
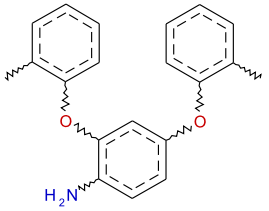
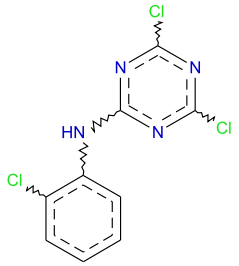
Bayesian Score: -2.4

Mahalanobis Distance: 7.4

Mahalanobis Distance p-value: 0.981

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

Structural Similar Compounds

Name	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	Aniline, 2,4-bis(o-methylphenoxy)-	s-Triazine, 2,4-dichloro-6-(o-chloroanilino)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.513	0.610	0.655
Reference	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,242,1	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986	34ZIAG* -,235,69

Model Applicability

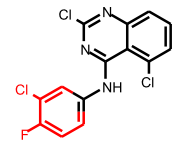
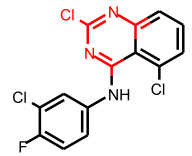
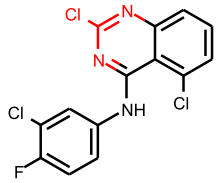
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

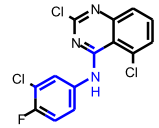
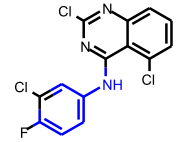
Feature Contribution

Top features for positive contribution

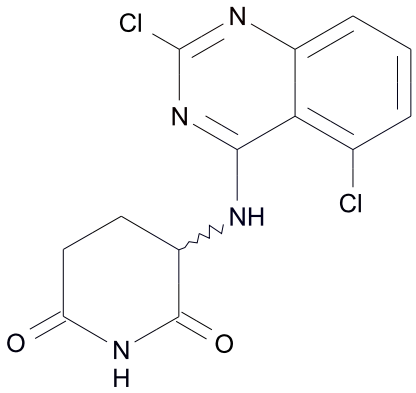
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1185376954	 F[c]1:[cH]:[cH]:[*]:[cH]:[c]:1Cl	0.0756	6 out of 6
FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	-1374842118	 [*]:n:[c](Cl):n:[*]	0.0703	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	839741273	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.708	4 out of 10
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.444	46 out of 79

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.953

Enrichment: 1.03

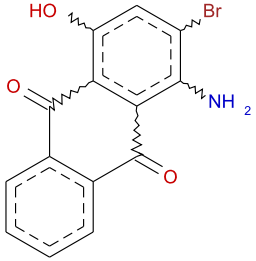
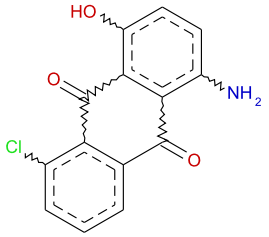
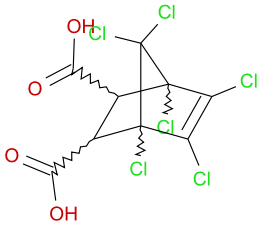
Bayesian Score: -1.8

Mahalanobis Distance: 8.93

Mahalanobis Distance p-value: 0.459

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

Structural Similar Compounds

Name	1-Amino-2-bromo-4-hydroxyanthraquinone	1-Amino-4-hydroxy-5-chloroanthraquinone	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.590	0.625	0.655
Reference	28ZPAK -,83,72	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986

Model Applicability

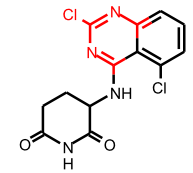
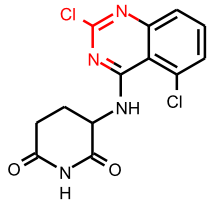
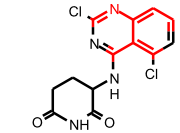
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

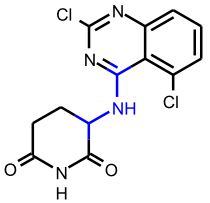
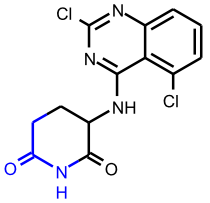
Feature Contribution

Top features for positive contribution

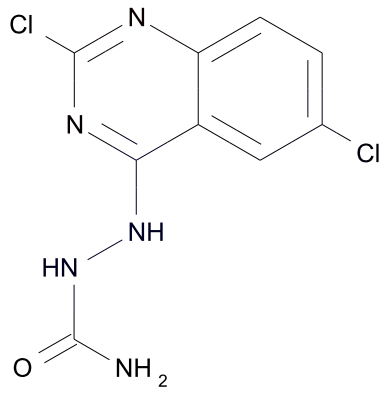
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	-1374842118	 [*]:n:[c](Cl):n:[*]	0.0703	4 out of 4
FCFP_12	-1716224640	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294255210	 [*]:[c](:[*])NC	-0.486	12 out of 22
FCFP_12	566058135	 [*]CC(=O)N[*]	-0.367	13 out of 21

TOPKAT_Skin_Irritancy_None_vs_Irritant



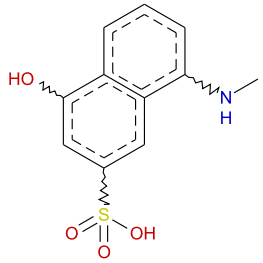
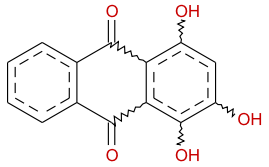
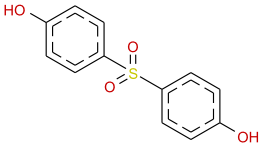
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09078
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Irritant
Probability: 0.958
Enrichment: 1.04
Bayesian Score: -1.63
Mahalanobis Distance: 6.44
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	8-Methylamino-4-hydroxy-2-naphthalene sulfonic acid	Anthraquinone, 1,2,4-trihydroxy-	Phenol, 4,4'-sulfonyldi-
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.584	0.615	0.681
Reference	28ZPAK -,190,72	28ZPAK -,103,72	BIOFX* BIOFAX Industrial Bio-Test Laboratories, Inc., Data Sheets. (1810 Fro ntage Rd., Northbrook, IL 60062) Volume(issue)/page/year: 601-05501,1974

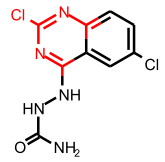
Model Applicability

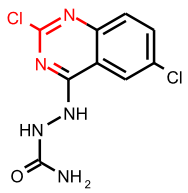
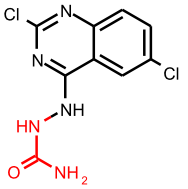
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

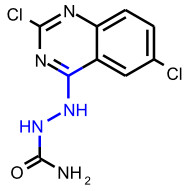
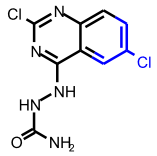
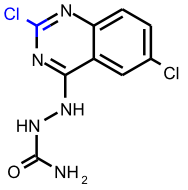
Feature Contribution

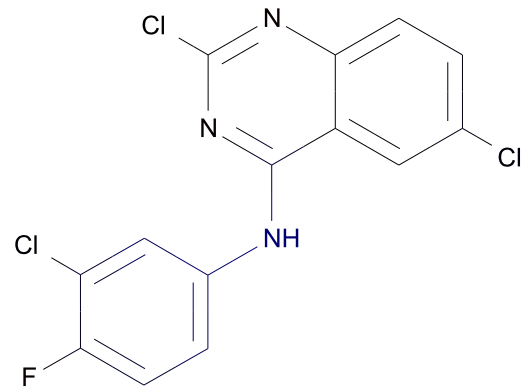
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	0.0703	4 out of 4

FCFP_12	-1374842118	 [*]:n:[c](Cl):n:[*]	0.0703	4 out of 4
FCFP_12	1499521844	 [*]NC(=O)N	0.0658	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	 [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	367998008	 [*][c](:[*]):[c](F):[cH]:[*]	-0.129	61 out of 76
FCFP_12	71476542	 [*]:[c](:[*])Cl	-0.12	64 out of 79



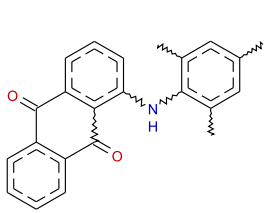
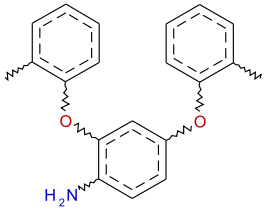
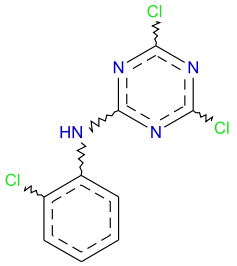
$C_{14}H_7Cl_3FN_3$
Molecular Weight: 342.58288
ALogP: 5.867
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: Non-Irritant
Probability: 0.931
Enrichment: 1.01
Bayesian Score: -2.32
Mahalanobis Distance: 7.4
Mahalanobis Distance p-value: 0.981

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

Structural Similar Compounds

Name	Anthraquinone, 1-(2,4,6-trimethylphenylamino)-	Aniline, 2,4-bis(o-methylphenoxy)-	s-Triazine, 2,4-dichloro-6-(o-chloroanilino)-
Structure			
Actual Endpoint	Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.518	0.615	0.664
Reference	28ZPAK "Sbornik Vysledku Toxixologickeho Vysetreni Latek A Pripravku," Marhol d, J.V., Institut Pro Vychovu Vedoucicn Pracovniku Chemickeho Prumyclu Praha, Cz echoslovakia, 1972 Volume(issue)/page/year: -,242,1	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,725,1986	34ZIAG* -,235,69

Model Applicability

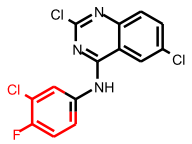
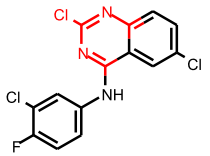
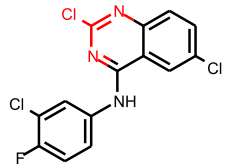
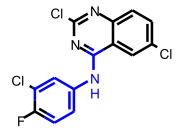
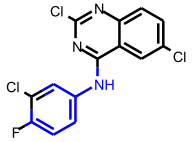
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

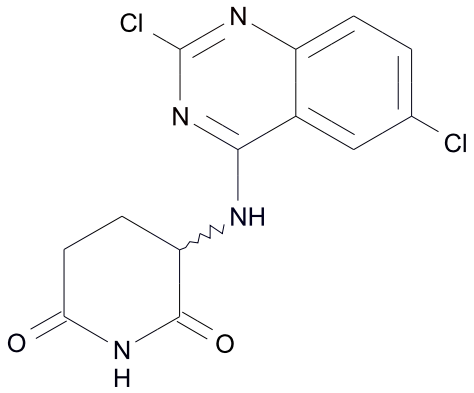
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1185376954	 F[c]1:[cH]:[cH]:[*]:[cH]:[c]:1Cl	0.0756	6 out of 6
FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	-1374842118	 [*]:n:[c](Cl):n:[*]	0.0703	4 out of 4
Top Features for negative contribution				
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**	**Irritant in training set**
FCFP_12	839741273	 [*][c]1:[*]:[cH]:[cH]:[c](N[c](:[*]):[*]):[cH]:1	-0.708	4 out of 10
FCFP_12	-773983804	 [*]N[c]1:[cH]:[*]:[c]([*]):[cH]:[cH]:1	-0.444	46 out of 79

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.955

Enrichment: 1.04

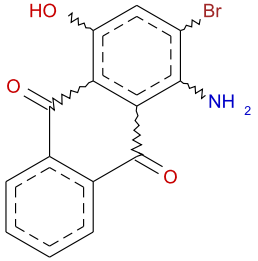
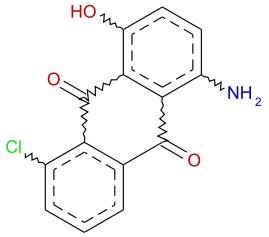
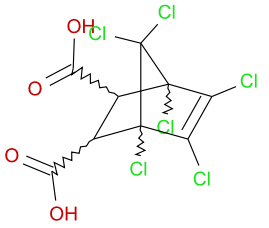
Bayesian Score: -1.72

Mahalanobis Distance: 8.93

Mahalanobis Distance p-value: 0.459

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

Structural Similar Compounds

Name	1-Amino-2-bromo-4-hydroxyanthraquinone	1-Amino-4-hydroxy-5-chloroanthraquinone	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Irritant
Distance	0.596	0.631	0.655
Reference	28ZPAK -,83,72	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986

Model Applicability

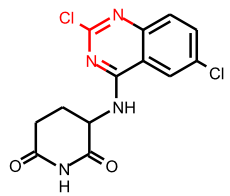
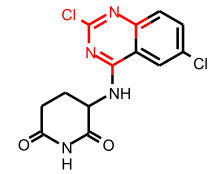
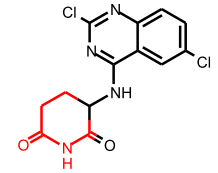
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

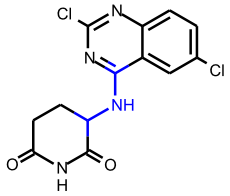
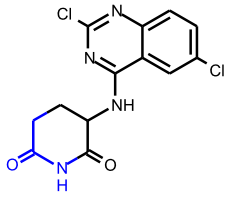
Feature Contribution

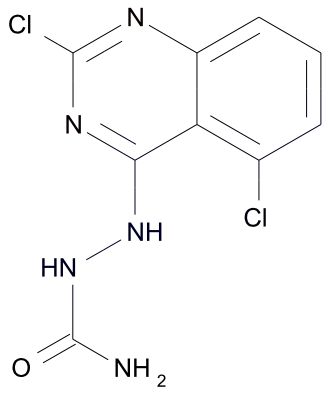
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1374842118	 [*]:n:[c](Cl):n:[*]	0.0703	4 out of 4
FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	-922480536	 [*]=C1[*]CCC(=O)N1	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294255210	 [*]:[c](:[*])NC	-0.486	12 out of 22
FCFP_12	566058135	 [*]CC(=O)N[*]	-0.367	13 out of 21



