Pred-CYP2D™

iS Tox Platform™
Machine-learning models for a new toxicology

Report

Client:
Altox

Username:
Tiago

Study Number:
PredCYP2D_Compound7_

Date:
2019/06/26 - 17:59:29

Program Version: 2.0

Molecular Query

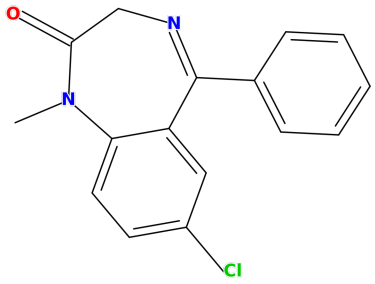
Name:
Compound 7

CAS:
NA

SMILES:
CN1c2ccc(Cl)cc2C(=NCC1=O)c1ccccc1

logK$_{ow}$:
3.15

logD:
2.75



Model Summary

Pred-CYP2D™ s a computational tool developed by combining knowledge-based and artificial intelligence models for prediction of the Molecular Metabolism Network (MMN), in this framework composed of Metabolism Reactions, Metabolites Structures, Metabolic Pathways, and Metabolic Inhibition.

In the Model 1 (Metabolism Reaction and Metabolite Structures) are predicted relevant metabolism reactions, the rate of metabolite formation and metabolite structures. Each predicted metabolite fires a search on our metabolite and bioassay databases, in order to identify correspondence with relevant tested molecules. The Model 2 (Metabolic Pathways) predicts if a molecule can be a substrate or not for relevant CYPs. The Model 3 (Metabolic Inhibition) predicts inhibition of specific cytochrome P450 enzymes.

Our metabolite database contains approximately 173,667 structures with annotations of biologically relevant metabolites and drug substances, including human (phase 1 and 2) and non-human metabolites.

The bioassay database contains 292,095 compounds categorized in approved drug substance, molecules that have been investigated in clinical trials, molecules not approved after clinical trials, active molecules in *in*

vitro assays at 10uM and molecules tested in *in vivo* assays.

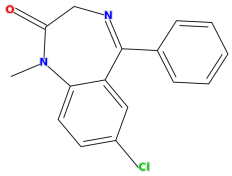
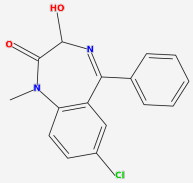
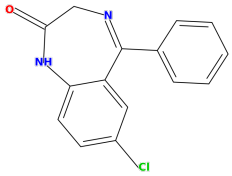
Cytochrome P-450 enzymes represent a large and diverse protein family with important functions in the metabolism of molecules.

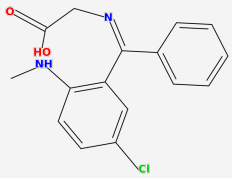
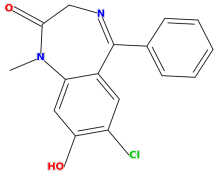
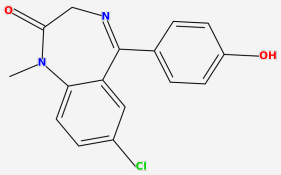
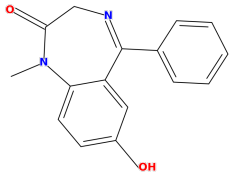
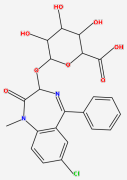
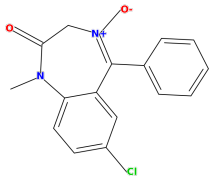
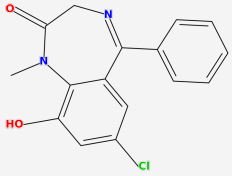
Inhibition and metabolism models were constructed for nine of the most popular enzymes from the CYP superfamily in human liver with the largest public dataset (~15,000 compounds). The nine enzymes for this study, namely CYP2D6, CYP2B6, CYP1A2, CYP2C8, CYP2C19, CYP2E1, CYP2A6, CYP2C9 and CYP3A4 account for 90% of the xenobiotic, drug metabolism and drug-drug interaction (DDI) in the human body.

