Support Information for:

# eMolTox: prediction of molecular toxicity with confidence

Changge Ji[1,2,*], Fredrik Svensson[2,3], Azedine Zoufir[2] and Andreas Bender[2,*]

[1]Shanghai Engineering Research Center for Molecular Therapeutics and New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062 China, [2]Center for Molecular Informatics, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, UK and [3]IOTA Pharmaceuticals, St Johns Innovation Centre, Cowley Road, Cambridge CB4 0WS, UK.

**Text S1.** Conformal Prediction

Conformal prediction evaluates the similarity between the new samples and the training data. The output represents the probability that the new molecule is either active, inactive or uncertain (in the case of the query molecule being outside the applicability domain of the model), given a user-defined significance level that sets the maximum allowable fraction of erroneous predictions. For the classification task, the problem of conformal prediction is how to estimate the confidence of predicting the class label. This is done by finding how strange (nonconformal) a new example is in comparison to the training set, by calculating the nonconformity measure. This is done by the following several steps:

First, we divide the training set $Z$ into two disjoint subsets: a proper training set $Z_t$ and a calibration set $Z_c$.

Second, fit a model $h$ using the training set $Z_t$. The underlying model is RandomForest.

Third, calculate the nonconformity score for each sample in the calibration set $Z_c$. To construct a conformal predictor's prediction, we need to define a nonconformity measure. The nonconformity function, we use is

$$\alpha(i) = 1-P_h(y_i|x_i),$$

where the $(y_i|x_i)$ is the predicted probability of a sample classified as a given class using the model $h$ we constructed in the second step.

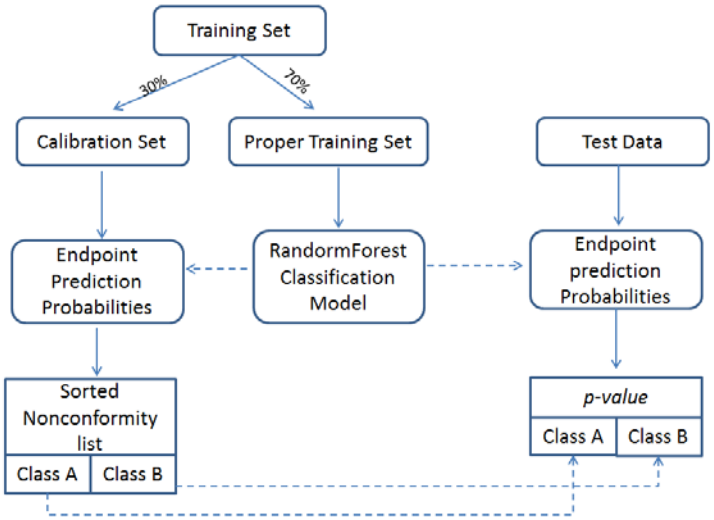
Fourth, to assess how different a new example $x_n$ is from old examples, we need to compare nonconformity $\alpha_n$ to the nonconformity scores $\alpha_j$ of previous examples $x_j$, $j=1,\ldots,k$ in the calibration sets. To make this comparison, we compute the fraction:

$$|\{j = 1, \ldots, k: \alpha_j \geq \alpha_n\}|/n$$

i.e. the number of $\alpha_j$ that are as large or larger than $\alpha_n$ divided by the total number of $\alpha_j$ which gives the fraction of examples that have a larger α than the new sample under investigation. This fraction is called the *p-value* (not to be confused with the traditional p-vlues from statistics) for the new sample $(x_n,y_n)$. If the p-value is

small, then ($x_n$, $y_n$) is very nonconforming, i.e. the new example is different from the previous examples because of its high α score compared to most of the examples in the calibration set. On the other hand, if the p-value is large, then ($x_n$, $y_n$) is very conforming,. i.e. very similar to the previous examples.

The following picture show a flow chart of conformal prediction:




For a given compound, a conformal predictor gives the p-value for active and inactive classes as $p_1$ and $p_0$ respectively. The output label under significance level $\mathcal{E}$ (maximum allowable fraction of erroneous prediction) can be defined as:

Active: $p_1 > \mathcal{E}$ and $p_0 \leq \mathcal{E}$

Inactive: $p_0 > \mathcal{E}$ and $p_1 \leq \mathcal{E}$

Uncertain: $p_1 > \mathcal{E}$ and $p_0 > \mathcal{E}$, $p_1 \leq \mathcal{E}$ and $p_0 \leq \mathcal{E}$

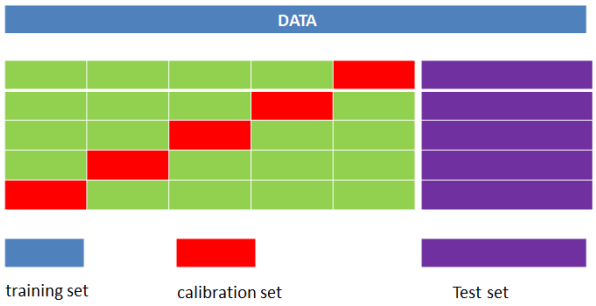
The associated output confidence for single labeled compound is defined as:

Confidence for compound labeled as Active: $1- p_0$

Confidence for compound labeled as inactive: $1- p_1$

**Text S2.** Model Construction and Performance Evaluation.

To evaluate the performance of each model, each data set was randomly divided into the training (70%) and test (30%) set. The training set was then further divided into a five-fold training set and calibrating set (see following figure). Each fold was used once as calibration set. The remaining training data were used to build underling machine learning models. The averaged *p-values* estimated from five folds were used for decision making(Sun, et al., 2017). The Random Forest (RF) module of the Scikit-Learn package (Pedregosa, et al., 2011) was used as underlying machine learning models to build binary classification RF models. The Inductive Conformal Prediction module of nonconfomist package (Linusson, 2017) was used for building Mondrial conformal predictor.




**References:**

Linusson, H. (2017) Nonconformist, https://github.com/donlnz/nonconformist.

Pedregosa, F., *et al.* (2011) Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.*, **12**, 2825-2830.

Sun, J., *et al.* (2017) Applying Mondrian Cross-Conformal Prediction To Estimate Prediction Confidence on Large Imbalanced Bioactivity Data Sets, *Journal of Chemical Information and Modeling*, **57**, 1591-1598.