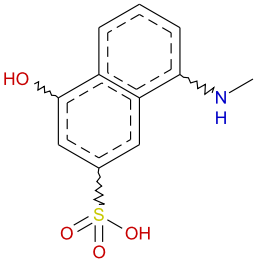
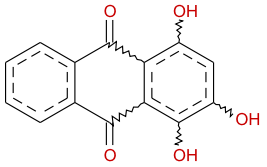
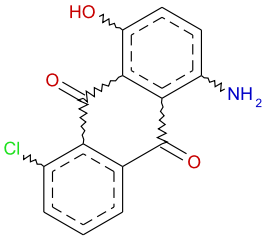
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: Non-Irritant
Probability: 0.955
Enrichment: 1.04
Bayesian Score: -1.71
Mahalanobis Distance: 6.44
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	8-Methylamino-4-hydroxy-2-naphthalene sulfonic acid	Anthraquinone, 1,2,4-trihydroxy-	1-Amino-4-hydroxy-5-chloroanthraquinone
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.577	0.608	0.680
Reference	28ZPAK -,190,72	28ZPAK -,103,72	28ZPAK -,83,72

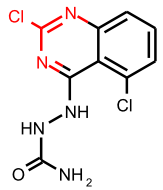
Model Applicability

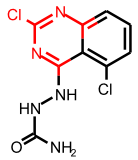
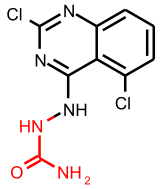
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

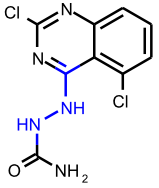
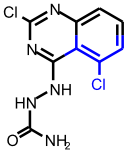
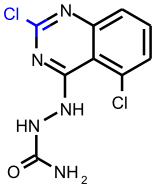
Feature Contribution

Top features for positive contribution

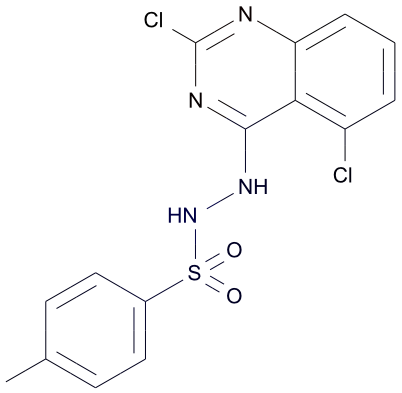
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	-1374842118	 [*]:n:[c](Cl):n:[*]	0.0703	4 out of 4

FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	1499521844	 [*]NC(=O)N	0.0658	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	 [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	367998008	 [*][c](:[*]):[c](F):[cH]:[*]	-0.129	61 out of 76
FCFP_12	71476542	 [*]:[c](:[*])Cl	-0.12	64 out of 79

TOPKAT_Skin_Irritancy_None_vs_Irritant



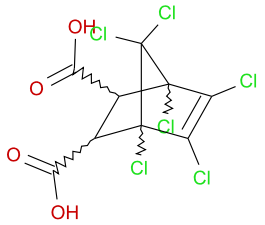
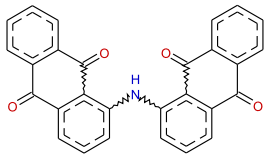
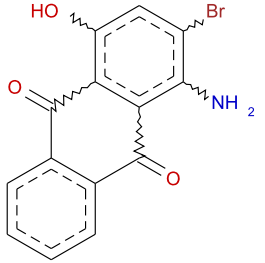
$C_{15}H_{12}Cl_2N_4O_2S$
Molecular Weight: 383.25237
ALogP: 4.443
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.927
Enrichment: 1.01
Bayesian Score: -2.38
Mahalanobis Distance: 6.03
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	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	Anthraquinone, 1,1'-iminodi-	1-Amino-2-bromo-4-hydroxyanthraquinone
Structure			
Actual Endpoint	Irritant	Irritant	Non-Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.722	0.734	0.755
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986	28ZPAK -,83,72

Model Applicability

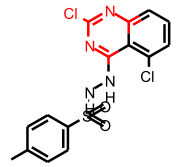
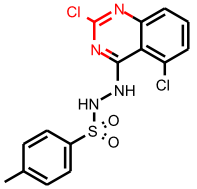
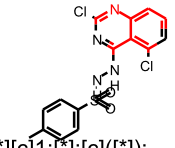
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

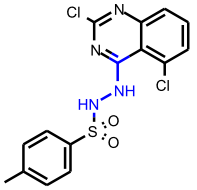
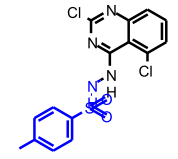
Feature Contribution

Top features for positive contribution

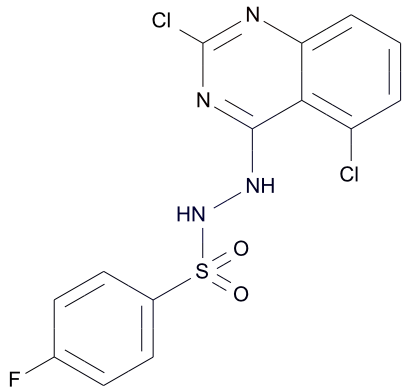
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]) :n:[c](Cl):n:1	0.0703	4 out of 4
FCFP_12	-1374842118	 [*]:n:[c](Cl):n:[*]	0.0703	4 out of 4
FCFP_12	-1716224640	 [*][c]1:[*]:[c]([*]): [c]2:[cH]:[*]:[cH]:[cH]:[c]:2:n:1	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	 [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	1789442672	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.347	1 out of 2

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{14}H_9Cl_2FN_4O_2S$

Molecular Weight: 387.21626

ALogP: 4.162

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.942

Enrichment: 1.02

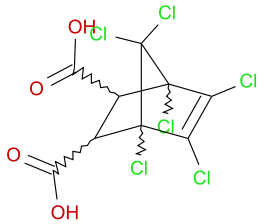
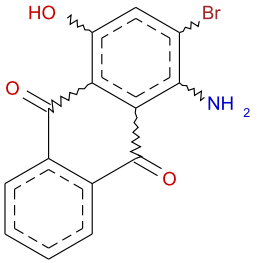
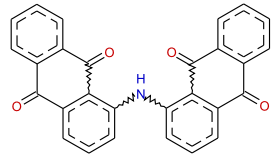
Bayesian Score: -2.09

Mahalanobis Distance: 5.98

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	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-2-bromo-4-hydroxyanthraquinone	Anthraquinone, 1,1'-iminodi-
Structure			
Actual Endpoint	Irritant	Non-Irritant	Irritant
Predicted Endpoint	Irritant	Non-Irritant	Non-Irritant
Distance	0.704	0.743	0.744
Reference	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,735,1986

Model Applicability

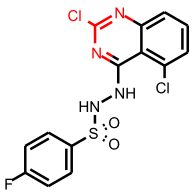
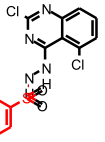
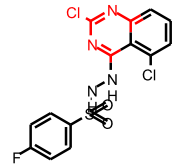
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

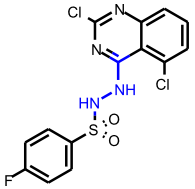
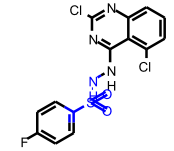
Feature Contribution

Top features for positive contribution

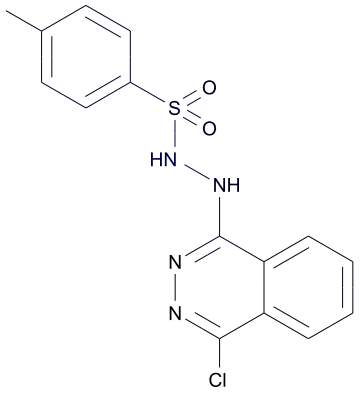
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	-1374842118	 [*]:n:[c](Cl):n:[*]	0.0703	4 out of 4
FCFP_12	-149636017	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.0703	4 out of 4
FCFP_12	-2069374281	 [*][c]1:[*]:[c](:[*]):n:[c](Cl):n:1	0.0703	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	 [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	-1096219292	 [*]NS(=O)(=O)[c](:[*]):[*]	-0.229	26 out of 36

TOPKAT_Skin_Irritancy_None_vs_Irritant



$C_{15}H_{13}ClN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: Non-Irritant

Probability: 0.916

Enrichment: 0.995

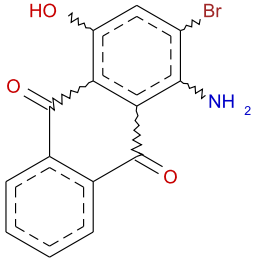
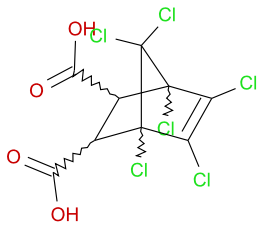
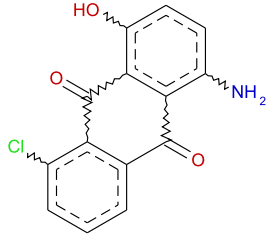
Bayesian Score: -2.55

Mahalanobis Distance: 6

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	1-Amino-2-bromo-4-hydroxyanthraquinone	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-4-hydroxy-5-chloroanthraquinone
Structure			
Actual Endpoint	Non-Irritant	Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.687	0.688	0.732
Reference	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72

Model Applicability

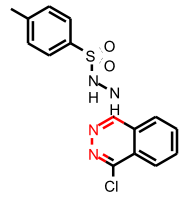
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

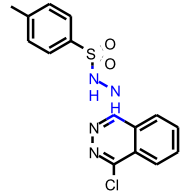
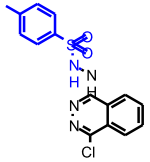
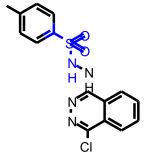
Feature Contribution

Top features for positive contribution

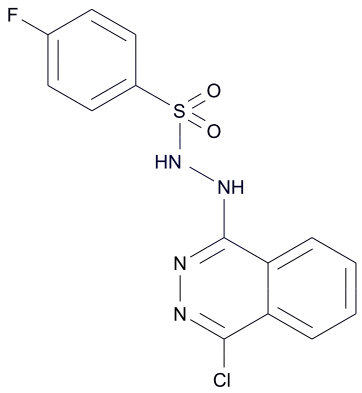
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	4427049	 [*][c](:[*]):n:n:[*]	0.0734	5 out of 5

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	 [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	1789442672	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[c](C):[cH]:[cH]:1	-0.347	1 out of 2
FCFP_12	-1096219292	 [*]NS(=O)(=O)[c](:[*]):[*]	-0.229	26 out of 36

TOPKAT_Skin_Irritancy_None_vs_Irritant



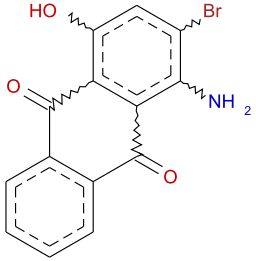
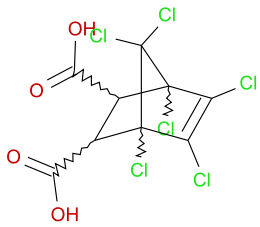
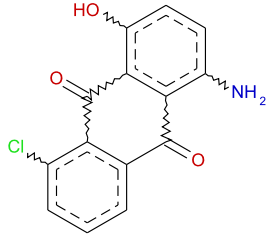
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.922
Enrichment: 1
Bayesian Score: -2.47
Mahalanobis Distance: 5.97
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	1-Amino-2-bromo-4-hydroxyanthraquinone	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	1-Amino-4-hydroxy-5-chloroanthraquinone
Structure			
Actual Endpoint	Non-Irritant	Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.674	0.679	0.717
Reference	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	28ZPAK -,83,72

Model Applicability

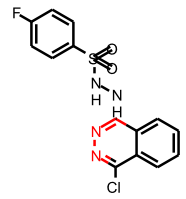
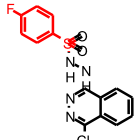
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

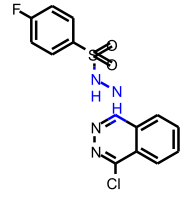
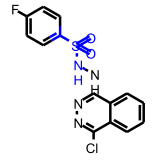
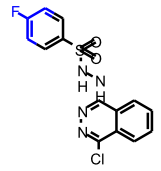
Feature Contribution

Top features for positive contribution

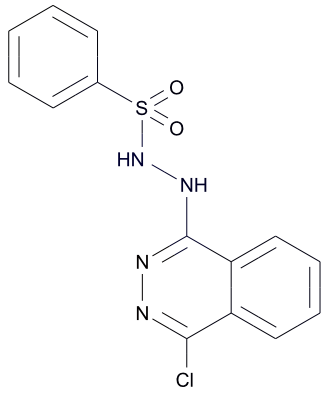
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	4427049	 [*][c](:[*]):n:n:[*]	0.0734	5 out of 5
FCFP_12	-149636017	 [*]S(=[*])(=[*])[c]1:[cH]:[cH]:[c](F):[cH]:[cH]:1	0.0703	4 out of 4

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	 [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	-1096219292	 [*]NS(=O)(=O)[c](:[*]):[*]	-0.229	26 out of 36
FCFP_12	367998008	 [*][c](:[*]):[c](F):[cH]:[*]	-0.129	61 out of 76

TOPKAT_Skin_Irritancy_None_vs_Irritant



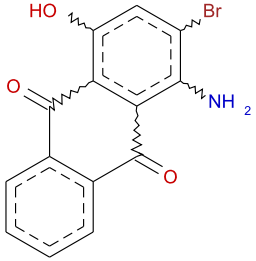
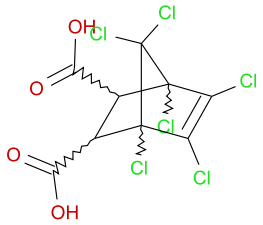
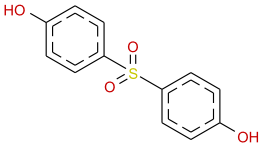
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: Non-Irritant
Probability: 0.94
Enrichment: 1.02
Bayesian Score: -2.14
Mahalanobis Distance: 6
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	1-Amino-2-bromo-4-hydroxyanthraquinone	5-Norbornene-2,3-dicarboxylic acid, 1,4,5,6,7,7-hexachloro-	Phenol, 4,4'-sulfonyldi-
Structure			
Actual Endpoint	Non-Irritant	Irritant	Irritant
Predicted Endpoint	Non-Irritant	Irritant	Non-Irritant
Distance	0.655	0.693	0.694
Reference	28ZPAK -,83,72	85JCAE "Prehled Prumyslove Toxikologie; Organicke Latky," Marhold, J., Prague , Czechoslovakia, Avicenum, 1986 Volume(issue)/page/year: -,581,1986	BIOFX* BIOFAX Industrial Bio-Test Laboratories, Inc., Data Sheets. (1810 Fro ntage Rd., Northbrook, IL 60062) Volume(issue)/page/year: 601-05501,1974