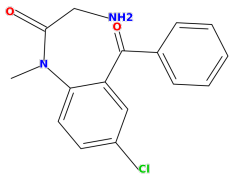
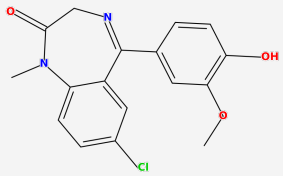
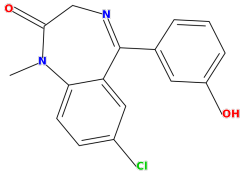
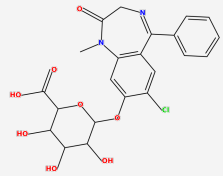
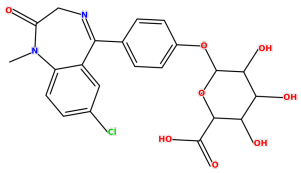
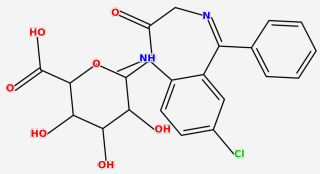
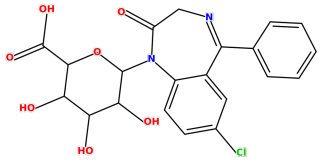
Results - Model 1 (Metabolism Reactions and Metabolite Structures)

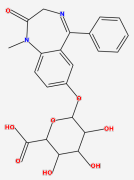
The predictions of the metabolite structures, metabolism reactions and rate of metabolite formation (score) are presented in the table below. Each predicted metabolite fires a search on our metabolite and bioassay databases, in order to identify correspondence with relevant tested molecules.

The knowledge-based predictions of the potential metabolites are based on a broad range of phase 1 and phase 2 reactions of the metabolism that occurs in humans for specific labile fragments (fragment-based). An empirical probability score is assigned to each rule in the predicted metabolites, using to refine and rank the predicted metabolites.

Predicted Metabolite (Name and/or SMILES)	Predicted metabolism reaction	Formation score	Data retrieval - Metabolite ocurrence	Data Retrieval - Bioassays
 (Diazepam) CN1C(=O)CN=C(c2ccccc2)c2cc(Cl)ccc21	-	-	Human endogenous metabolite	Approved Drugs, Compounds that have been investigated (clinical trials)
 (Temazepam) CN1C(=O)C(O)N=C(c2ccccc2)c2cc(Cl)ccc21	Aliphatic Hydroxylation	42.1 %	Human endogenous metabolite	Approved Drugs, Compounds that have been investigated (clinical trials)
 O=C1CN=C(c2ccccc2)c2cc(Cl)ccc2N1	N-Demethylation	41.8 %	-	-

Predicted Metabolite (Name and/or SMILES)	Predicted metabolism reaction	Formation score	Data retrieval - Metabolite ocurrence	Data Retrieval - Bioassays
 CNc1ccc(Cl)cc1C(=NCC(=O)O)c1ccccc1	Hydrolysis	9.6 %	-	-
 CN1C(=O)CN=C(c2ccccc2)c2cc(Cl)c(O)cc21	Aromatic Hydroxylation	6.1 %	-	-
 CN1C(=O)CN=C(c2ccc(O)cc2)c2cc(Cl)ccc21	Aromatic Hydroxylation	6.1 %	-	-
 CN1C(=O)CN=C(c2ccccc2)c2cc(O)ccc21	Aromatic Dechlorination	4.5 %	-	-
 () CN1C(=O)C(OC2OC(C(=O)O)C(O)C(O)C2O)N=C(c2ccccc2)c2cc(Cl)ccc21	Aliphatic Hydroxylation	4.3 %	-	*in vitro* active at 10 uM
 CN1C(=O)C[N+]([O-])=C(c2ccccc2)c2cc(Cl)ccc21	N-Oxidation	3.6 %	-	-
 CN1C(=O)CN=C(c2ccccc2)c2cc(Cl)cc(O)c21	Aromatic Hydroxylation	3.0 %	-	-