**Text S3.** List of RDKit physochemical descriptors used in this study.

'BalabanJ', 'BertzCT', 'Chi0', 'Chi0n', 'Chi0v', 'Chi1', 'Chi1n', 'Chi1v', 'Chi2n', 'Chi2v', 'Chi3n', 'Chi3v', 'Chi4n', 'Chi4v', 'EState_VSA1', 'EState_VSA10', 'EState_VSA11', 'EState_VSA2', 'EState_VSA3', 'EState_VSA4', 'EState_VSA5', 'EState_VSA6', 'EState_VSA7', 'EState_VSA8', 'EState_VSA9', 'ExactMolWt', 'FractionCSP3', 'HallKierAlpha', 'HeavyAtomCount', 'HeavyAtomMolWt', 'Ipc', 'Kappa1', 'Kappa2', 'Kappa3', 'LabuteASA', 'MaxAbsEStateIndex', 'MaxAbsPartialCharge', 'MaxEStateIndex', 'MaxPartialCharge', 'MinAbsEStateIndex', 'MinAbsPartialCharge', 'MinEStateIndex', 'MinPartialCharge', 'MolLogP', 'MolMR', 'MolWt', 'NHOHCount', 'NOCount', 'NumAliphaticCarbocycles', 'NumAliphaticHeterocycles', 'NumAliphaticRings', 'NumAromaticCarbocycles', 'NumAromaticHeterocycles', 'NumAromaticRings', 'NumHAcceptors', 'NumHDonors', 'NumHeteroatoms', 'NumRadicalElectrons', 'NumRotatableBonds', 'NumSaturatedCarbocycles', 'NumSaturatedHeterocycles', 'NumSaturatedRings', 'NumValenceElectrons', 'PEOE_VSA1', 'PEOE_VSA10', 'PEOE_VSA11', 'PEOE_VSA12', 'PEOE_VSA13', 'PEOE_VSA14', 'PEOE_VSA2', 'PEOE_VSA3', 'PEOE_VSA4', 'PEOE_VSA5', 'PEOE_VSA6', 'PEOE_VSA7', 'PEOE_VSA8', 'PEOE_VSA9', 'RingCount', 'SMR_VSA1', 'SMR_VSA10', 'SMR_VSA2', 'SMR_VSA3', 'SMR_VSA4', 'SMR_VSA5', 'SMR_VSA6', 'SMR_VSA7', 'SMR_VSA8', 'SMR_VSA9', 'SlogP_VSA1', 'SlogP_VSA10', 'SlogP_VSA11', 'SlogP_VSA12', 'SlogP_VSA2', 'SlogP_VSA3', 'SlogP_VSA4', 'SlogP_VSA5', 'SlogP_VSA6', 'SlogP_VSA7', 'SlogP_VSA8', 'SlogP_VSA9', 'TPSA', 'VSA_EState1', 'VSA_EState10', 'VSA_EState2', 'VSA_EState3', 'VSA_EState4', 'VSA_EState5', 'VSA_EState6', 'VSA_EState7', 'VSA_EState8', 'VSA_EState9', 'fr_Al_COO', 'fr_Al_OH', 'fr_Al_OH_noTert', 'fr_ArN', 'fr_Ar_COO', 'fr_Ar_N', 'fr_Ar_NH', 'fr_Ar_OH', 'fr_COO', 'fr_COO2', 'fr_C_O', 'fr_C_O_noCOO', 'fr_C_S', 'fr_HOCCN', 'fr_Imine', 'fr_NH0', 'fr_NH1', 'fr_NH2', 'fr_N_O', 'fr_Ndealkylation1', 'fr_Ndealkylation2', 'fr_Nhpyrrole', 'fr_SH', 'fr_aldehyde', 'fr_alkyl_carbamate', 'fr_alkyl_halide', 'fr_allylic_oxid', 'fr_amide', 'fr_amidine', 'fr_aniline', 'fr_aryl_methyl', 'fr_azide', 'fr_azo', 'fr_barbitur', 'fr_benzene', 'fr_benzodiazepine', 'fr_bicyclic', 'fr_diazo', 'fr_dihydropyridine', 'fr_epoxide', 'fr_ester', 'fr_ether', 'fr_furan', 'fr_guanido', 'fr_halogen', 'fr_hdrzine', 'fr_hdrzone', 'fr_imidazole', 'fr_imide', 'fr_isocyan', 'fr_isothiocyan', 'fr_ketone', 'fr_ketone_Topliss', 'fr_lactam', 'fr_lactone', 'fr_methoxy', 'fr_morpholine', 'fr_nitrile', 'fr_nitro', 'fr_nitro_arom', 'fr_nitro_arom_nonortho', 'fr_nitroso', 'fr_oxazole', 'fr_oxime', 'fr_para_hydroxylation', 'fr_phenol', 'fr_phenol_noOrthoHbond', 'fr_phos_acid', 'fr_phos_ester', 'fr_piperdine', 'fr_piperzine', 'fr_priamide', 'fr_prisulfonamd', 'fr_pyridine', 'fr_quatN', 'fr_sulfide', 'fr_sulfonamd', 'fr_sulfone', 'fr_term_acetylene', 'fr_tetrazole', 'fr_thiazole', 'fr_thiocyan', 'fr_thiophene', 'fr_unbrch_alkane', 'fr_urea'

**Table S1.** Details of data sets for toxicity predictor construction. Data sets were collected from Chembl, Pubchem, Toxnet, eChemPortal databases and literature. For off-target data sets, compounds and their corresponding bioactivity information were derived from CHEMBL database and bioactivity values of 1 μ M or lower is classified as active. For those targets with very few inactive compounds, we borrow inactive compounds from their homology neighbors.

| Model_ID | Data Description | Potential Injury | N_active | N_inact |
|---|---|---|---|---|
| M_001 | Modulator of Sodium channel protein type V alpha subunit | Heart | 122 | 610 |
| M_002 | Modulator of Calcitonin gene-related peptide type 1 receptor | Heart, Nervous system, Gastrointestinal, blood | 554 | 2451 |
| M_003 | Modulator of Receptor protein-tyrosine kinase erbB-2 | Heart | 1035 | 5171 |
| M_004 | Modulator of Dopamine D1 receptor | Central nervous system, Kidney, Heart | 764 | 3824 |
| M_005 | Modulator of Sodium channel protein type IX alpha subunit | Heart | 2159 | 8554 |
| M_006 | Modulator of P2X purinoceptor 3 | Nervous system, immune, Kidney | 848 | 565 |
| M_007 | Modulator of Endothelin receptor ET-B | Heart, respiratory, blood, Kidney, Gastrointestinal | 513 | 1077 |
| M_008 | Modulator of Neuronal acetylcholine receptor protein alpha-4 subunit | Central nervous system | 178 | 890 |
| M_009 | Modulator of Serotonin transporter | Central nervous system, blood, Heart, Gastrointestinal | 3502 | 11228 |
| M_010 | Modulator of Adenosine A2a receptor | Heart, blood, Central nervous system | 2997 | 12876 |
| M_011 | Modulator of Bradykinin B2 receptor | Heart, respiratory, Kidney, Central nervous system | 383 | 1928 |
| M_012 | Modulator of Neuronal acetylcholine receptor; alpha3/beta4 | Central nervous system | 230 | 1150 |
| M_013 | Modulator of Serotonin 2c (5-HT2c) receptor | Nervous system | 2325 | 11615 |
| M_014 | Modulator of HMG-CoA reductase | Kidney | 192 | 959 |
| M_015 | Modulator of Cholecystokinin A receptor | Gastrointestinal, gallbladder, Nervous system | 265 | 1325 |
| M_016 | Modulator of TNF-alpha | Immune | 349 | 1745 |
| M_017 | Modulator of P2X purinoceptor 7 | Nervous system, immune, Kidney | 1872 | 4075 |
| M_018 | Modulator of Serotonin 2a (5-HT2a) receptor | Nervous system, blood, Heart | 3086 | 13014 |
| M_019 | Modulator of Beta-1 adrenergic receptor | Heart | 883 | 4410 |
| M_020 | Modulator of Platelet-derived growth factor receptor | Heart | 171 | 854 |
| M_021 | Modulator of GABA-A receptor; anion channel | Central nervous system | 974 | 2739 |
| M_022 | Modulator of Vascular endothelial growth factor receptor 1 | Heart | 1041 | 5198 |
| M_023 | Modulator of Platelet-derived growth factor receptor alpha | Heart | 331 | 1655 |
| M_024 | Modulator of Melatonin receptor 1B | Central nervous system | 739 | 1540 |
| M_025 | Modulator of Delta opioid receptor | Central nervous system | 2398 | 11186 |
| M_026 | Modulator of GABA-A receptor; alpha-2/beta-3/gamma-2 | Central nervous system | 492 | 1888 |
| M_027 | Modulator of Peroxisome proliferator-activated receptor gamma | Kidney, Heart, immune | 2099 | 10499 |
| M_028 | Modulator of Adenosine A1 receptor | Nervous system, Kidney, Heart | 2584 | 12934 |
| M_029 | Modulator of Muscarinic acetylcholine receptor M4 | Nervous system, Heart | 448 | 2240 |
| M_030 | Modulator of Vascular endothelial growth factor receptor | Heart | 141 | 705 |
| M_031 | Modulator of Glucocorticoid receptor | Endocrine, immune, Nervous system | 1875 | 9373 |
| M_032 | Modulator of Cyclooxygenase-2 | Kidney | 1553 | 7794 |
| M_033 | Modulator of Kappa opioid receptor | Nervous system, Heart, Gastrointestinal | 2831 | 13273 |
| M_034 | Modulator of Platelet activating factor receptor | Immune, respiratory, blood, Heart, Kidney | 310 | 688 |