Model Applicability

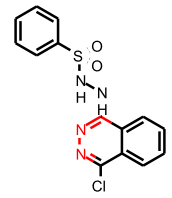
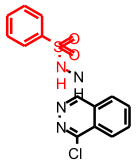
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

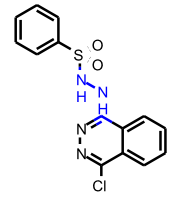
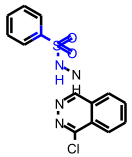
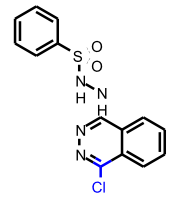
Feature Contribution

Top features for positive contribution

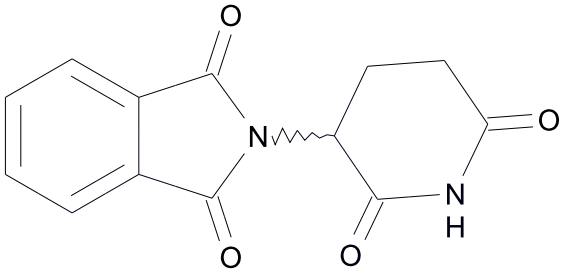
Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set

FCFP_12	4427049	 [*][c](:[*]):n:n:[*]	0.0734	5 out of 5
FCFP_12	-1089199451	 [*]NS(=O)(=O)[c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	0.0658	3 out of 3

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	1294344583	 [*]NN[c](:[*]):[*]	-0.65	0 out of 1
FCFP_12	-1096219292	 [*]NS(=O)(=O)[c](:[*]):[*]	-0.229	26 out of 36
FCFP_12	71476542	 [*]:[c](:[*])Cl	-0.12	64 out of 79

Thalidomide



$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

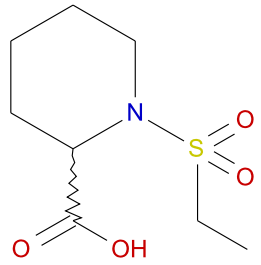
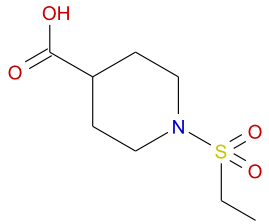
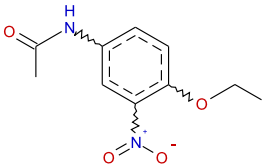
Model Prediction

Prediction: Non-Irritant
Probability: 0.964
Enrichment: 1.05
Bayesian Score: -1.37
Mahalanobis Distance: 11.6
Mahalanobis Distance p-value: 0.000214

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

TOPKAT_Skin_Irritancy_None_vs_Irritant

Structural Similar Compounds

Name	1-Methylsulfonyl-2-pipecoline	1-(ethylsulfonyl)-4-pipecoline	p-Acetophenetidide, 3'-nitro-
Structure			
Actual Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Predicted Endpoint	Non-Irritant	Non-Irritant	Non-Irritant
Distance	0.591	0.592	0.627
Reference	US ARMY	US ARMY	28ZPAK -,115,72

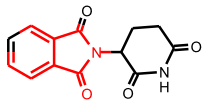
Model Applicability

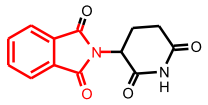
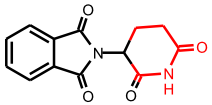
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

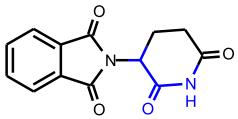
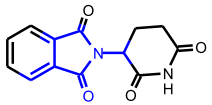
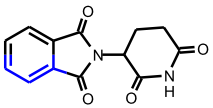
Feature Contribution

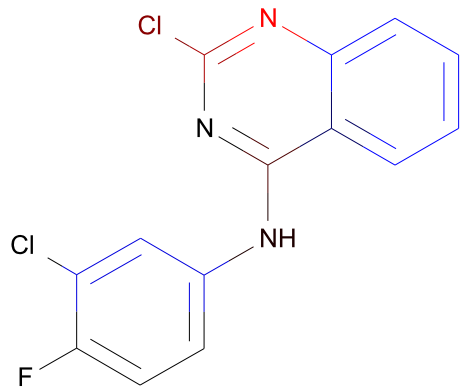
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	2139882011	 [*]N1C(=[*])[c]2:[cH]:[*]:[cH]:[cH]:[c]:2C1=O	0.081	11 out of 11

FCFP_12	1393189956	 [*]N1C(=[*])[c]2:[cH] :[cH]:[cH]:[cH]:[c]: 2C1=O	0.081	11 out of 11
FCFP_12	-922480536	 [*]=C1[*]CCC(=O)N1	0.0583	2 out of 2

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score	Irritant in training set
FCFP_12	566058135	 [*]CC(=O)N[*]	-0.367	13 out of 21
FCFP_12	392594677	 [*]C([*])N1C(=[*])[c] (:[*]):[c](:[cH]:[*])C1=O	-0.207	8 out of 11
FCFP_12	1618154665	 [*]:[cH]:[cH]:[cH]:[*]	-0.0845	412 out of 490



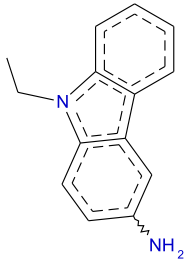
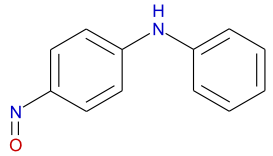
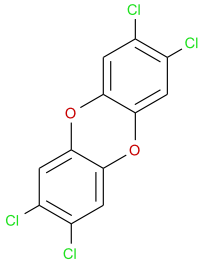
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 10.9
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 10.5
Mahalanobis Distance p-value: 0.0088

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	148	p-Nitrosodiphenylamine	2,3,7,8-Tetrachlorodibenzo-p-dioxin
Structure			
Actual Endpoint (-log C)	3.7362	2.76567	9.31469
Predicted Endpoint (-log C)	3.75282	3.37945	4.75869
Distance	0.685	0.701	0.704
Reference	CPDB	CPDB	CPDB

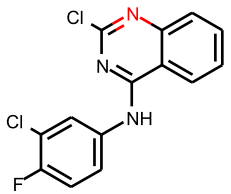
Model Applicability

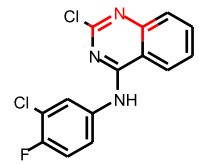
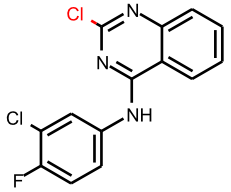
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

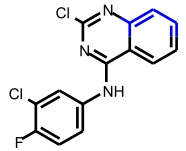
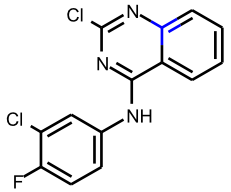
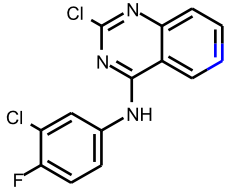
Feature Contribution

Top features for positive contribution

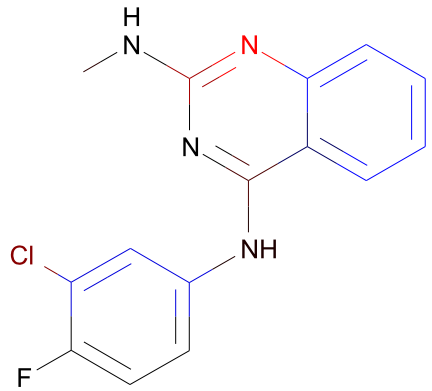
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	 [*]Cl	0.129

Top Features for negative contribution

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: 8.65

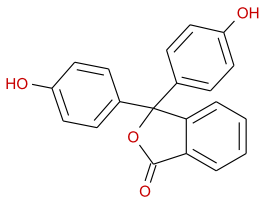
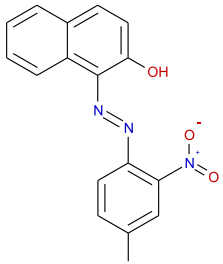
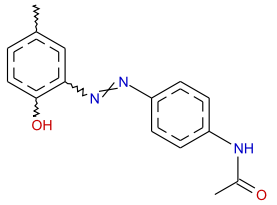
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 10.9

Mahalanobis Distance p-value: 0.00191

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

Structural Similar Compounds

Name	Phenolphthalein	646	44
Structure			
Actual Endpoint (-log C)	2.43468	0.937339	2.42163
Predicted Endpoint (-log C)	3.66084	3.26294	2.85113
Distance	0.646	0.696	0.698
Reference	CPDB	CPDB	CPDB

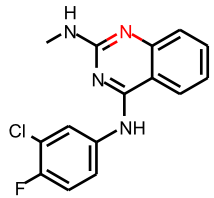
Model Applicability

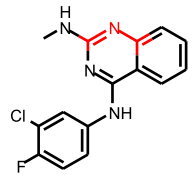
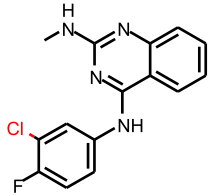
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

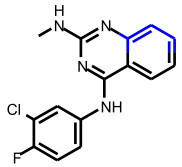
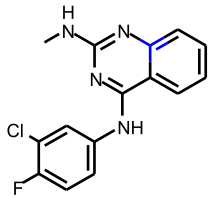
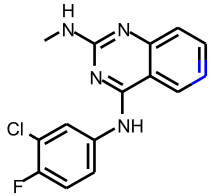
Feature Contribution

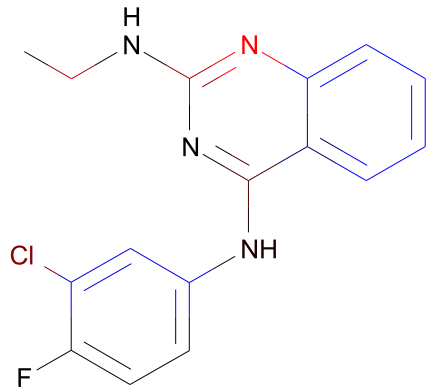
Top features for positive contribution

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

ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	 [*]Cl	0.129

Top Features for negative contribution

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



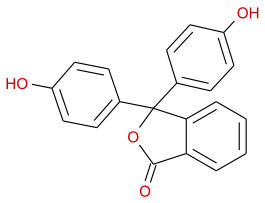
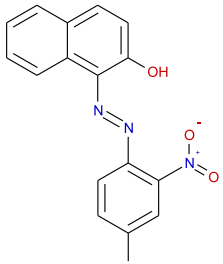
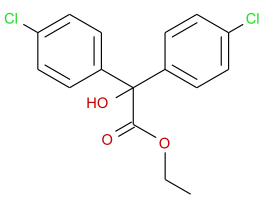
$C_{16}H_{14}ClFN_4$
Molecular Weight: 316.76056
ALogP: 5.025
Rotatable Bonds: 4
Acceptors: 4
Donors: 2

Model Prediction

Prediction: 5.92
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11
Mahalanobis Distance p-value: 0.00123

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

Structural Similar Compounds

Name	Phenolphthalein	646	Chlorobenzilate
Structure			
Actual Endpoint (-log C)	2.43468	0.937339	3.53947
Predicted Endpoint (-log C)	3.66084	3.26294	3.34564
Distance	0.672	0.700	0.704
Reference	CPDB	CPDB	CPDB

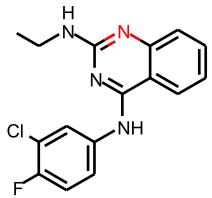
Model Applicability

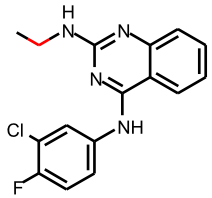
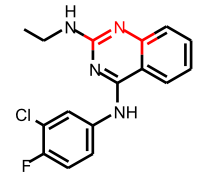
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

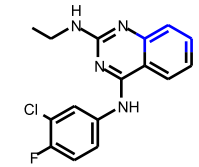
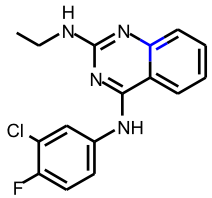
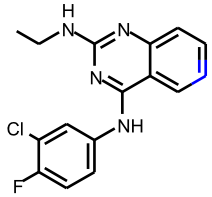
Feature Contribution

Top features for positive contribution

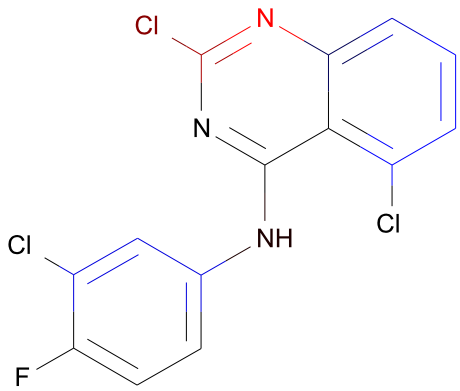
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	1559650422	 [*]C[*]	0.203
ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163

Top Features for negative contribution

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: 7.12

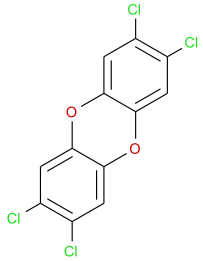
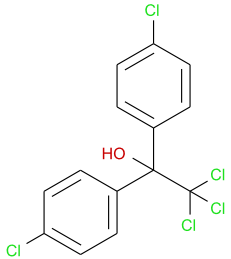
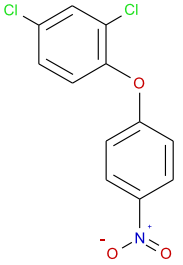
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 9.93

Mahalanobis Distance p-value: 0.0434

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	2,3,7,8-Tetrachlorodibenzo-p-dioxin	Dicofol	Nitrofen
Structure			
Actual Endpoint (-log C)	9.31469	4.05158	3.39277
Predicted Endpoint (-log C)	4.75869	3.80707	3.41579
Distance	0.702	0.715	0.735
Reference	CPDB	CPDB	CPDB