| | | | | |
|---|---|---|---|---|
| M_035 | Modulator of Sigma opioid receptor | Nervous system | 1862 | 9303 |
| M_036 | Modulator of Androgen Receptor | Endocrine, Central nervous system | 1378 | 6890 |
| M_037 | Modulator of Neuropeptide Y receptor type 1 | Gastrointestinal, immune, Nervous system | 398 | 1990 |
| M_038 | Modulator of Monoamine oxidase A | Pharmaco-kinetics, Central nervous system | 450 | 2250 |
| M_039 | Modulator of Dopamine D2 receptor | Central nervous system, Gastrointestinal, Heart, Kidney | 4693 | 11918 |
| M_040 | Modulator of Vascular endothelial growth factor receptor 2 | Heart | 4642 | 9422 |
| M_041 | Modulator of Alpha-1a adrenergic receptor | Heart | 1437 | 7176 |
| M_042 | Modulator of Glutamate receptor ionotropic, AMPA | Central nervous system | 135 | 675 |
| M_043 | Modulator of Prostanoid EP2 receptor | Immune, urogenital | 162 | 810 |
| M_044 | Modulator of Serotonin 3a (5-HT3a) receptor | Gastrointestinal, Heart, Central nervous system | 365 | 1824 |
| M_045 | Modulator of Vasopressin V1a receptor | Heart, Kidney, Central nervous system | 533 | 2665 |
| M_046 | Modulator of Cannabinoid CB1 receptor | Nervous system, Heart | 2844 | 11626 |
| M_047 | Modulator of Serotonin 2b (5-HT2b) receptor | Heart, Nervous system, Gastrointestinal | 1111 | 5550 |
| M_048 | Modulator of Cannabinoid CB2 receptor | Immune, Nervous system | 3674 | 11038 |
| M_049 | Modulator of Adenosine A2b receptor | Immune, Peripheral nervous system | 1434 | 5003 |
| M_050 | Modulator of Angiotensin II type 2 (AT-2) receptor | Kidney | 288 | 1224 |
| M_051 | Modulator of Alpha-1b adrenergic receptor | Heart, Nervous system, Kidney | 1084 | 5416 |
| M_052 | Modulator of Phosphodiesterase 3 | Heart, blood | 179 | 895 |
| M_053 | Modulator of Type-1 angiotensin II receptor | Kidney, Heart | 626 | 2367 |
| M_054 | Modulator of HERG | Heart | 1606 | 8024 |
| M_055 | Modulator of Neurokinin 2 receptor | Gastrointestinal, immune, gonad | 728 | 3639 |
| M_056 | Modulator of Mu opioid receptor | Nervous system, respiratory, Gastrointestinal, Heart, pancreas, Kidney | 2783 | 12209 |
| M_057 | Modulator of GABA-A receptor; alpha-5/beta-3/gamma-2 | Central nervous system | 604 | 1887 |
| M_058 | Modulator of Alpha-2a adrenergic receptor | Heart, Nervous system, Gastrointestinal, pancreas | 620 | 3104 |
| M_059 | Modulator of Muscarinic acetylcholine receptor M5 | Nervous system | 325 | 1624 |
| M_060 | Modulator of Serotonin 4 (5-HT4) receptor | Gastrointestinal, Heart, Central nervous system | 445 | 2225 |
| M_061 | Modulator of Beta-3 adrenergic receptor | Heart | 1324 | 6614 |
| M_062 | Modulator of Serotonin 1a (5-HT1a) receptor | Nervous system, Heart, endocrine | 3257 | 11716 |
| M_063 | Modulator of Neurokinin 1 receptor | Nervous system, immune, Gastrointestinal, gonad | 2214 | 9498 |
| M_064 | Modulator of Angiotensin-converting enzyme | Heart, Kidney | 376 | 1880 |
| M_065 | Modulator of Urotensin II receptor | Kidney, Heart | 292 | 1458 |
| M_066 | Modulator of Muscarinic acetylcholine receptor M3 | Gastrointestinal, Nervous system, respiratory, Heart, liver | 1363 | 6815 |
| M_067 | Modulator of Platelet-derived growth factor receptor beta | Heart | 720 | 3598 |
| M_068 | Modulator of Neuronal acetylcholine receptor; alpha4/beta2 | Central nervous system | 592 | 2960 |
| M_069 | Modulator of Dopamine transporter | Nervous system | 1795 | 8974 |
| M_070 | Modulator of Serotonin 1b (5-HT1b) receptor | Nervous system, Heart | 907 | 4535 |
| M_071 | Modulator of Glutamate NMDA receptor | Central nervous system | 267 | 1335 |
| M_072 | Modulator of Histamine H1 receptor | Immune, Nervous system, Heart, Gastrointestinal | 910 | 4548 |
| M_073 | Modulator of Neuronal acetylcholine receptor protein alpha-7 subunit | Central nervous system | 384 | 1920 |
| M_074 | Modulator of Muscarinic acetylcholine receptor M2 | Heart, respiratory, Nervous system | 1145 | 5722 |

| | | | | |
|---|---|---|---|---|
| M_075 | Modulator of Histamine H2 receptor | Gastrointestinal, immune, Heart | 118 | 590 |
| M_076 | Modulator of Acetylcholinesterase | Central nervous system | 1846 | 9245 |
| M_077 | Modulator of Endothelin receptor ET-A | Heart, Kidney, Nervous system, Gastrointestinal | 1099 | 1007 |
| M_078 | Modulator of Muscarinic acetylcholine receptor M1 | Nervous system, respiratory, Gastrointestinal, Heart, liver | 1186 | 5935 |
| M_079 | Modulator of GABA-A receptor; alpha-1/beta-3/gamma-2 | Central nervous system | 573 | 1910 |
| M_080 | Modulator of Beta-2 adrenergic receptor | Heart, blood | 1095 | 5469 |
| M_081 | Modulator of GABA-A receptor; alpha-3/beta-3/gamma-2 | Central nervous system | 617 | 1889 |
| M_082 | Modulator of Adenosine A3 receptor | Nervous system, respiratory | 2711 | 8890 |
| M_083 | Modulator of Serotonin 7 (5-HT7) receptor | Nervous system, immune | 1417 | 7081 |
| M_084 | Modulator of Norepinephrine transporter | Heart, Nervous system | 2232 | 11150 |
| M_085 | Modulator of Alpha-2b adrenergic receptor | Heart, Nervous system | 353 | 1769 |
| M_086 | Modulator of Vascular endothelial growth factor receptor 3 | Heart | 364 | 1820 |
| M_087 | Modulator of Cyclooxygenase-1 | Kidney | 296 | 1479 |
| M_088 | Differential cytotoxicity (isogenic chicken DT40 cell lines) | Genotoxicity | 1939 | 5132 |
| M_089 | Antagonist of the glucocorticoid receptor (GR) signaling pathway | Endocrine, immune, Nervous system | 373 | 6143 |
| M_090 | Differential cytotoxicity (isogenic chicken DT40 Rev3 mutant cell line) | Genotoxicity | 2063 | 4980 |
| M_091 | Agonist of H2AX | DNA damage | 411 | 6814 |
| M_092 | Agonist of the glucocorticoid receptor (GR) signaling pathway | Endocrine, immune, Nervous system | 178 | 7158 |
| M_093 | Induce genotoxicity in human embryonic kidney cells | Genotoxicity | 284 | 7444 |
| M_094 | Cytotoxicity in HEK293 cells - 40 hour | Kidney | 702 | 6482 |
| M_095 | Activators of the human pregnane X receptor (PXR) signaling pathway | Liver | 236 | 1749 |
| M_096 | Inhibitors of Hepatocyte nuclear factor 4 (HNF4) dimerization | Liver | 1733 | 513 |
| M_097 | Antagonist of the androgen receptor (AR) signaling pathway | Endocrine, Central nervous system | 532 | 6207 |
| M_098 | Disruptors of the mitochondrial membrane potential | Liver | 992 | 5409 |
| M_099 | Agonist of the antioxidant response element (ARE) signaling pathway | Liver | 967 | 5113 |
| M_100 | Cytotoxicity in HepG2 cells - 40 hour | Liver | 854 | 6282 |
| M_101 | Antagonist of the estrogen receptor alpha (ER-alpha) signaling pathway | Endocrine | 376 | 6361 |
| M_102 | Inhibitors and Substrates of Cytochrome P450 3A4 | Liver | 3435 | 7013 |
| M_103 | Inhibitors and Substrates of Cytochrome P450 2C9 | Liver | 1266 | 6902 |
| M_104 | Antagonist of the farnesoid-X-receptor (FXR) signaling pathway | Liver | 315 | 6464 |
| M_105 | Activator the aryl hydrocarbon receptor (AhR) signaling pathway | Liver | 873 | 6352 |
| M_106 | Antagonist of the constitutive androstane receptor (CAR) signaling pathway | Liver | 460 | 6823 |
| M_107 | Activators of the heat shock response signaling pathway | Liver | 412 | 6370 |
| M_108 | Inhibitors and Substrates of Cytochrome P450 2D6 | Liver | 1616 | 6295 |
| M_109 | Differential cytotoxicity against isogenic chicken DT40 cell lines with known DNA damage response pathways - Rad54/Ku70 mutant cell line | Genotoxicity | 2070 | 5230 |
| M_110 | Agonist of the p53 signaling pathway | DNA damage | 510 | 6922 |
| M_111 | Antagonist of the thyroid receptor (TR) signaling pathway | Endocrine, Heart | 370 | 5602 |
| M_112 | Antagonist of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway | Kidney, Heart, immune | 375 | 5856 |
| M_113 | Cytotoxicity in HepG2 cells - 16 hour | Liver | 487 | 6820 |
| M_114 | Antagonist of the retinoic acid receptor (RAR) signaling | Immune | 576 | 5095 |

| | | | | |
|---|---|---|---|---|
| | pathway | | | |
| M_115 | Cytotoxicity in HEK293 cells - 32 hour | Kidney | 583 | 6610 |
| M_116 | Cytotoxicity in HepG2 cells - 8 hour | Liver | 329 | 7058 |
| M_117 | Antagonist of the vitamin D receptor (VDR) signaling pathway | Liver | 364 | 6035 |
| M_118 | Antagonist of the estrogen receptor alpha (ER-alpha) signaling pathway | Endocrine | 391 | 6508 |
| M_119 | Agonist of the androgen receptor (AR) signaling pathway | Endocrine, Central nervous system | 249 | 7126 |
| M_120 | Agonist of the AP-1 signaling pathway | Cancer | 599 | 6406 |
| M_121 | Cytotoxicity in HepG2 cells - 32 hour | Liver | 758 | 6383 |
| M_122 | Cytotoxicity in HepG2 cells - 0 hour | Liver | 111 | 7419 |
| M_123 | Cytotoxicity in HEK293 cells - 24 hour | Kidney | 477 | 6792 |
| M_124 | Antagonist of the retinoid-related orphan receptor gamma (ROR-gamma) signaling pathway | Immune | 529 | 4999 |
| M_125 | Agonist of the estrogen receptor alpha (ER-alpha) signaling pathway | Endocrine | 422 | 7256 |
| M_126 | Agonist of the thyroid stimulating hormone receptor (TSHR) signaling pathway | Endocrine, Heart | 330 | 6822 |
| M_127 | Agonist of the constitutive androstane receptor (CAR) signaling pathway | Liver | 893 | 6092 |
| M_128 | Agonist of the RXR signaling pathway | Liver | 217 | 5865 |
| M_129 | Antagonist of the farnesoid-X-receptor (FXR) signaling pathway | Liver | 226 | 6154 |
| M_130 | Cytotoxicity in HEK293 cells - 8 hour | Kidney | 282 | 7062 |
| M_131 | Substrates of Cytochrome P450 2C19 | Liver | 1895 | 6400 |
| M_132 | Activators of Cytochrome P450 3A4 | Liver | 165 | 12631 |
| M_133 | Antagonist of the androgen receptor (AR) signaling pathway | Endocrine, Central nervous system | 406 | 6256 |
| M_134 | Inhibit CYP1A2 Activity | Liver | 4164 | 3639 |
| M_135 | Cytotoxicity in HEK293 cells - 16 hour | Heart | 343 | 6982 |
| M_136 | Agonist of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway | Kidney, Heart, immune | 216 | 6851 |
| M_137 | Cytotoxicity in HepG2 cells - 24 hour | Liver | 648 | 6547 |
| M_138 | Agonist of the farnesoid-X-receptor (FXR) signaling pathway | Liver | 446 | 5829 |
| M_139 | Aromatase inhibitors | Endocrine | 323 | 6075 |
| M_140 | Antagonist of the constitutive androstane receptor (CAR) signaling pathway | Liver | 166 | 5239 |
| M_141 | Activators of Cytochrome P450 2A9 | Liver | 1368 | 13678 |
| M_142 | Activator Alzheimer's amyloid precursor | Central nervous system | 1987 | 19870 |
| M_143 | Modulators of myocardial damage | Heart | 2272 | 22716 |
| M_144 | Inhibit CYP2C9 Acitivity | Liver | 18731 | 18731 |
| M_145 | Inhibit CYP2C19 Activity | Liver | 20295 | 20295 |
| M_146 | Acute Oral Toxicity (rodent, LD50 < 5mg/kg) | Rodent | 751 | 11072 |
| M_147 | Induce Phospholipidosis | Liver | 442 | 789 |
| M_148 | Block OATP1B1 Transporter | Liver | 1640 | 200 |
| M_149 | Agonist of Liver X receptor beta | Liver | 560 | 3415 |
| M_150 | Mutagenicity | Mutagenicity | 3296 | 2752 |
| M_151 | Acute Oral Toxicity (rodent, LD50 < 300mg/kg) | Rodent | 6640 | 5063 |
| M_152 | Acute Oral Toxicity (rodent, LD50 < 50mg/kg) | Rodent | 2664 | 9069 |
| M_153 | Agonist of Liver X receptor alpha | Liver | 436 | 3415 |
| M_154 | Block OATP1B3 transporter | Liver | 1726 | 130 |
| M_155 | Block Bile Salt Export Pump | Liver | 361 | 309 |

| | | | | |
|---|---|---|---|---|
| M_156 | Acute Oral Toxicity (Rat, LD50 < 300mg/kg) | Rat | 2700 | 9227 |
| M_157 | Acute Oral Toxicity (Rat, LD50 < 5mg/kg) | Rat | 237 | 11690 |
| M_158 | Acute Oral Toxicity (Rat, LD50 < 50mg/kg) | Rat | 1002 | 10925 |
| M_159 | Carcinogenic Potency (Salmonella Mutagenicity) | carcinogenicity | 392 | 426 |
| M_160 | Carcinogenic Potency ( SingleCellCall) | carcinogenicity | 780 | 675 |
| M_161 | Carcinogenic Potency (Rat) | carcinogenicity | 571 | 597 |
| M_162 | Acute Oral Toxicity (Mouse, LD50 < 300mg/kg) | Mouse | 4623 | 20233 |
| M_163 | Acute Oral Toxicity (Mouse, LD50 < 5mg/kg) | Mouse | 149 | 24707 |
| M_164 | Acute Oral Toxicity (Mouse, LD50 < 50mg/kg) | Mouse | 844 | 24012 |
| M_165 | Acute Oral Toxicity (Rabbit, LD50 < 300mg/kg) | Rabbit | 227 | 695 |
| M_166 | Drug Induced Liver Injury | Liver | 780 | 1042 |
| M_167 | Developmental Toxicity | develop | 712 | 127 |
| M_168 | Endocine Disruption | reproduction | 648 | 430 |
| M_169 | Eye Corrosion | eye | 883 | 1412 |
| M_170 | Eye Irritation | eye | 3861 | 1346 |
| M_171 | Fish Tox (high risk) | fish | 637 | 1387 |
| M_172 | Fish Tox (moderate risk) | fish | 1158 | 866 |
| M_173 | Skin Sensitization | skin | 343 | 329 |
| M_174 | Aquatic Tox(T pyriformis) | T pyriformis | 1143 | 307 |

**Table S2.** Summary of Structural Alerts used in eMolTox analysis. Alerts were collected from Chembl, ToxAlerts and literatures.