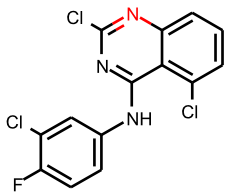
Model Applicability

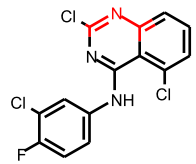
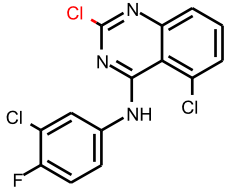
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

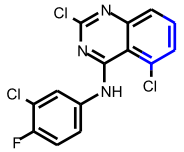
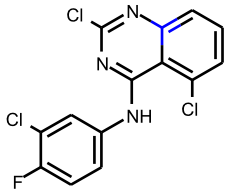
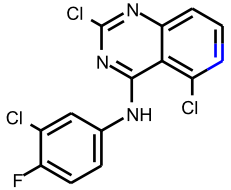
Feature Contribution

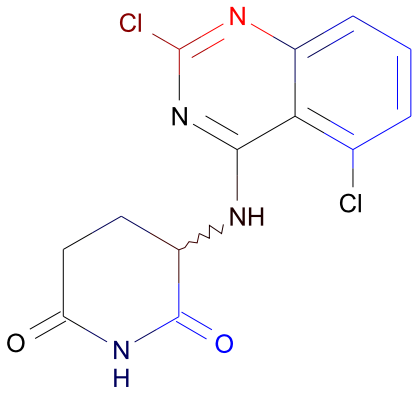
Top features for positive contribution

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

ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	 [*]Cl	0.129

Top Features for negative contribution

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



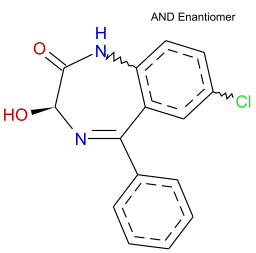
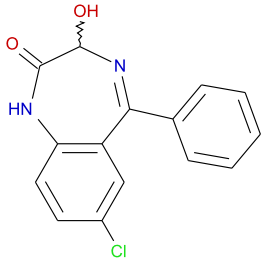
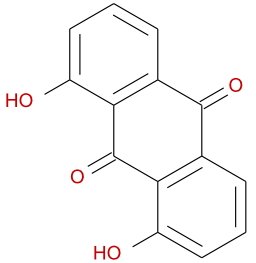
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 20.5
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.7
Mahalanobis Distance p-value: 5.83e-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	693	Oxazepam	Chrysazin
Structure			
Actual Endpoint (-log C)	3.90356	3.90356	3.0774
Predicted Endpoint (-log C)	3.39677	3.39677	3.07832
Distance	0.610	0.610	0.615
Reference	CPDB	CPDB	CPDB

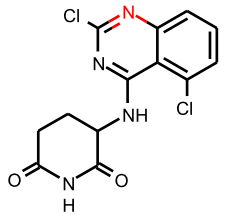
Model Applicability

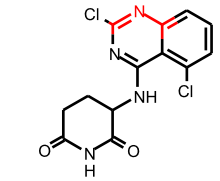
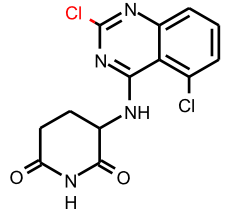
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

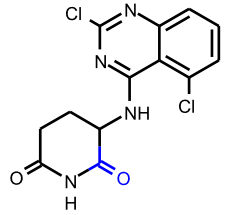
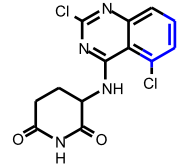
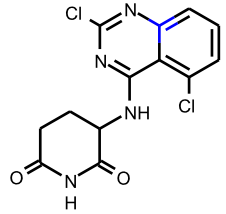
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1237219435: [*]C([*])N[c](:[*]):[*]
4. Unknown ECFP_2 feature: -2097159651: [*]CC(N[*])C(=[*])[*]

Feature Contribution

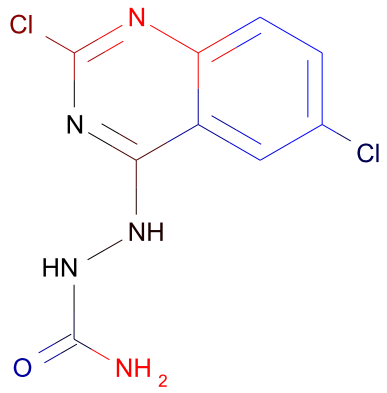
Top features for positive contribution

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

ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	 [*]Cl	0.129

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09078

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 10.1

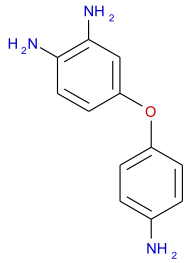
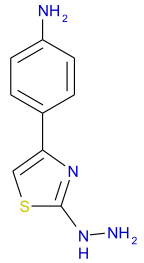
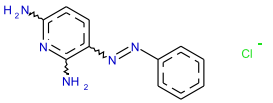
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 12.6

Mahalanobis Distance p-value: 1.32e-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	3,4,4´-Triaminodiphenyl ether	2-Hydrazino-4-(p-aminophenyl) thi-azole	215
Structure			
Actual Endpoint (-log C)	4.20909	4.26135	3.477
Predicted Endpoint (-log C)	3.19224	4.32504	3.97642
Distance	0.589	0.596	0.618
Reference	CPDB	CPDB	CPDB

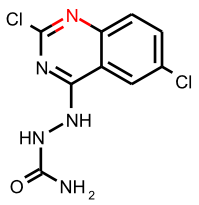
Model Applicability

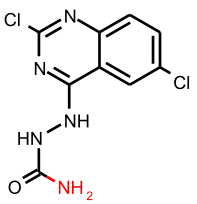
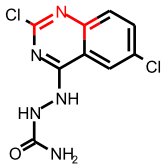
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC18 out of range. Value: -3.6042. Training min, max, SD, explained variance: -3.5308, 6.5936, 1.215, 0.0184.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

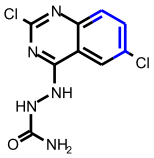
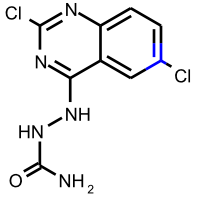
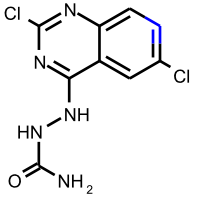
Feature Contribution

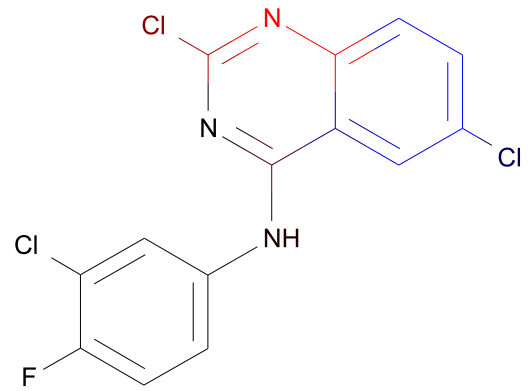
Top features for positive contribution

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

ECFP_6	1572579716	 [*]N	0.225
ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163

Top Features for negative contribution

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



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: 4.57

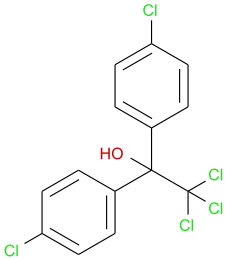
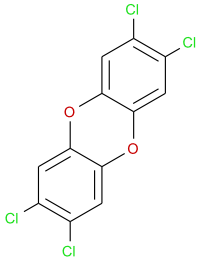
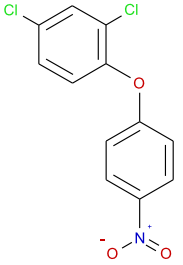
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.3

Mahalanobis Distance p-value: 0.00044

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	Dicofol	2,3,7,8-Tetrachlorodibenzo-p-dioxin	Nitrofen
Structure			
Actual Endpoint (-log C)	4.05158	9.31469	3.39277
Predicted Endpoint (-log C)	3.80707	4.75869	3.41579
Distance	0.700	0.702	0.716
Reference	CPDB	CPDB	CPDB

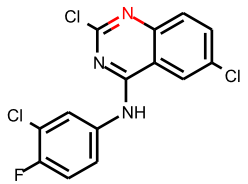
Model Applicability

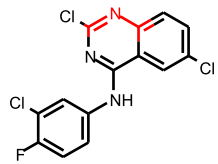
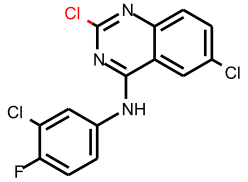
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC16 out of range. Value: 4.2781. Training min, max, SD, explained variance: -3.1026, 4.016, 1.245, 0.0193.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1311285389: [*][c](:[*]):[c](F):[cH]:[*]

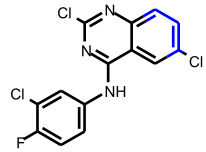
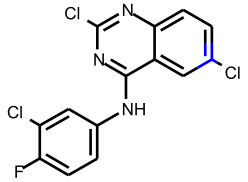
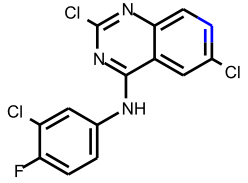
Feature Contribution

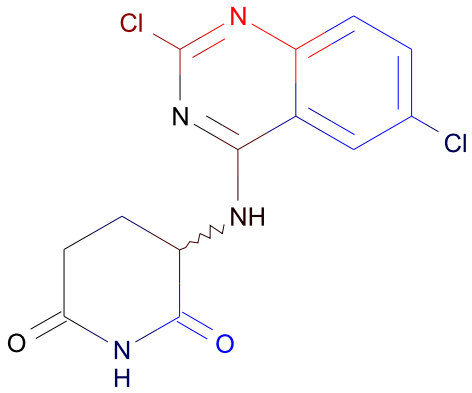
Top features for positive contribution

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

ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	 [*]Cl	0.129

Top Features for negative contribution

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



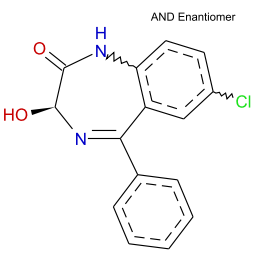
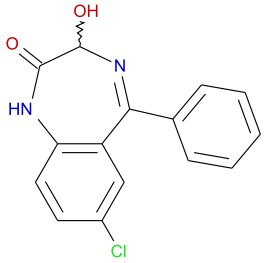
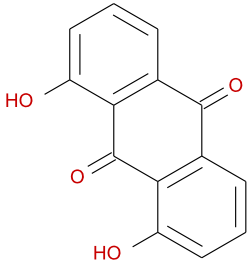
$C_{13}H_{10}Cl_2N_4O_2$
Molecular Weight: 325.15009
ALogP: 2.49
Rotatable Bonds: 2
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 21.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.6
Mahalanobis Distance p-value: 3.06e-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	693	Oxazepam	Chrysazin
Structure			
Actual Endpoint (-log C)	3.90356	3.90356	3.0774
Predicted Endpoint (-log C)	3.39677	3.39677	3.07832
Distance	0.592	0.592	0.628
Reference	CPDB	CPDB	CPDB

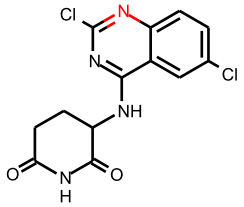
Model Applicability

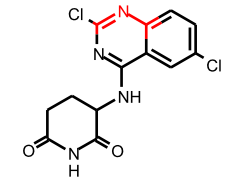
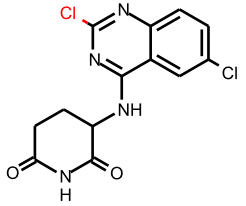
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC18 out of range. Value: -3.6033. Training min, max, SD, explained variance: -3.5308, 6.5936, 1.215, 0.0184.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: -1237219435: [*]C([*])N[c](:[*]):[*]
4. Unknown ECFP_2 feature: -2097159651: [*]CC(N[*])C(=[*])[*]

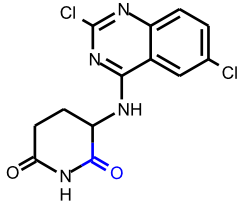
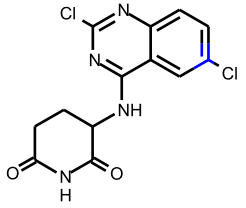
Feature Contribution

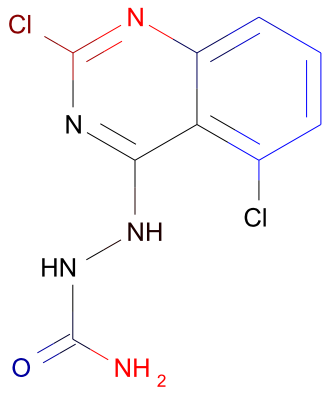
Top features for positive contribution

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

ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	 [*]Cl	0.129

Top Features for negative contribution

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



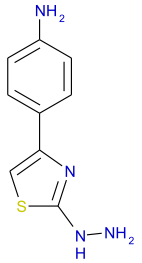
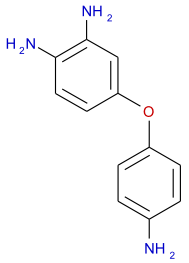
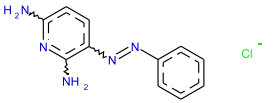
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: 9.53
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 10.8
Mahalanobis Distance p-value: 0.00281

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	2-Hydrazino-4-(p-aminophenyl) thi-azole	3,4,4´-Triaminodiphenyl ether	215
Structure			
Actual Endpoint (-log C)	4.26135	4.20909	3.477
Predicted Endpoint (-log C)	4.32504	3.19224	3.97642
Distance	0.596	0.596	0.611
Reference	CPDB	CPDB	CPDB

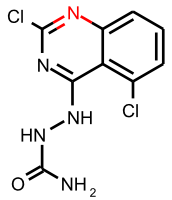
Model Applicability

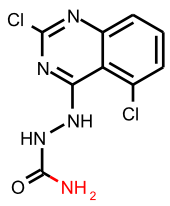
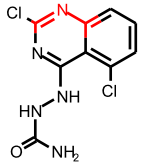
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]

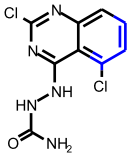
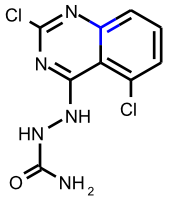
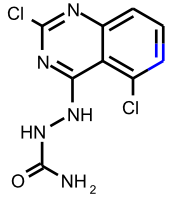
Feature Contribution

Top features for positive contribution

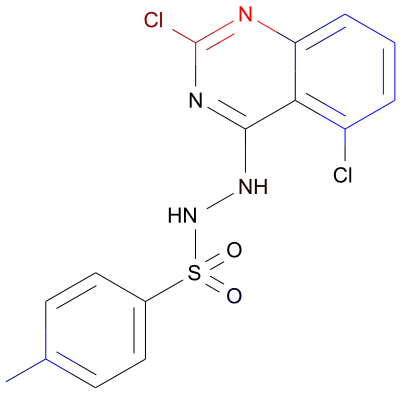
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	1572579716	 [*]N	0.225
ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163

Top Features for negative contribution

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{15}H_{12}Cl_2N_4O_2S$

Molecular Weight: 383.25237

ALogP: 4.443

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 8.91

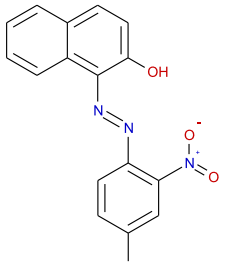
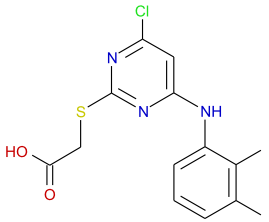
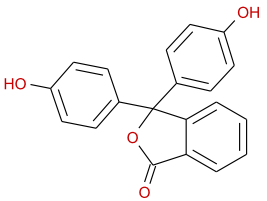
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 11.3

Mahalanobis Distance p-value: 0.000538

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

Structural Similar Compounds

Name	646	[4-Chloro-6-(2,3-xylidino)-2-pyri-midinylthio]acetic acid s	Phenolphthalein
Structure			
Actual Endpoint (-log C)	0.937339	4.47685	2.43468
Predicted Endpoint (-log C)	3.26294	3.8529	3.66084
Distance	0.627	0.662	0.678
Reference	CPDB	CPDB	CPDB

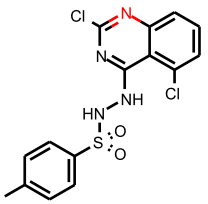
Model Applicability

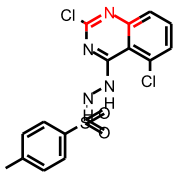
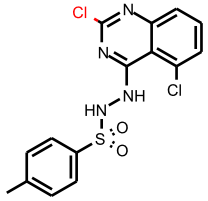
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

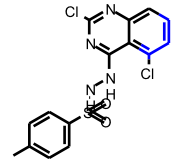
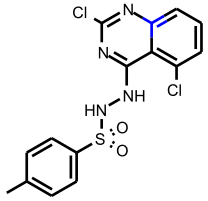
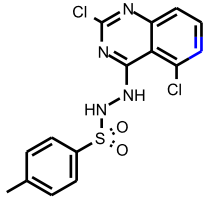
Feature Contribution

Top features for positive contribution

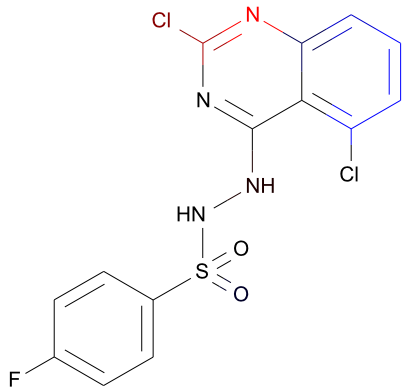
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	 [*]Cl	0.129

Top Features for negative contribution

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{14}H_9Cl_2FN_4O_2S$

Molecular Weight: 387.21626

ALogP: 4.162

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 6.89

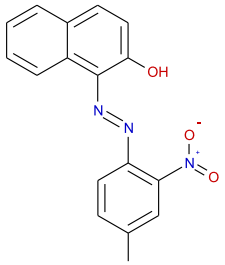
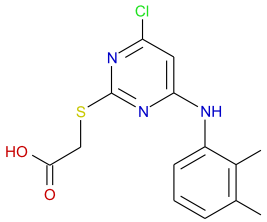
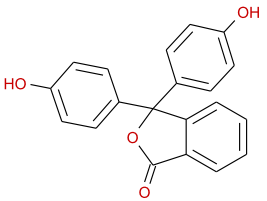
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 10.2

Mahalanobis Distance p-value: 0.0208

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

Structural Similar Compounds

Name	646	[4-Chloro-6-(2,3-xylidino)-2-pyri-midinylthio]acetic acid s	Phenolphthalein
Structure			
Actual Endpoint (-log C)	0.937339	4.47685	2.43468
Predicted Endpoint (-log C)	3.26294	3.8529	3.66084
Distance	0.656	0.676	0.677
Reference	CPDB	CPDB	CPDB

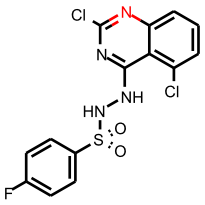
Model Applicability

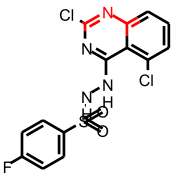
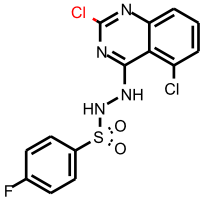
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: 1126732433: [*]:n:[c](Cl):n:[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

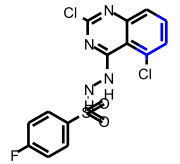
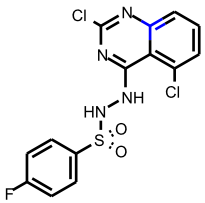
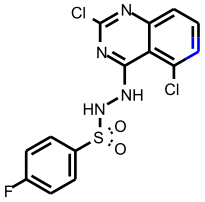
Feature Contribution

Top features for positive contribution

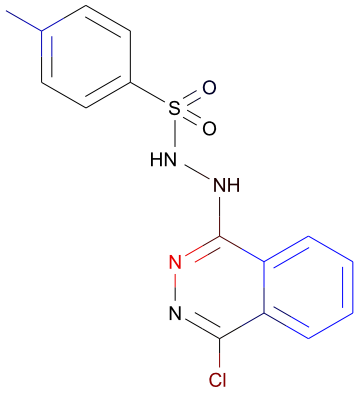
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	834876373	 [*][c](:[*]):n:[c](:[*]):[*]	0.163
ECFP_6	-817402818	 [*]Cl	0.129

Top Features for negative contribution

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



$C_{15}H_{13}ClN_4O_2S$

Molecular Weight: 348.80731

ALogP: 3.529

Rotatable Bonds: 4

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 21.2

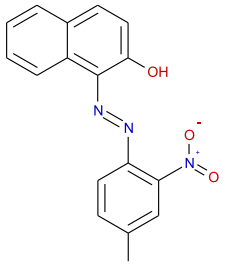
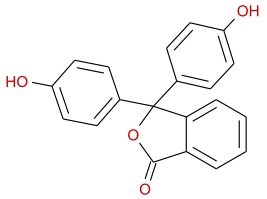
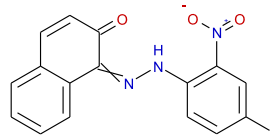
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 12.5

Mahalanobis Distance p-value: 2.15e-006

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

Structural Similar Compounds

Name	646	Phenolphthalein	C.I. pigment red 3
Structure			
Actual Endpoint (-log C)	0.937339	2.43468	0.937339
Predicted Endpoint (-log C)	3.26294	3.66084	3.17837
Distance	0.618	0.659	0.660
Reference	CPDB	CPDB	CPDB

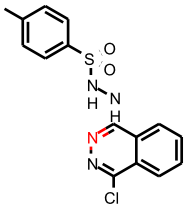
Model Applicability

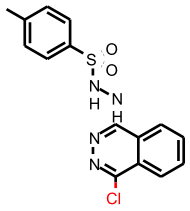
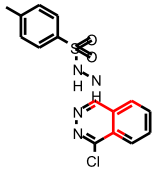
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC26 out of range. Value: -3.4447. Training min, max, SD, explained variance: -2.9667, 3.5042, 1.009, 0.0127.
2. Unknown ECFP_2 feature: 1050351974: [*]:n:[c](Cl):[c](:[*]):[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

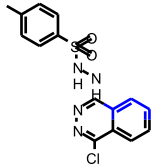
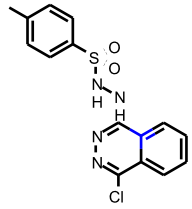
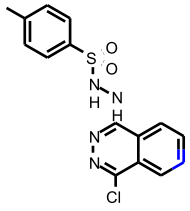
Feature Contribution

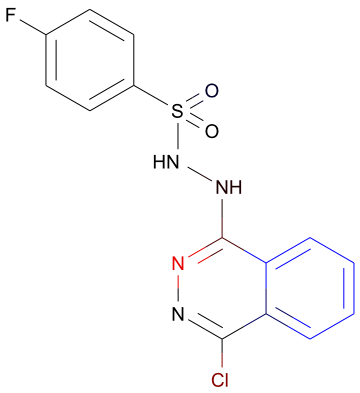
Top features for positive contribution

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

ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	1333660716	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.0746

Top Features for negative contribution

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



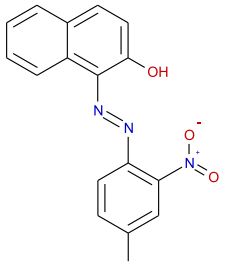
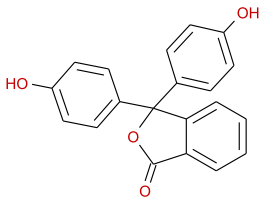
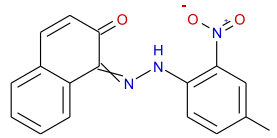
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 16.4
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.000448

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

Structural Similar Compounds

Name	646	Phenolphthalein	C.I. pigment red 3
Structure			
Actual Endpoint (-log C)	0.937339	2.43468	0.937339
Predicted Endpoint (-log C)	3.26294	3.66084	3.17837
Distance	0.655	0.664	0.686
Reference	CPDB	CPDB	CPDB

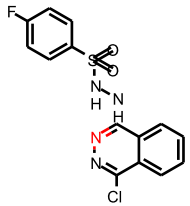
Model Applicability

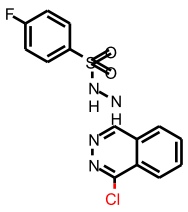
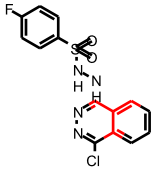
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC26 out of range. Value: -3.2042. Training min, max, SD, explained variance: -2.9667, 3.5042, 1.009, 0.0127.
2. Unknown ECFP_2 feature: 1050351974: [*]:n:[c](Cl):[c](:[*]):[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

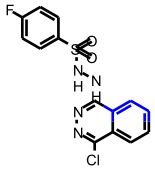
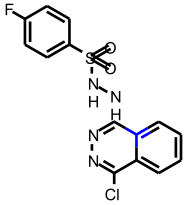
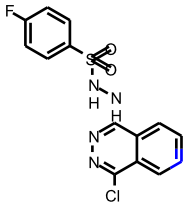
Feature Contribution

Top features for positive contribution

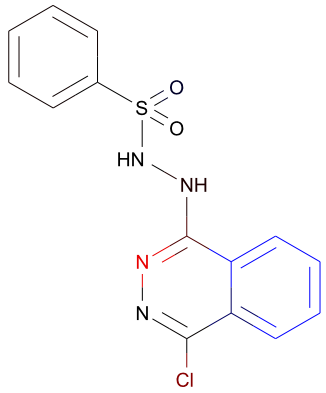
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	1333660716	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.0746

Top Features for negative contribution

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

TOPKAT_Carcinogenic_Potency_TD50_Mouse



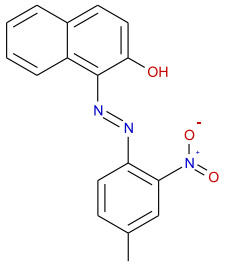
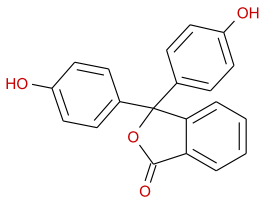
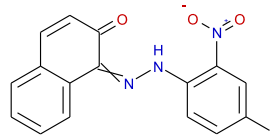
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 17.2
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 12.3
Mahalanobis Distance p-value: 6.1e-006

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

Structural Similar Compounds

Name	646	Phenolphthalein	C.I. pigment red 3
Structure			
Actual Endpoint (-log C)	0.937339	2.43468	0.937339
Predicted Endpoint (-log C)	3.26294	3.66084	3.17837
Distance	0.659	0.664	0.684
Reference	CPDB	CPDB	CPDB

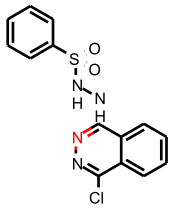
Model Applicability

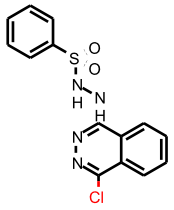
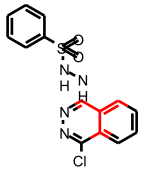
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC26 out of range. Value: -2.9692. Training min, max, SD, explained variance: -2.9667, 3.5042, 1.009, 0.0127.
2. Unknown ECFP_2 feature: 1050351974: [*]:n:[c](Cl):[c](:[*]):[*]
3. Unknown ECFP_2 feature: 944235823: [*]NNS(=[*])(=[*])[*]

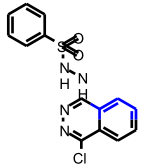
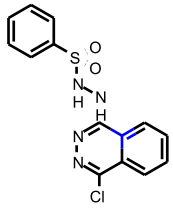
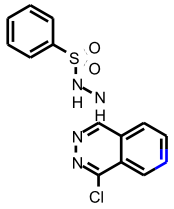
Feature Contribution