| ID | Alert Name | Number of Alerts | Reference |
|---|---|---|---|
| 0 | Chelating agents | 55 | 1. Agrawal A, Johnson SL, Jacobsen JA, Miller MT, Chen LH, Pellecchia M, Cohen SM. Chelator fragment libraries for targeting metalloproteinases. ChemMedChem. 2010 Feb 1;5(2):195-9. |
| 1 | Acute Aquatic Toxicity | 54 | 1. Hermens JL. Electrophiles and acute toxicity to fish. Environmental Health Perspectives. 1990;87:219-225.<br>2. Henk J.M. Verhaar, Cees J. van Leeuwen, Joop L.M. Hermens, Classifying environmental pollutants, Chemosphere, Volume 25, Issue 4, 1992, Pages 471-491. |
| 2 | Covalent Bind With Protein | 108 | 1. Enoch SJ, Ellison CM, Schultz TW, Cronin MTD. (2011). A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. Crit Rev Toxicol 41:783-802. |
| 3 | Respiratory Sensitization | 52 | 1. Enoch SJ et al, Development of Mechanism-Based Structural Alerts for Respiratory Sensitization Hazard Identification. Chem. Res. Toxicol., 2012, 25 (11), pp 2490–2498 |
| 5 | Mitochondrial Toxicity | 17 | 1. Nelms MD, Mellor CL, Cronin MT, Madden JC, Enoch SJ. Development of an in silico profiler for mitochondrial toxicity. Chemical research in toxicology. 2015 Sep 24;28(10):1891-902. |
| 6 | Non genotoxic carcinogenicity | 22 | 1. Benigni R, Bossa C, Tcheremenskaia O. Nongenotoxic carcinogenicity of chemicals: mechanisms of action and early recognition through a new set of structural alerts. Chemical reviews. 2013 Mar 8;113(5):2940-57. |
| 8 | Skin sensitization | 135 | 1. Payne MP, Walsh PT. Structure-activity relationships for skin sensitization potential: development of structural alerts for use in knowledge-based toxicity prediction systems. Journal of chemical information and computer sciences. 1994 Jan;34(1):154-61.<br>2. Barratt MD, Basketter DA, Chamberlain M, Admans GD, Langowski JJ. An expert system rulebase for identifying contact allergens. Toxicology in vitro. 1994 Oct 1;8(5):1053-60.<br>3. Enoch SJ, Madden JC, Cronin MT. Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. SAR and QSAR in Environmental Research. 2008 Jul 1;19(5-6):555-78.<br>4. Gerner I, Barratt MD, Zinke S, Schlegel K, Schlede E. Development and prevalidation of a list of structure-activity relationship rules to be used in expert systems for prediction of the skin-sensitising properties of chemicals. Alternatives to laboratory animals: ATLA. 2004 Nov;32(5):487-509. |
| 9 | Alert from Top200 Drug | 82 | 1. Stepan AF, Walker DP, Bauman J, Price DA, Baillie TA, Kalgutkar AS, Aleo MD. Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States. Chemical research in toxicology. 2011 Jul 11;24(9):1345-410. |
| 10 | Genotoxic carcinogenicity mutagenicity | 81 | 1. Ashby J, Tennant RW. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the US NCI/NTP. Mutation Research/Genetic Toxicology. 1988 Jan 1;204(1):17-15.<br>2. Kazius J, McGuire R, Bursi R. Derivation and validation of toxicophores for mutagenicity prediction. Journal of medicinal chemistry. 2005 Jan 13;48(1):312-20.<br>3. Bailey AB, Chanderbhan R, Collazo-Braier N, Cheeseman MA, Twaroski ML. The use of structure–activity relationship analysis in the food contact notification program. Regulatory Toxicology and Pharmacology. 2005 Jul 31;42(2):225-35.<br>4. Benigni R, Bossa C. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutation Research/Reviews in Mutation Research. 2008 Oct 31;659(3):248-61. |
| 11 | Kidney Toxicity | 6 | 1. Pizzo F, Gadaleta D, Lombardo A, Nicolotti O, Benfenati E. Identification of structural alerts for liver and kidney toxicity using repeated dose toxicity data. Chemistry Central Journal. 2015 Dec 1;9(1):62. |
| 12 | Hepatotoxicity | 36 | 1. Liu R, Yu X, Wallqvist A. Data-driven identification of structural alerts for mitigating the risk of drug-induced human liver injuries. Journal of cheminformatics. 2015 Dec 1;7(1):4.<br>2. Pizzo F, Gadaleta D, Lombardo A, Nicolotti O, Benfenati E. Identification of structural alerts for liver and kidney toxicity using repeated dose toxicity data. Chemistry Central Journal. 2015 Dec 1;9(1):62.<br>3. Hewitt M, Enoch SJ, Madden JC, Przybylak KR, Cronin MT. Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism (s) of action. Critical reviews in toxicology. 2013 Aug 1;43(7):537-58. |
| 13 | Developmental and mitochondrial toxicity | 12 | 1. Structural alerts for devlopmental toxicity and mitochondrial toxicity molecular initiating events (Lhasa), Mukesh,P et al. 2014 |

| | | | |
|---|---|---|---|
| 14 | Covalent Bind With DNA | 111 | 1. Benigni R and Bossa C. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutation Research 2008, 659, 248-261.<br>2. Enoch SJ, Cronin MTD. A review of the electrophilic reaction chemistry involved in covalent DNA binding. Crit Rev Toxicol 2010,40(8),728-748.<br>3. Kalgutkar AS et al. A comprehensive listing of bioactivation pathways of organic functional groups. Current Drug Metabolism 2005; 6, 161-225.<br>4. Kazius J et al. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J Med Chem 2005; 48, 312-320.<br>5. Mekenyan O et al. Identification of the Structural Requirements for Mutagenicity by Incorporating Molecular Flexibility and Metabolic Activation of Chemicals I: TA100 Model. Chem Res Toxicol 2004, 17, 753-766.<br>6. Mekenyan O et al. Identifying the Structural Requirements for Chromosomal Aberration by Incorporating Molecular Flexibility and Metabolic Activation of Chemicals. Chem Res Toxicol 2007, 20, 1927-1941.<br>7. Serafimova R et al. Identification of the Structural Requirements for Mutagencitiy, by Incorporating Molecular Flexibility and Metabolic Activation of Chemicals. II. General Ames Mutagenicity Model. Chem Res Toxicol 2007, 20, 662-676. |
| 15 | Idiosyncratic toxicity Metabolic activation | 32 | 1. Kalgutkar AS, Soglia JR. Minimising the potential for metabolic activation in drug discovery. Expert opinion on drug metabolism & toxicology. 2005 Jun 1;1(1):91-142. |
| 16 | Potential electrophilic agents | 119 | 1. Enoch SJ, Ellison CM, Schultz TW, Cronin MT. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. Critical reviews in toxicology. 2011 Oct 1;41(9):783-802. |