Top features for positive contribution

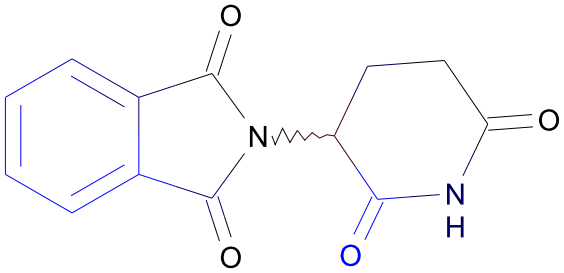
Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	655739385	 [*]:n:[*]	0.229

ECFP_6	-817402818	 [*]Cl	0.129
ECFP_6	1333660716	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.0746

Top Features for negative contribution

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

Thalidomide



$C_{13}H_{10}N_2O_4$
Molecular Weight: 258.22949
ALogP: 9.7e-002
Rotatable Bonds: 1
Acceptors: 4
Donors: 1

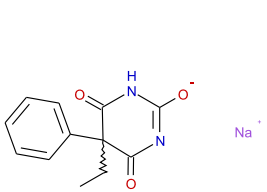
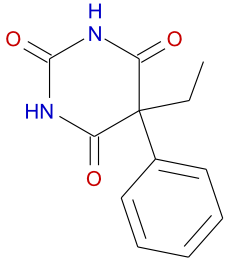
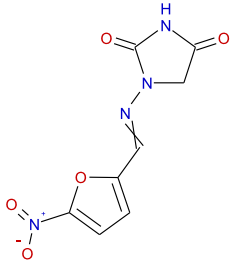
Model Prediction

Prediction: 367
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 9.93
Mahalanobis Distance p-value: 0.043

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	581	Phenobarbital s	1-[(5-Nitrofurfuryli-dene)ami-no]hy-dantoin
Structure			
Actual Endpoint (-log C)	3.89128	4.49846	2.23074
Predicted Endpoint (-log C)	3.31105	3.14828	3.18468
Distance	0.590	0.638	0.660
Reference	CPDB	CPDB	CPDB

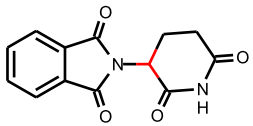
Model Applicability

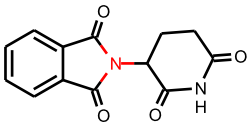
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

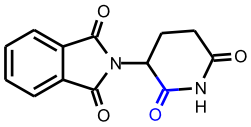
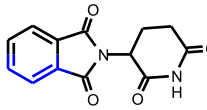
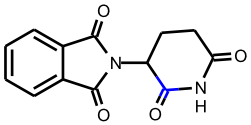
1. All properties and OPS components are within expected ranges.
2. Unknown ECFP_2 feature: -957084426: [*]C([*])N1C(=[*])[*]:[*]C1=[*]

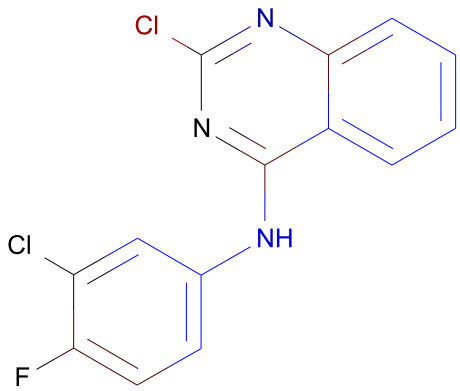
Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
ECFP_6	-167460056	 [*]C([*])[*]	0.0596

ECFP_6	670515721	 [*]N([*])[*]	0.00735

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



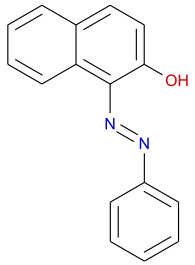
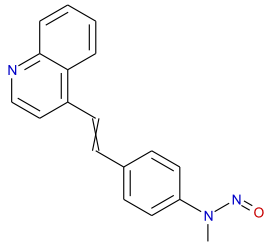
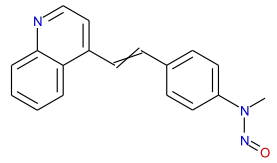
$C_{14}H_8Cl_2FN_3$
Molecular Weight: 308.13782
ALogP: 5.202
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 18.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.7
Mahalanobis Distance p-value: 0.00147

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	45	357	4-(4-N-Methyl-N-nitrosamino-styryl)quinoline
Structure			
Actual Endpoint (-log C)	3.92659	5.61692	5.61692
Predicted Endpoint (-log C)	3.28325	5.49128	5.49128
Distance	0.530	0.601	0.601
Reference	CPDB	CPDB	CPDB

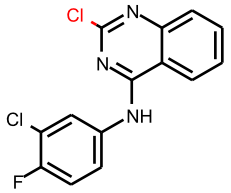
Model Applicability

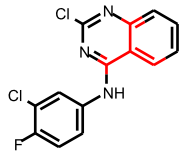
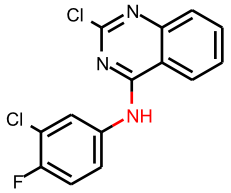
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

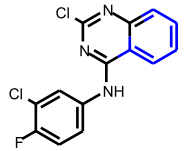
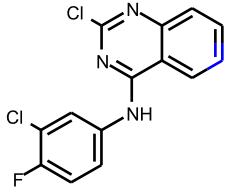
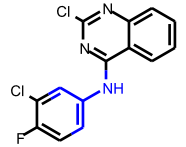
Feature Contribution

Top features for positive contribution

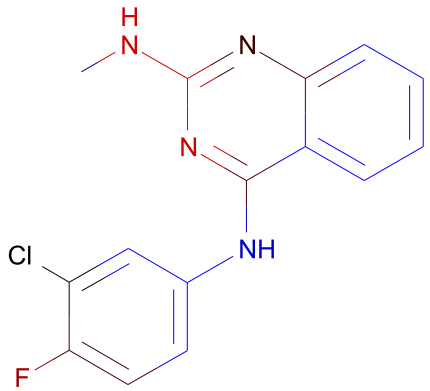
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	32	 [*]Cl	0.154

FCFP_6	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121
FCFP_6	3	 [*]N[*]	0.064

Top Features for negative contribution

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

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{15}H_{12}ClFN_4$

Molecular Weight: 302.73398

ALogP: 4.677

Rotatable Bonds: 3

Acceptors: 4

Donors: 2

Model Prediction

Prediction: 12

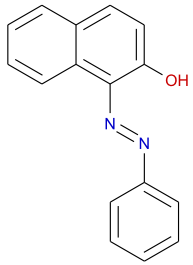
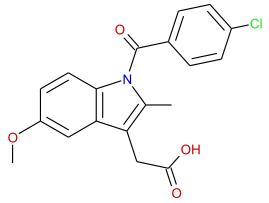
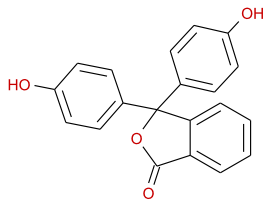
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13.9

Mahalanobis Distance p-value: 4.08e-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	45	Indomethacin	Phenolphthalein
Structure			
Actual Endpoint (-log C)	3.92659	5.49293	2.54766
Predicted Endpoint (-log C)	3.28325	4.9569	3.7508
Distance	0.618	0.629	0.630
Reference	CPDB	CPDB	CPDB

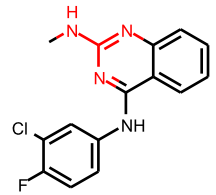
Model Applicability

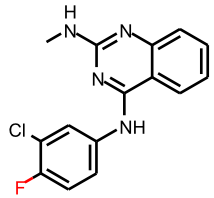
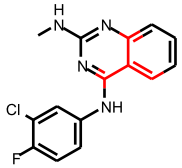
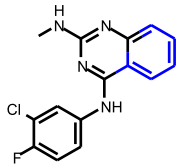
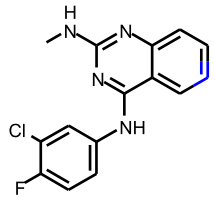
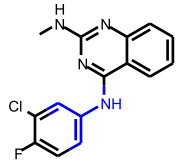
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

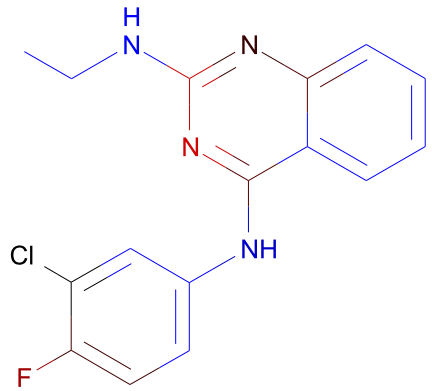
1. OPS PC18 out of range. Value: 4.9898. Training min, max, SD, explained variance: -4.1023, 4.8669, 1.305, 0.0164.

Feature Contribution

Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1151914249	 [*]N[c](:n:[*]):n:[*]	0.204

FCFP_6	32	 [*]Cl	0.154
FCFP_6	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121
Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.323



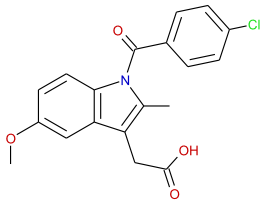
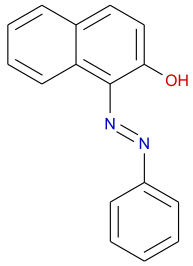
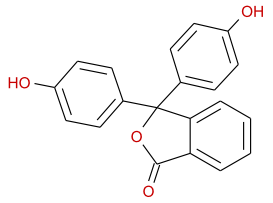
$C_{16}H_{14}ClFN_4$
Molecular Weight: 316.76056
ALogP: 5.025
Rotatable Bonds: 4
Acceptors: 4
Donors: 2

Model Prediction

Prediction: 35.7
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.1
Mahalanobis Distance p-value: 9.65e-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	Indomethacin	45	Phenolphthalein
Structure			
Actual Endpoint (-log C)	5.49293	3.92659	2.54766
Predicted Endpoint (-log C)	4.9569	3.28325	3.7508
Distance	0.623	0.644	0.650
Reference	CPDB	CPDB	CPDB

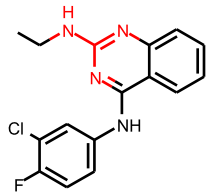
Model Applicability

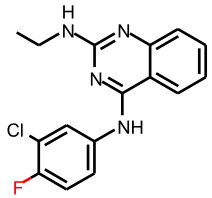
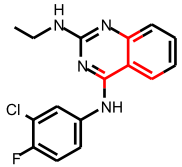
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC18 out of range. Value: 5.121. Training min, max, SD, explained variance: -4.1023, 4.8669, 1.305, 0.0164.

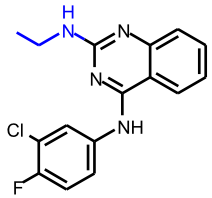
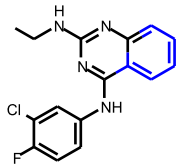
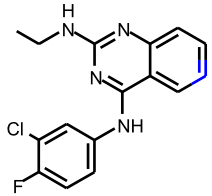
Feature Contribution

Top features for positive contribution

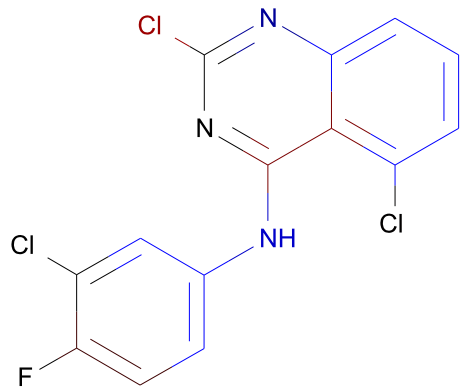
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1151914249	 [*]N[c](:n:[*]):n:[*]	0.204

FCFP_6	32	 [*]Cl	0.154
FCFP_6	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1272709286	 [*]NCC	-0.526
FCFP_6	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*]:[cH]:[*]	-0.354

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{14}H_7Cl_3FN_3$

Molecular Weight: 342.58288

ALogP: 5.867

Rotatable Bonds: 2

Acceptors: 3

Donors: 1

Model Prediction

Prediction: 16

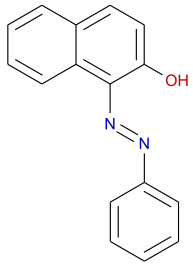
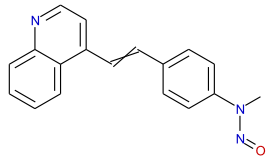
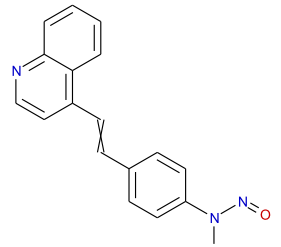
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 14

Mahalanobis Distance p-value: 1.83e-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	45	4-(4-N-Methyl-N-nitrosamino-styryl)quinoline	357
Structure			
Actual Endpoint (-log C)	3.92659	5.61692	5.61692
Predicted Endpoint (-log C)	3.28325	5.49128	5.49128
Distance	0.596	0.651	0.651
Reference	CPDB	CPDB	CPDB

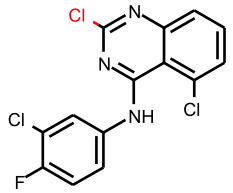
Model Applicability

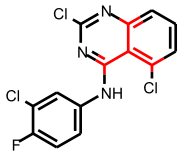
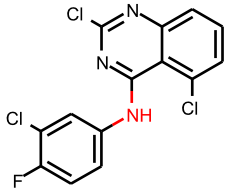
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

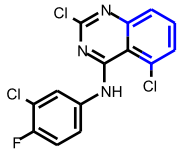
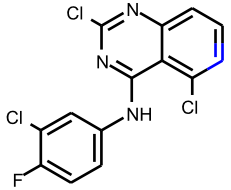
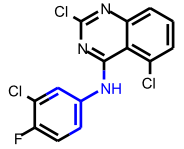
Feature Contribution

Top features for positive contribution

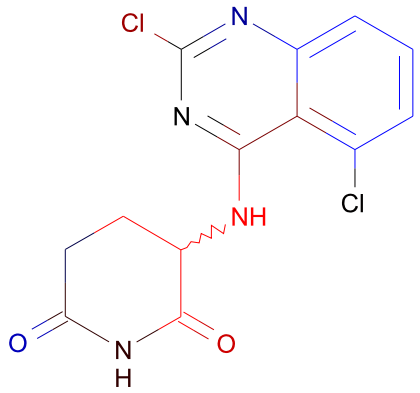
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	32	 [*]Cl	0.154