**Table S3.** Performance of Mondral Conformal Predictors for all 174 Data Sets at three significance levels.

| Model_ID | Ɛ = 0.05 | | | Ɛ = 0.10 | | | Ɛ = 0.20 | | |
|---|---|---|---|---|---|---|---|---|---|
| | efficiency | validity (positive set) | validity (negative set) | efficiency | validity (positive set) | validity (negative set) | efficiency | validity (positive set) | validity (negative set) |
| M_001 | 0.44 | 1.00 | 0.97 | 0.94 | 0.92 | 0.90 | 0.80 | 0.62 | 0.83 |
| M_002 | 0.98 | 0.98 | 0.98 | 0.94 | 0.96 | 0.93 | 0.85 | 0.85 | 0.86 |
| M_003 | 0.98 | 0.97 | 0.94 | 0.91 | 0.89 | 0.89 | 0.81 | 0.80 | 0.80 |
| M_004 | 0.99 | 0.90 | 0.96 | 0.92 | 0.86 | 0.91 | 0.82 | 0.80 | 0.82 |
| M_005 | 0.97 | 0.95 | 0.96 | 0.93 | 0.92 | 0.91 | 0.82 | 0.82 | 0.81 |
| M_006 | 0.97 | 0.97 | 0.97 | 0.92 | 0.92 | 0.90 | 0.81 | 0.84 | 0.78 |
| M_007 | 0.97 | 0.96 | 0.94 | 0.92 | 0.90 | 0.88 | 0.81 | 0.84 | 0.79 |
| M_008 | 0.94 | 0.96 | 0.96 | 0.99 | 0.89 | 0.96 | 0.91 | 0.78 | 0.90 |
| M_009 | 1.00 | 0.95 | 0.96 | 0.93 | 0.90 | 0.91 | 0.81 | 0.81 | 0.81 |
| M_010 | 0.99 | 0.94 | 0.96 | 0.93 | 0.91 | 0.91 | 0.83 | 0.82 | 0.82 |
| M_011 | 0.96 | 0.97 | 0.96 | 0.92 | 0.94 | 0.92 | 0.84 | 0.79 | 0.85 |
| M_012 | 0.80 | 0.94 | 0.97 | 0.99 | 0.87 | 0.93 | 0.84 | 0.80 | 0.82 |
| M_013 | 0.99 | 0.96 | 0.96 | 0.93 | 0.91 | 0.91 | 0.83 | 0.82 | 0.82 |
| M_014 | 0.97 | 0.88 | 0.96 | 0.92 | 0.88 | 0.90 | 0.80 | 0.79 | 0.79 |
| M_015 | 0.99 | 0.96 | 0.96 | 0.95 | 0.91 | 0.94 | 0.82 | 0.84 | 0.81 |
| M_016 | 0.91 | 0.98 | 0.98 | 0.94 | 0.92 | 0.94 | 0.87 | 0.76 | 0.88 |
| M_017 | 0.97 | 0.95 | 0.96 | 0.92 | 0.91 | 0.92 | 0.82 | 0.82 | 0.81 |
| M_018 | 0.98 | 0.97 | 0.96 | 0.93 | 0.92 | 0.91 | 0.82 | 0.82 | 0.81 |
| M_019 | 0.98 | 0.95 | 0.95 | 0.92 | 0.89 | 0.91 | 0.83 | 0.82 | 0.83 |
| M_020 | 0.61 | 1.00 | 0.94 | 0.98 | 0.96 | 0.89 | 0.86 | 0.82 | 0.84 |
| M_021 | 1.00 | 0.97 | 0.95 | 0.93 | 0.93 | 0.90 | 0.82 | 0.84 | 0.79 |
| M_022 | 0.63 | 0.96 | 0.96 | 0.85 | 0.92 | 0.92 | 0.92 | 0.83 | 0.82 |
| M_023 | 0.53 | 0.96 | 0.95 | 0.71 | 0.91 | 0.90 | 0.96 | 0.77 | 0.78 |
| M_024 | 0.97 | 0.98 | 0.95 | 0.92 | 0.92 | 0.91 | 0.81 | 0.83 | 0.80 |
| M_025 | 0.98 | 0.96 | 0.95 | 0.93 | 0.92 | 0.91 | 0.82 | 0.82 | 0.81 |
| M_026 | 0.98 | 0.95 | 0.97 | 0.90 | 0.88 | 0.91 | 0.78 | 0.77 | 0.78 |
| M_027 | 0.98 | 0.96 | 0.95 | 0.92 | 0.91 | 0.91 | 0.82 | 0.82 | 0.80 |
| M_028 | 0.99 | 0.96 | 0.96 | 0.93 | 0.91 | 0.91 | 0.82 | 0.79 | 0.82 |
| M_029 | 0.69 | 0.97 | 0.97 | 0.96 | 0.92 | 0.93 | 0.84 | 0.78 | 0.81 |
| M_030 | 0.79 | 0.98 | 0.97 | 0.95 | 0.95 | 0.90 | 0.81 | 0.83 | 0.79 |
| M_031 | 0.96 | 0.95 | 0.96 | 0.91 | 0.89 | 0.91 | 0.81 | 0.81 | 0.81 |
| M_032 | 0.90 | 0.95 | 0.94 | 0.98 | 0.89 | 0.89 | 0.85 | 0.78 | 0.81 |
| M_033 | 0.99 | 0.95 | 0.95 | 0.92 | 0.91 | 0.91 | 0.82 | 0.81 | 0.81 |
| M_034 | 0.94 | 1.00 | 0.94 | 0.93 | 0.96 | 0.89 | 0.83 | 0.86 | 0.82 |
| M_035 | 0.96 | 0.96 | 0.95 | 0.95 | 0.91 | 0.91 | 0.84 | 0.82 | 0.82 |
| M_036 | 0.98 | 0.95 | 0.95 | 0.92 | 0.90 | 0.91 | 0.83 | 0.81 | 0.83 |
| M_037 | 0.98 | 0.97 | 0.97 | 0.92 | 0.87 | 0.92 | 0.82 | 0.79 | 0.82 |
| M_038 | 0.75 | 0.99 | 0.97 | 0.91 | 0.94 | 0.94 | 0.87 | 0.87 | 0.82 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| M_039 | 0.99 | 0.96 | 0.95 | 0.93 | 0.92 | 0.90 | 0.82 | 0.82 | 0.80 |
| M_040 | 0.93 | 0.96 | 0.95 | 0.97 | 0.92 | 0.91 | 0.84 | 0.81 | 0.82 |
| M_041 | 0.99 | 0.96 | 0.95 | 0.93 | 0.91 | 0.91 | 0.83 | 0.80 | 0.82 |
| M_042 | 0.64 | 1.00 | 0.96 | 0.97 | 1.00 | 0.89 | 0.82 | 0.83 | 0.80 |
| M_043 | 0.90 | 0.90 | 0.95 | 0.96 | 0.86 | 0.91 | 0.79 | 0.65 | 0.79 |
| M_044 | 0.81 | 0.99 | 0.96 | 0.94 | 0.93 | 0.90 | 0.79 | 0.76 | 0.79 |
| M_045 | 0.98 | 0.96 | 0.97 | 0.94 | 0.94 | 0.93 | 0.85 | 0.87 | 0.83 |
| M_046 | 1.00 | 0.95 | 0.96 | 0.93 | 0.91 | 0.90 | 0.83 | 0.83 | 0.82 |
| M_047 | 0.83 | 0.98 | 0.96 | 0.99 | 0.93 | 0.91 | 0.82 | 0.84 | 0.79 |
| M_048 | 1.00 | 0.96 | 0.95 | 0.93 | 0.91 | 0.91 | 0.83 | 0.83 | 0.82 |
| M_049 | 0.99 | 0.97 | 0.96 | 0.94 | 0.93 | 0.91 | 0.83 | 0.84 | 0.81 |
| M_050 | 0.97 | 0.94 | 0.97 | 0.93 | 0.91 | 0.91 | 0.78 | 0.73 | 0.79 |
| M_051 | 0.99 | 0.97 | 0.96 | 0.92 | 0.92 | 0.91 | 0.83 | 0.83 | 0.82 |
| M_052 | 0.75 | 1.00 | 0.98 | 0.96 | 0.93 | 0.95 | 0.85 | 0.82 | 0.85 |
| M_053 | 0.98 | 0.98 | 0.96 | 0.91 | 0.90 | 0.91 | 0.80 | 0.81 | 0.80 |
| M_054 | 0.62 | 0.97 | 0.95 | 0.83 | 0.95 | 0.90 | 0.91 | 0.83 | 0.81 |
| M_055 | 0.99 | 0.99 | 0.97 | 0.92 | 0.94 | 0.92 | 0.82 | 0.78 | 0.82 |
| M_056 | 1.00 | 0.97 | 0.95 | 0.93 | 0.91 | 0.91 | 0.82 | 0.79 | 0.82 |
| M_057 | 0.97 | 0.97 | 0.96 | 0.92 | 0.95 | 0.91 | 0.83 | 0.83 | 0.82 |
| M_058 | 0.83 | 0.95 | 0.95 | 0.98 | 0.90 | 0.91 | 0.85 | 0.81 | 0.82 |
| M_059 | 0.58 | 1.00 | 0.96 | 0.83 | 0.98 | 0.93 | 0.89 | 0.87 | 0.83 |
| M_060 | 0.97 | 0.96 | 0.96 | 0.92 | 0.93 | 0.92 | 0.83 | 0.82 | 0.83 |
| M_061 | 0.97 | 0.97 | 0.96 | 0.92 | 0.95 | 0.91 | 0.84 | 0.86 | 0.84 |
| M_062 | 1.00 | 0.96 | 0.94 | 0.92 | 0.93 | 0.89 | 0.81 | 0.82 | 0.80 |
| M_063 | 0.96 | 0.93 | 0.96 | 0.92 | 0.89 | 0.92 | 0.82 | 0.81 | 0.82 |
| M_064 | 0.99 | 0.94 | 0.95 | 0.91 | 0.88 | 0.90 | 0.81 | 0.77 | 0.81 |
| M_065 | 0.95 | 0.97 | 0.95 | 0.89 | 0.90 | 0.89 | 0.79 | 0.84 | 0.79 |
| M_066 | 0.99 | 0.96 | 0.96 | 0.93 | 0.91 | 0.92 | 0.83 | 0.81 | 0.83 |
| M_067 | 0.70 | 0.98 | 0.97 | 0.93 | 0.93 | 0.92 | 0.85 | 0.76 | 0.82 |
| M_068 | 0.95 | 0.96 | 0.95 | 0.95 | 0.93 | 0.90 | 0.83 | 0.85 | 0.80 |
| M_069 | 0.99 | 0.95 | 0.95 | 0.93 | 0.89 | 0.90 | 0.81 | 0.77 | 0.81 |
| M_070 | 0.99 | 0.95 | 0.96 | 0.91 | 0.88 | 0.91 | 0.83 | 0.81 | 0.83 |
| M_071 | 0.92 | 0.98 | 0.98 | 0.97 | 0.88 | 0.94 | 0.86 | 0.86 | 0.85 |
| M_072 | 0.95 | 0.96 | 0.96 | 0.95 | 0.92 | 0.91 | 0.84 | 0.81 | 0.83 |
| M_073 | 0.82 | 0.97 | 0.97 | 0.98 | 0.93 | 0.90 | 0.83 | 0.80 | 0.81 |
| M_074 | 0.90 | 0.97 | 0.95 | 0.95 | 0.91 | 0.90 | 0.82 | 0.79 | 0.81 |
| M_075 | 0.60 | 0.97 | 0.94 | 0.87 | 0.89 | 0.90 | 0.92 | 0.77 | 0.84 |
| M_076 | 0.95 | 0.94 | 0.96 | 0.96 | 0.91 | 0.91 | 0.85 | 0.83 | 0.83 |
| M_077 | 0.86 | 0.97 | 0.98 | 0.99 | 0.92 | 0.92 | 0.85 | 0.84 | 0.82 |
| M_078 | 0.89 | 0.96 | 0.96 | 0.97 | 0.90 | 0.91 | 0.84 | 0.81 | 0.81 |
| M_079 | 0.97 | 0.97 | 0.94 | 0.91 | 0.93 | 0.90 | 0.82 | 0.86 | 0.80 |
| M_080 | 0.98 | 0.95 | 0.95 | 0.93 | 0.89 | 0.90 | 0.82 | 0.79 | 0.82 |
| M_081 | 0.97 | 0.95 | 0.96 | 0.91 | 0.88 | 0.91 | 0.81 | 0.76 | 0.82 |
| M_082 | 0.99 | 0.97 | 0.95 | 0.94 | 0.92 | 0.90 | 0.81 | 0.78 | 0.81 |