FCFP_6	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121
FCFP_6	3	 [*]N[*]	0.064

Top Features for negative contribution

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

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 0.679

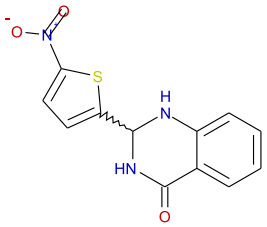
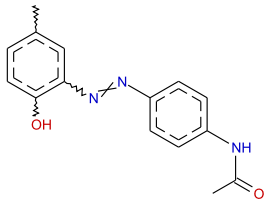
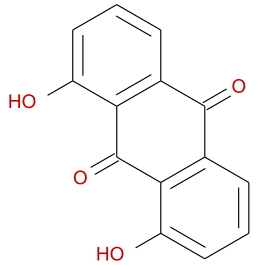
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 15.3

Mahalanobis Distance p-value: 4.62e-012

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

Structural Similar Compounds

Name	1,2-Dihydro-2-(5-nitro-2-thi-enyl) quinazolin-4(3H)-one	44	Chrysazin
Structure			
Actual Endpoint (-log C)	5.25509	2.85045	2.99143
Predicted Endpoint (-log C)	3.89291	2.7768	3.29868
Distance	0.593	0.616	0.627
Reference	CPDB	CPDB	CPDB

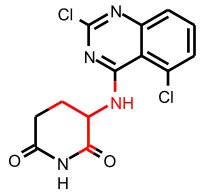
Model Applicability

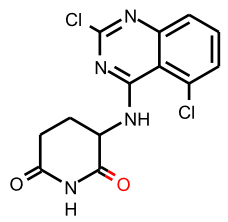
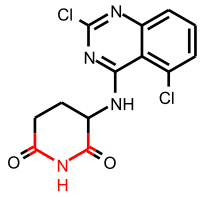
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC18 out of range. Value: 4.9912. Training min, max, SD, explained variance: -4.1023, 4.8669, 1.305, 0.0164.

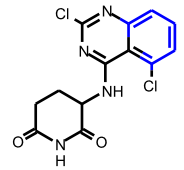
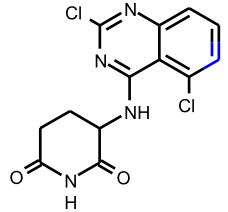
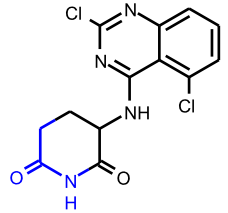
Feature Contribution

Top features for positive contribution

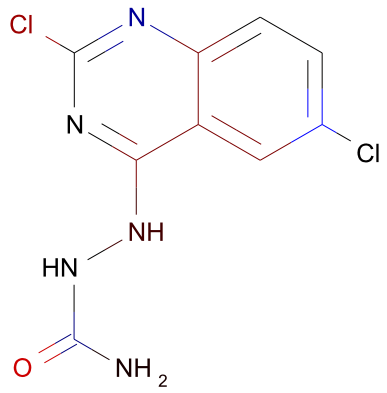
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1043250487	 [*]CC(N[*])C(=[*])[*]	1.15

FCFP_6	1	 [*]=O	0.234
FCFP_6	-885550502	 [*]C(=[*])NC(=[*])[*]	0.229

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	566058135	 [*]CC(=O)N[*]	-0.182

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_9H_7Cl_2N_5O$

Molecular Weight: 272.09078

ALogP: 2.379

Rotatable Bonds: 2

Acceptors: 4

Donors: 3

Model Prediction

Prediction: 11.2

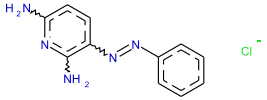
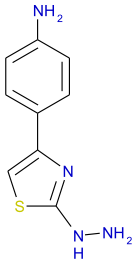
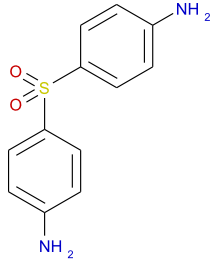
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 13.3

Mahalanobis Distance p-value: 8.94e-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	215	2-Hydrazino-4-(p-aminophenyl) thi-azole	Dapsone
Structure			
Actual Endpoint (-log C)	2.84742	5.30159	4.04473
Predicted Endpoint (-log C)	3.32496	4.70335	4.05717
Distance	0.598	0.615	0.621
Reference	CPDB	CPDB	CPDB

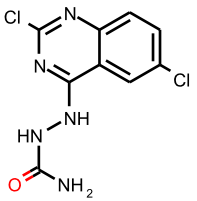
Model Applicability

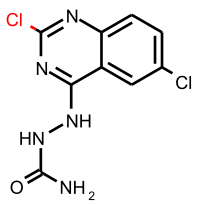
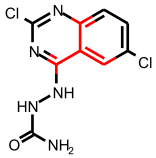
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

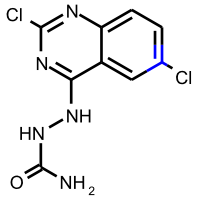
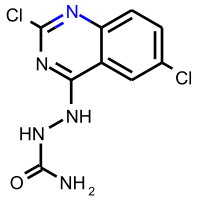
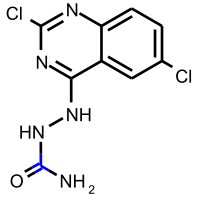
Feature Contribution

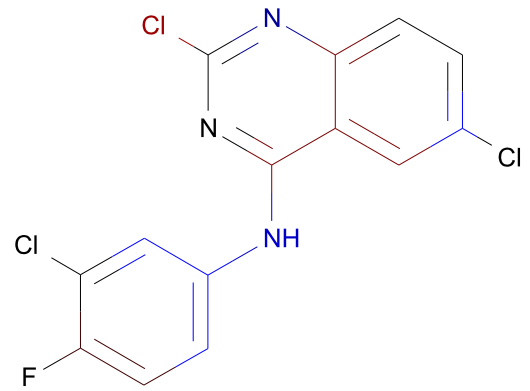
Top features for positive contribution

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

FCFP_6	32	 [*]Cl	0.154
FCFP_6	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	17	 [*]:n:[*]	-0.149
FCFP_6	0	 [*]C	-0.115



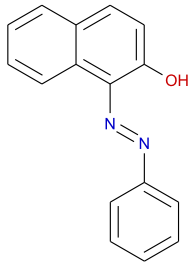
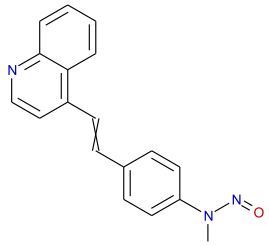
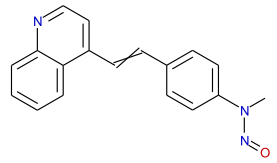
$C_{14}H_7Cl_3FN_3$
Molecular Weight: 342.58288
ALogP: 5.867
Rotatable Bonds: 2
Acceptors: 3
Donors: 1

Model Prediction

Prediction: 6.05
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14.2
Mahalanobis Distance p-value: 4.74e-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	45	357	4-(4-N-Methyl-N-nitrosamino-styryl)quinoline
Structure			
Actual Endpoint (-log C)	3.92659	5.61692	5.61692
Predicted Endpoint (-log C)	3.28325	5.49128	5.49128
Distance	0.603	0.664	0.664
Reference	CPDB	CPDB	CPDB

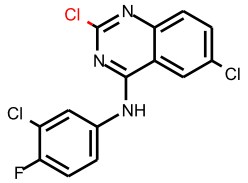
Model Applicability

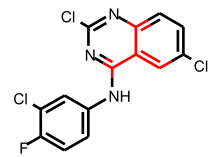
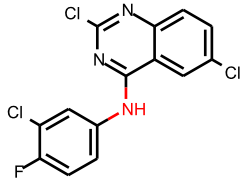
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

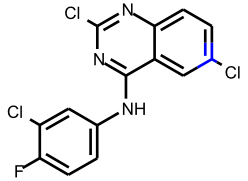
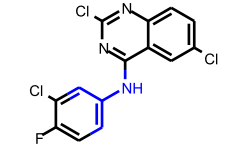
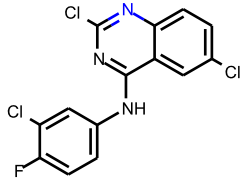
1. All properties and OPS components are within expected ranges.

Feature Contribution

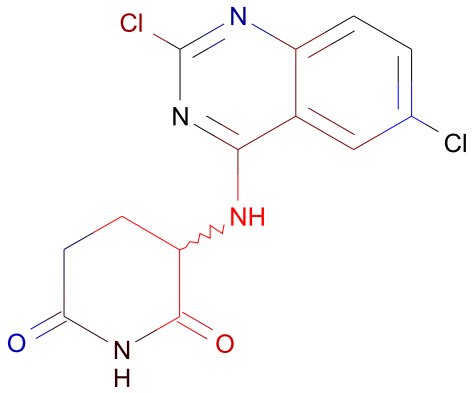
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	32	 [*]Cl	0.154

FCFP_6	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121
FCFP_6	3	 [*]N[*]	0.064

Top Features for negative contribution			
Fingerprint	**Bit/Smiles**	**Feature Structure**	**Score**
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	590925877	 [*]N[c](:[cH]:[*]):[cH]:[*]	-0.323
FCFP_6	17	 [*]:n:[*]	-0.149

TOPKAT_Carcinogenic_Potency_TD50_Rat



$C_{13}H_{10}Cl_2N_4O_2$

Molecular Weight: 325.15009

ALogP: 2.49

Rotatable Bonds: 2

Acceptors: 5

Donors: 2

Model Prediction

Prediction: 0.257

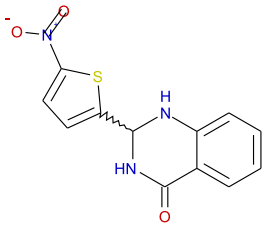
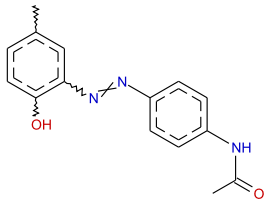
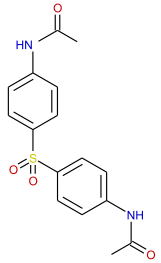
Unit: mg/kg_body_weight/day

Mahalanobis Distance: 15.1

Mahalanobis Distance p-value: 2.07e-011

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

Structural Similar Compounds

Name	1,2-Dihydro-2-(5-nitro-2-thi-enyl) quinazolin-4(3H)-one	44	4,4´-Sulfonylbisacetanilide
Structure			
Actual Endpoint (-log C)	5.25509	2.85045	3.77655
Predicted Endpoint (-log C)	3.89291	2.7768	3.55337
Distance	0.599	0.616	0.628
Reference	CPDB	CPDB	CPDB

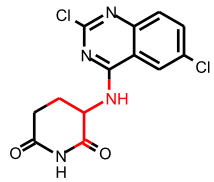
Model Applicability

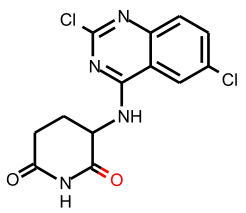
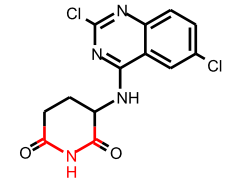
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC18 out of range. Value: 4.9468. Training min, max, SD, explained variance: -4.1023, 4.8669, 1.305, 0.0164.

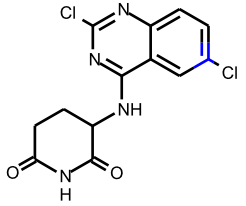
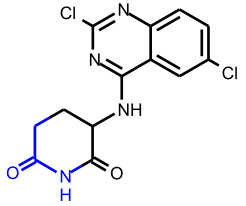
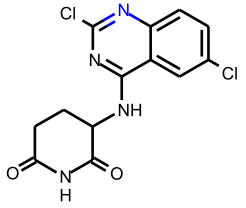
Feature Contribution

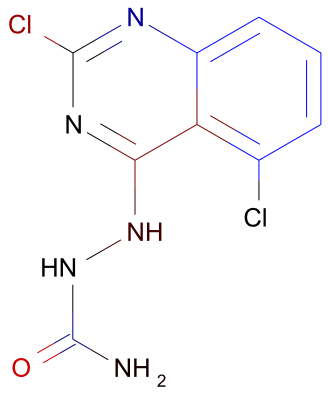
Top features for positive contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	-1043250487	 [*]CC(N[*])C(=[*])[*]	1.15

FCFP_6	1	 [*]=O	0.234
FCFP_6	-885550502	 [*]C(=[*])NC(=[*])[*]	0.229

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	566058135	 [*]CC(=O)N[*]	-0.182
FCFP_6	17	 [*]:n:[*]	-0.149



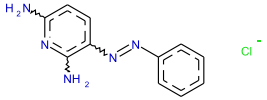
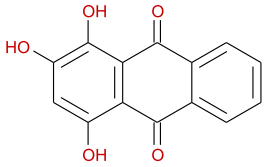
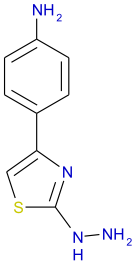
$C_9H_7Cl_2N_5O$
Molecular Weight: 272.09077
ALogP: 2.379
Rotatable Bonds: 2
Acceptors: 4
Donors: 3

Model Prediction

Prediction: 29.6
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.6
Mahalanobis Distance p-value: 2.08e-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	215	Purpurin	2-Hydrazino-4-(p-aminophenyl) thi-azole
Structure			
Actual Endpoint (-log C)	2.84742	2.57737	5.30159
Predicted Endpoint (-log C)	3.32496	3.49183	4.70335
Distance	0.590	0.613	0.615
Reference	CPDB	CPDB	CPDB

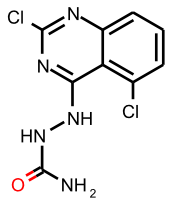
Model Applicability

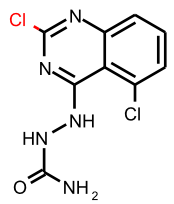
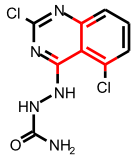
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