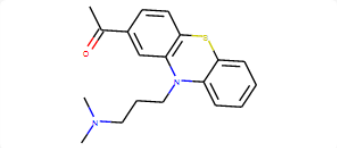
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| M_083 | 0.97 | 0.97 | 0.96 | 0.93 | 0.91 | 0.91 | 0.82 | 0.81 | 0.81 |
| M_084 | 0.99 | 0.96 | 0.95 | 0.92 | 0.91 | 0.90 | 0.82 | 0.79 | 0.82 |
| M_085 | 0.77 | 0.96 | 0.95 | 0.93 | 0.91 | 0.92 | 0.88 | 0.86 | 0.80 |
| M_086 | 0.48 | 0.93 | 0.96 | 0.73 | 0.88 | 0.90 | 1.00 | 0.78 | 0.78 |
| M_087 | 0.45 | 0.99 | 0.95 | 0.80 | 0.96 | 0.89 | 0.95 | 0.81 | 0.80 |
| M_088 | 0.49 | 0.95 | 0.97 | 0.71 | 0.90 | 0.92 | 0.97 | 0.79 | 0.81 |
| M_089 | 0.39 | 0.98 | 0.98 | 0.67 | 0.96 | 0.93 | 0.97 | 0.85 | 0.83 |
| M_090 | 0.39 | 0.97 | 0.96 | 0.61 | 0.93 | 0.91 | 0.90 | 0.83 | 0.81 |
| M_091 | 0.34 | 0.98 | 0.96 | 0.59 | 0.91 | 0.91 | 0.89 | 0.83 | 0.81 |
| M_092 | 0.20 | 1.00 | 0.97 | 0.50 | 0.98 | 0.94 | 0.84 | 0.93 | 0.84 |
| M_093 | 0.27 | 0.99 | 0.97 | 0.50 | 0.98 | 0.93 | 0.85 | 0.86 | 0.84 |
| M_094 | 0.24 | 0.99 | 0.96 | 0.49 | 0.95 | 0.91 | 0.82 | 0.86 | 0.83 |
| M_095 | 0.48 | 0.96 | 0.97 | 0.76 | 0.92 | 0.91 | 1.00 | 0.83 | 0.81 |
| M_096 | 0.38 | 0.96 | 0.98 | 0.61 | 0.93 | 0.90 | 0.90 | 0.82 | 0.82 |
| M_097 | 0.52 | 0.96 | 0.96 | 0.72 | 0.93 | 0.91 | 0.99 | 0.81 | 0.82 |
| M_098 | 0.61 | 0.98 | 0.96 | 0.82 | 0.95 | 0.91 | 0.94 | 0.87 | 0.79 |
| M_099 | 0.34 | 0.98 | 0.96 | 0.58 | 0.91 | 0.91 | 0.90 | 0.79 | 0.80 |
| M_100 | 0.26 | 0.96 | 0.96 | 0.48 | 0.93 | 0.92 | 0.81 | 0.84 | 0.81 |
| M_101 | 0.36 | 0.97 | 0.96 | 0.66 | 0.91 | 0.92 | 0.96 | 0.80 | 0.82 |
| M_102 | 0.68 | 0.96 | 0.95 | 0.87 | 0.92 | 0.90 | 0.93 | 0.83 | 0.81 |
| M_103 | 0.36 | 0.97 | 0.96 | 0.65 | 0.92 | 0.91 | 0.96 | 0.83 | 0.80 |
| M_104 | 0.46 | 0.96 | 0.97 | 0.69 | 0.90 | 0.93 | 0.97 | 0.80 | 0.83 |
| M_105 | 0.57 | 0.95 | 0.96 | 0.78 | 0.90 | 0.90 | 0.99 | 0.83 | 0.80 |
| M_106 | 0.28 | 0.94 | 0.97 | 0.51 | 0.88 | 0.92 | 0.80 | 0.78 | 0.83 |
| M_107 | 0.13 | 0.95 | 0.96 | 0.32 | 0.94 | 0.92 | 0.72 | 0.76 | 0.82 |
| M_108 | 0.57 | 0.98 | 0.96 | 0.77 | 0.93 | 0.91 | 0.98 | 0.82 | 0.81 |
| M_109 | 0.45 | 0.96 | 0.95 | 0.66 | 0.90 | 0.90 | 0.95 | 0.80 | 0.79 |
| M_110 | 0.41 | 0.95 | 0.97 | 0.62 | 0.92 | 0.93 | 0.92 | 0.78 | 0.81 |
| M_111 | 0.45 | 0.96 | 0.96 | 0.65 | 0.93 | 0.90 | 0.91 | 0.85 | 0.80 |
| M_112 | 0.40 | 0.96 | 0.97 | 0.61 | 0.92 | 0.92 | 0.91 | 0.83 | 0.81 |
| M_113 | 0.23 | 0.95 | 0.96 | 0.48 | 0.88 | 0.92 | 0.83 | 0.80 | 0.82 |
| M_114 | 0.39 | 0.94 | 0.96 | 0.61 | 0.91 | 0.91 | 0.93 | 0.81 | 0.79 |
| M_115 | 0.27 | 0.96 | 0.96 | 0.55 | 0.90 | 0.92 | 0.89 | 0.77 | 0.82 |
| M_116 | 0.14 | 0.99 | 0.97 | 0.33 | 0.93 | 0.92 | 0.69 | 0.80 | 0.84 |
| M_117 | 0.44 | 0.93 | 0.96 | 0.68 | 0.87 | 0.91 | 0.98 | 0.77 | 0.80 |
| M_118 | 0.51 | 0.97 | 0.97 | 0.71 | 0.94 | 0.92 | 0.95 | 0.87 | 0.81 |
| M_119 | 0.10 | 1.00 | 0.97 | 0.24 | 0.99 | 0.95 | 0.72 | 0.91 | 0.88 |
| M_120 | 0.31 | 0.99 | 0.95 | 0.53 | 0.95 | 0.91 | 0.85 | 0.88 | 0.82 |
| M_121 | 0.27 | 0.95 | 0.95 | 0.50 | 0.93 | 0.90 | 0.86 | 0.81 | 0.81 |
| M_122 | 0.02 | 1.00 | 0.99 | 0.04 | 1.00 | 0.97 | 0.17 | 1.00 | 0.91 |
| M_123 | 0.18 | 1.00 | 0.97 | 0.45 | 0.94 | 0.92 | 0.84 | 0.80 | 0.83 |
| M_124 | 0.48 | 0.95 | 0.96 | 0.71 | 0.91 | 0.91 | 0.98 | 0.85 | 0.80 |
| M_125 | 0.26 | 0.93 | 0.97 | 0.45 | 0.89 | 0.92 | 0.84 | 0.76 | 0.83 |
| M_126 | 0.30 | 0.98 | 0.97 | 0.55 | 0.94 | 0.93 | 0.87 | 0.83 | 0.83 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| M_127 | 0.63 | 0.96 | 0.96 | 0.81 | 0.91 | 0.92 | 0.96 | 0.80 | 0.81 |
| M_128 | 0.04 | 1.00 | 0.98 | 0.13 | 0.95 | 0.95 | 0.39 | 0.75 | 0.87 |
| M_129 | 0.45 | 0.99 | 0.96 | 0.69 | 0.93 | 0.91 | 0.95 | 0.82 | 0.82 |
| M_130 | 0.18 | 0.95 | 0.97 | 0.36 | 0.91 | 0.92 | 0.70 | 0.81 | 0.83 |
| M_131 | 0.44 | 0.96 | 0.96 | 0.67 | 0.92 | 0.91 | 0.95 | 0.85 | 0.81 |
| M_132 | 0.21 | 1.00 | 0.96 | 0.64 | 0.96 | 0.92 | 0.94 | 0.84 | 0.85 |
| M_133 | 0.44 | 0.95 | 0.97 | 0.66 | 0.93 | 0.92 | 0.94 | 0.76 | 0.81 |
| M_134 | 0.57 | 0.95 | 0.96 | 0.75 | 0.91 | 0.93 | 0.99 | 0.82 | 0.81 |
| M_135 | 0.20 | 0.95 | 0.97 | 0.44 | 0.92 | 0.93 | 0.79 | 0.83 | 0.84 |
| M_136 | 0.24 | 0.99 | 0.97 | 0.52 | 0.92 | 0.92 | 0.88 | 0.75 | 0.83 |
| M_137 | 0.23 | 0.96 | 0.96 | 0.46 | 0.91 | 0.93 | 0.81 | 0.83 | 0.83 |
| M_138 | 0.49 | 0.94 | 0.96 | 0.71 | 0.89 | 0.93 | 0.97 | 0.78 | 0.84 |
| M_139 | 0.29 | 1.00 | 0.97 | 0.57 | 0.97 | 0.91 | 0.89 | 0.91 | 0.80 |
| M_140 | 0.14 | 1.00 | 0.97 | 0.40 | 0.96 | 0.93 | 0.75 | 0.88 | 0.84 |
| M_141 | 0.42 | 0.96 | 0.96 | 0.67 | 0.89 | 0.92 | 0.95 | 0.79 | 0.82 |
| M_142 | 0.67 | 0.94 | 0.96 | 0.86 | 0.90 | 0.91 | 0.94 | 0.81 | 0.82 |
| M_143 | 0.60 | 0.97 | 0.96 | 0.80 | 0.92 | 0.91 | 0.96 | 0.83 | 0.81 |
| M_144 | 0.32 | 0.95 | 0.96 | 0.50 | 0.91 | 0.91 | 0.79 | 0.82 | 0.80 |
| M_145 | 0.29 | 0.96 | 0.96 | 0.46 | 0.92 | 0.92 | 0.74 | 0.83 | 0.82 |
| M_146 | 0.61 | 0.95 | 0.95 | 0.84 | 0.89 | 0.91 | 0.94 | 0.82 | 0.82 |
| M_147 | 0.36 | 0.99 | 0.97 | 0.60 | 0.91 | 0.95 | 0.94 | 0.85 | 0.87 |
| M_148 | 0.32 | 0.95 | 0.97 | 0.59 | 0.89 | 0.92 | 0.85 | 0.82 | 0.87 |
| M_149 | 0.96 | 0.96 | 0.95 | 0.91 | 0.89 | 0.91 | 0.82 | 0.77 | 0.83 |
| M_150 | 0.51 | 0.97 | 0.96 | 0.73 | 0.91 | 0.90 | 0.98 | 0.80 | 0.81 |
| M_151 | 0.44 | 0.96 | 0.96 | 0.65 | 0.92 | 0.91 | 0.92 | 0.81 | 0.79 |
| M_152 | 0.47 | 0.96 | 0.97 | 0.68 | 0.91 | 0.92 | 0.95 | 0.81 | 0.81 |
| M_153 | 0.96 | 0.98 | 0.95 | 0.91 | 0.95 | 0.90 | 0.81 | 0.82 | 0.81 |
| M_154 | 0.46 | 0.95 | 0.95 | 0.64 | 0.91 | 0.95 | 0.95 | 0.79 | 0.92 |
| M_155 | 0.61 | 0.96 | 0.97 | 0.86 | 0.93 | 0.88 | 0.96 | 0.86 | 0.80 |
| M_156 | 0.45 | 0.97 | 0.95 | 0.65 | 0.93 | 0.91 | 0.94 | 0.82 | 0.81 |
| M_157 | 0.63 | 0.90 | 0.96 | 0.87 | 0.83 | 0.92 | 0.93 | 0.73 | 0.84 |
| M_158 | 0.45 | 0.97 | 0.96 | 0.73 | 0.96 | 0.91 | 0.98 | 0.86 | 0.83 |
| M_159 | 0.35 | 1.00 | 0.92 | 0.55 | 0.98 | 0.88 | 0.85 | 0.89 | 0.74 |
| M_160 | 0.21 | 0.98 | 0.97 | 0.35 | 0.92 | 0.94 | 0.63 | 0.83 | 0.85 |
| M_161 | 0.21 | 0.94 | 0.95 | 0.35 | 0.90 | 0.93 | 0.57 | 0.84 | 0.84 |
| M_162 | 0.50 | 0.96 | 0.96 | 0.71 | 0.91 | 0.91 | 0.97 | 0.80 | 0.82 |
| M_163 | 0.04 | 1.00 | 0.98 | 0.26 | 0.96 | 0.95 | 0.86 | 0.78 | 0.89 |
| M_164 | 0.47 | 0.97 | 0.96 | 0.78 | 0.92 | 0.92 | 0.96 | 0.83 | 0.83 |
| M_165 | 0.13 | 0.99 | 0.97 | 0.27 | 0.94 | 0.95 | 0.58 | 0.84 | 0.88 |
| M_166 | 0.18 | 0.97 | 0.98 | 0.40 | 0.94 | 0.94 | 0.69 | 0.87 | 0.86 |
| M_167 | 0.06 | 0.98 | 0.97 | 0.18 | 0.94 | 0.95 | 0.38 | 0.87 | 0.92 |
| M_168 | 0.14 | 1.00 | 0.94 | 0.30 | 0.95 | 0.90 | 0.60 | 0.81 | 0.80 |
| M_169 | 0.97 | 0.98 | 0.96 | 0.97 | 0.96 | 0.92 | 0.86 | 0.86 | 0.84 |
| M_170 | 0.90 | 0.95 | 0.94 | 0.99 | 0.91 | 0.89 | 0.83 | 0.81 | 0.78 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| M_171 | 0.44 | 0.98 | 0.96 | 0.70 | 0.90 | 0.89 | 0.97 | 0.83 | 0.81 |
| M_172 | 0.55 | 0.97 | 0.95 | 0.75 | 0.91 | 0.89 | 0.99 | 0.84 | 0.77 |
| M_173 | 0.40 | 0.92 | 0.98 | 0.67 | 0.86 | 0.92 | 0.97 | 0.75 | 0.85 |
| M_174 | 0.81 | 0.97 | 1.00 | 0.93 | 0.92 | 0.94 | 0.88 | 0.81 | 0.83 |