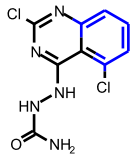
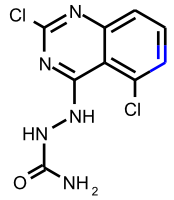
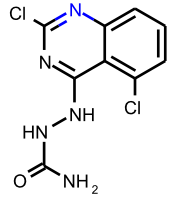
Feature Contribution

Top features for positive contribution

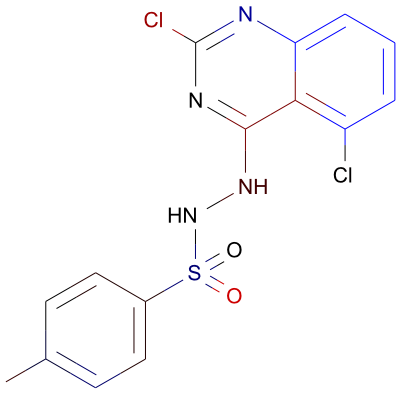
Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	1	 [*]=O	0.234

FCFP_6	32	 [*]Cl	0.154
FCFP_6	307419094	 [*][c](:[*]):[c](:[cH]:[*]):[c](:[*]):[*]	0.121

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	17	 [*]:n:[*]	-0.149

TOPKAT_Carcinogenic_Potency_TD50_Rat



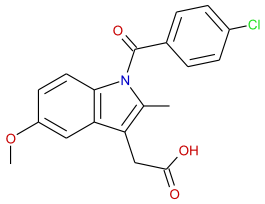
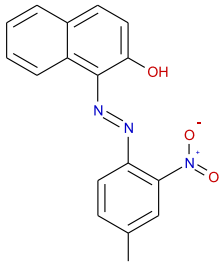
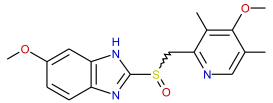
$C_{15}H_{12}Cl_2N_4O_2S$
Molecular Weight: 383.25237
ALogP: 4.443
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 10.2
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14
Mahalanobis Distance p-value: 1.55e-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	Indomethacin	646	Omeprazole
Structure			
Actual Endpoint (-log C)	5.49293	2.41938	3.4628
Predicted Endpoint (-log C)	4.9569	3.77987	4.7324
Distance	0.568	0.603	0.609
Reference	CPDB	CPDB	CPDB

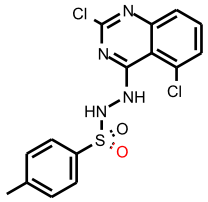
Model Applicability

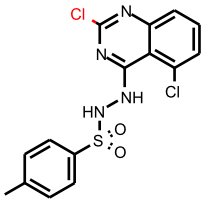
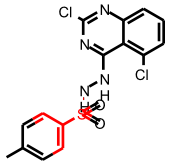
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

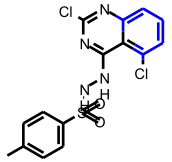
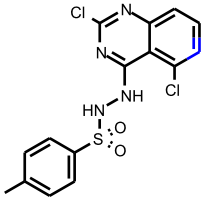
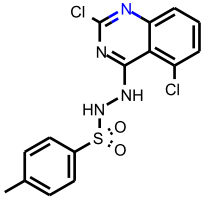
Feature Contribution

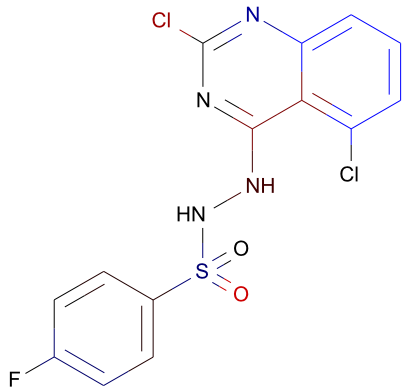
Top features for positive contribution

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

FCFP_6	32	 [*]Cl	0.154
FCFP_6	203677720	 [*]S(=[*])(=[*])[c](:[cH]:[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	17	 [*]:n:[*]	-0.149



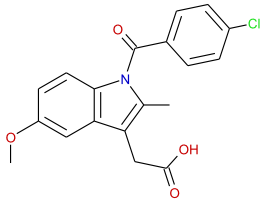
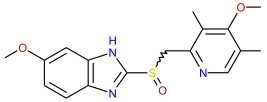
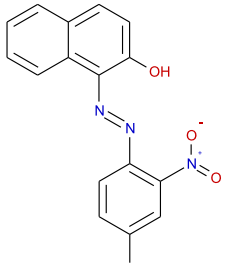
$C_{14}H_9Cl_2FN_4O_2S$
Molecular Weight: 387.21626
ALogP: 4.162
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 12.2
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 13.5
Mahalanobis Distance p-value: 2.98e-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	Indomethacin	Omeprazole	646
Structure			
Actual Endpoint (-log C)	5.49293	3.4628	2.41938
Predicted Endpoint (-log C)	4.9569	4.7324	3.77987
Distance	0.554	0.605	0.623
Reference	CPDB	CPDB	CPDB

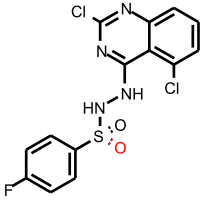
Model Applicability

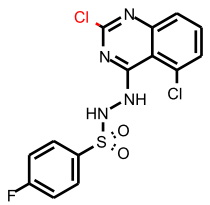
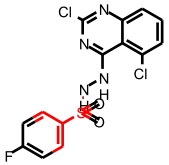
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

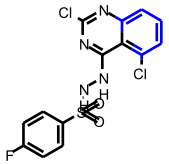
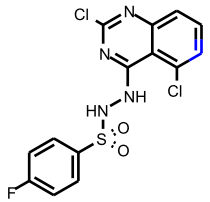
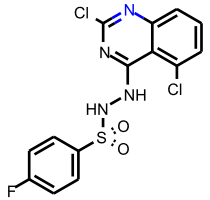
Feature Contribution

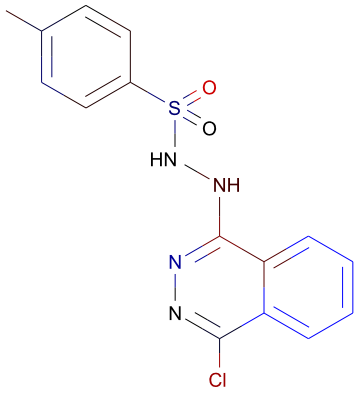
Top features for positive contribution

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

FCFP_6	32	 [*]Cl	0.154
FCFP_6	203677720	 [*]S(=[*])(=[*])[c](:[cH]:[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	17	 [*]:n:[*]	-0.149



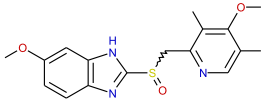
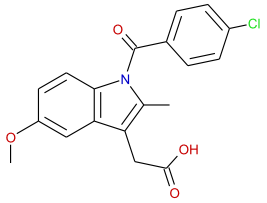
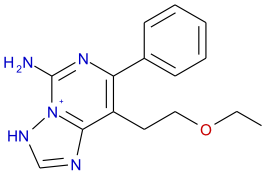
$C_{15}H_{13}ClN_4O_2S$
Molecular Weight: 348.80731
ALogP: 3.529
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 12.7
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.9
Mahalanobis Distance p-value: 0.000885

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	Omeprazole	Indomethacin	796
Structure			
Actual Endpoint (-log C)	3.4628	5.49293	2.71505
Predicted Endpoint (-log C)	4.7324	4.9569	4.45918
Distance	0.592	0.595	0.603
Reference	CPDB	CPDB	CPDB

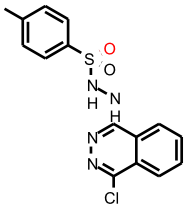
Model Applicability

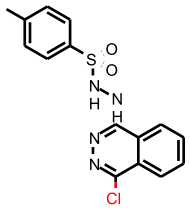
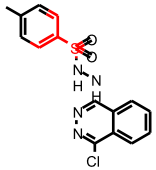
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

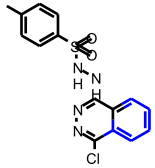
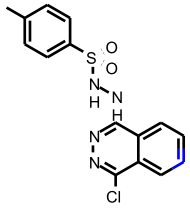
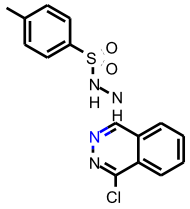
Feature Contribution

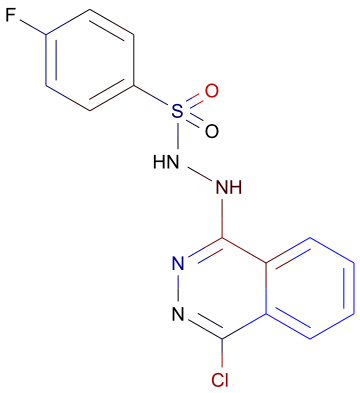
Top features for positive contribution

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

FCFP_6	32	 [*]Cl	0.154
FCFP_6	203677720	 [*]S(=[*])(=[*])[c](:[cH]:[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	17	 [*]:n:[*]	-0.149



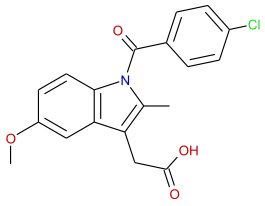
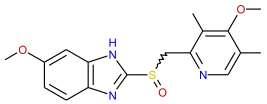
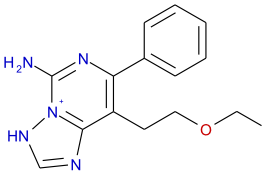
$C_{14}H_{10}ClFN_4O_2S$
Molecular Weight: 352.7712
ALogP: 3.248
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 15.2
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 11.1
Mahalanobis Distance p-value: 0.0125

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	Indomethacin	Omeprazole	796
Structure			
Actual Endpoint (-log C)	5.49293	3.4628	2.71505
Predicted Endpoint (-log C)	4.9569	4.7324	4.45918
Distance	0.576	0.593	0.593
Reference	CPDB	CPDB	CPDB

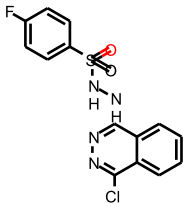
Model Applicability

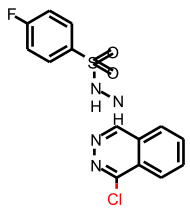
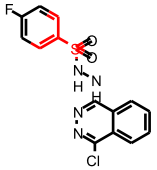
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

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

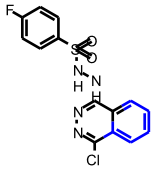
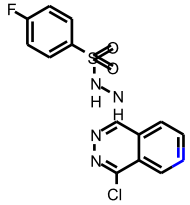
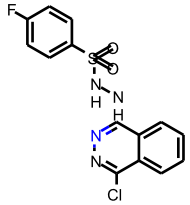
Feature Contribution

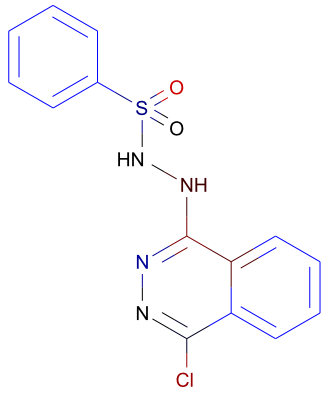
Top features for positive contribution

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

FCFP_6	32	 [*]Cl	0.154
FCFP_6	203677720	 [*]S(=[*])(=[*])[c](:[cH]:[*]):[cH]:[*]	0.137

Top Features for negative contribution

Fingerprint	Bit/Smiles	Feature Structure	Score
FCFP_6	991735244	 [*]:[c]1:[*]:[cH]:[cH]:[cH]:[cH]:1	-0.422
FCFP_6	16	 [*]:[cH]:[*]	-0.354
FCFP_6	17	 [*]:n:[*]	-0.149



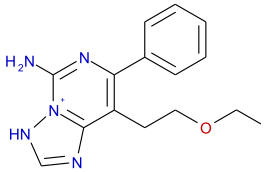
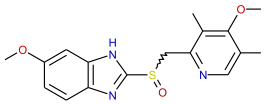
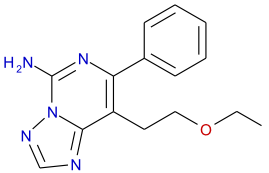
$C_{14}H_{11}ClN_4O_2S$
Molecular Weight: 334.78073
ALogP: 3.043
Rotatable Bonds: 4
Acceptors: 5
Donors: 2

Model Prediction

Prediction: 38.1
Unit: mg/kg_body_weight/day
Mahalanobis Distance: 14
Mahalanobis Distance p-value: 2.75e-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	796	Omeprazole	Bemitradine
Structure			
Actual Endpoint (-log C)	2.71505	3.4628	2.71351
Predicted Endpoint (-log C)	4.45918	4.7324	4.65043
Distance	0.566	0.594	0.599
Reference	CPDB	CPDB	CPDB

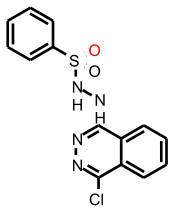
Model Applicability

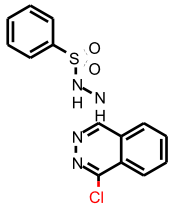
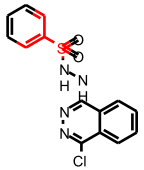
Unknown features are fingerprint features in the query molecule, but not found or appearing too infrequently in the training set.

1. OPS PC11 out of range. Value: 5.1391. Training min, max, SD, explained variance: -5.9328, 4.6461, 1.629, 0.0255.

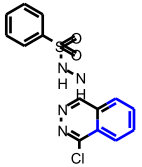
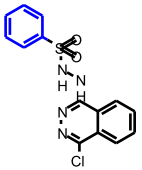
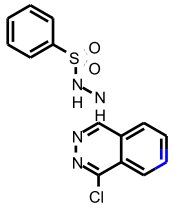
Feature Contribution

Top features for positive contribution

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

FCFP_6	32	 [*]Cl	0.154
FCFP_6	203677720	 [*]S(=[*])(=[*])[c](:[cH]:[*]):[cH]:[*]	0.137

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	-2093839777	 [*][c]1:[cH]:[cH]:[cH]:[cH]:[cH]:1	-0.378
FCFP_6	16	 [*]:[cH]:[*]	-0.354