**Figure S1.** Example figure of result table from conformal prediction.

**Molecule Depiction**



Mol

**Molecule properties:**

| Property | Value |
|---|---|
| Num Aliphatic Carbocycles | 0 |
| Num Heteroatoms | 4 |
| Num Hydrogen Bond Acceptor | 4 |
| Num Aromatic Carbocycles | 2 |
| LogP | 4.444 |
| Num Rotatable Bonds | 5 |
| Num Aromatic Heterocycles | 0 |
| Num Lipinski HB Donor | 0 |
| Num Atoms | 23 |
| Num Lipinski HB Acceptor | 3 |
| Num Rings | 3 |

**Potential Toxicity:**

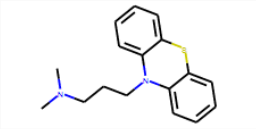
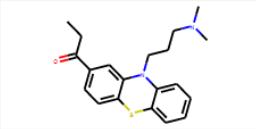
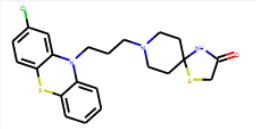
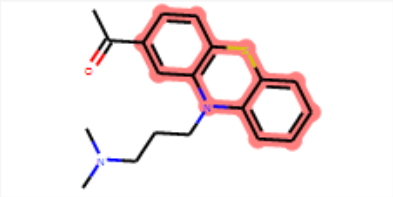
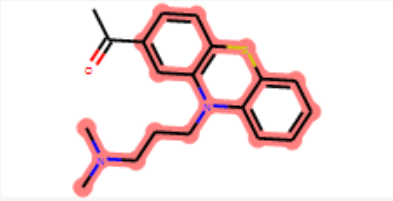
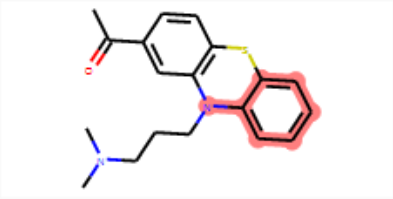
| Index | Action | Injury | Confidence | Similar active compound in database |
|---|---|---|---|---|
| 000 | Modulator of Serotonin 2b (5-HT2b) receptor | Heart, Nervous system, Gastrointestinal | 0.995 |  |
| 001 | Antagonist of the retinoic acid receptor (RAR) signaling pathway | Immune | 0.981 |  |
| 002 | Disruptors of the mitochondrial membrane potential | Liver | 0.928 |  |

**Figure S2.** Example figure of result table from toxic substructure analysis.

**Toxic substructure:**

| Index | Alert Name | Structure |
|---|---|---|
| 000 | Hepatotoxicity |  |
| 001 | Hepatotoxicity |  |
| 002 | Alert from Top200 Drug |  |