Predicted Metabolite (Name and/or SMILES)	Predicted metabolism reaction	Formation score	Data retrieval - Metabolite ocurrence	Data Retrieval - Bioassays
 CN(C(=O)CN)c1ccc(Cl)cc1C(=O)c1ccccc1	Imine Hydrolysis	2.7 %	-	-
 COc1cc(C2=NCC(=O)N(C)c3ccc(Cl)cc32)ccc1O	Aromatic Oxidation	2.1 %	-	-
 CN1C(=O)CN=C(c2cccc(O)c2)c2cc(Cl)ccc21	Aromatic Hydroxylation	1.6 %	-	-
 CN1C(=O)CN=C(c2ccccc2)c2cc(Cl)c(OC3OC(C(=O)O)C(O)C(O)C3O)cc21	Aromatic Hydroxylation	1.5 %	-	-
 CN1C(=O)CN=C(c2ccc(OC3OC(C(=O)O)C(O)C(O)C3O)cc2)c2cc(Cl)ccc21	Aromatic Hydroxylation	1.5 %	-	-
 CNc1ccc(Cl)cc1C(=NCC(=O)OC1OC(C(=O)O)C(O)C(O)C1O)c1ccccc1	Hydrolysis	1.4 %	-	-
 O=C(O)C1OC(N2C(=O)CN=C(c3ccccc3)c3cc(Cl)ccc32)C(O)C(O)C1O	N-Demethylation	1.2 %	-	-

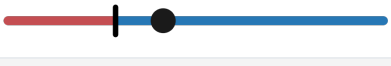
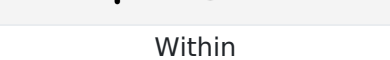
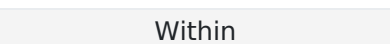
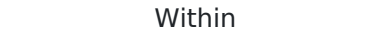
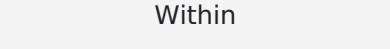
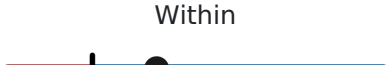
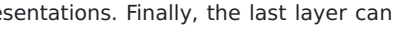
Predicted Metabolite (Name and/or SMILES)	Predicted metabolism reaction	Formation score	Data retrieval - Metabolite ocurrence	Data Retrieval - Bioassays
 CN1C(=O)CN=C(c2ccccc2)c2cc(OC3OC(C(=O)O)C(O)C(O)C3O)ccc21	Aromatic Dechlorination O-Glucuronidation	1.1 %	-	-

The Expert Knowledge algorithm is based on a broad range of metabolic phase 1 reactions reported in the human Metabolite Database. The "Formation score" is an empirical probability score assigned to each rule representing the fraction of correctly predicted metabolites.

Results - Model 2 (Metabolic Pathways)

The predicted metabolic pathways, with an analysis for the molecule as being a substrate or not for relevant CYPs, are presented in the table below.

The individual predictions were obtained for the eight fractions of Cytochrome P450 (CYPs) metabolism by Artificial Neural Networks (ANN). The ANN assign a category non-substrate (-) or substrate (+) with confidence levels (OECD Principle 2) and Applicability Domain (OECD Principle 2).

CYP	Predicted category	Confidence	Applicability Domain
CYP2D6	Substrate (+)	78.9%	Within 
CYP2B6	Substrate (+)	85.1%	Within 
CYP1A2	Substrate (+)	79.4%	Within 
CYP3A4	Substrate (+)	89.5%	Within 
CYP2C19	Substrate (+)	92.7%	Within 
CYP2E1	Non-substrate (-)	86.4%	Within 
CYP2A6	Non-substrate (-)	82.1%	Within 
CYP2C9	Substrate (+)	91.5%	Within 
CYP2C8	Substrate (+)	88.5%	Within 

The deep learning model has components with multiple hidden layers that can learn increasingly abstract representations of the chemical hybrid descriptors. Then, every 4 subsequent layers learn more complex representations. Finally, the last layer can evaluate the Cytochrome P450 (CYPs) inhibition of the given compound.

The applicability domain (AD) is the chemical and toxicological space encoded by the developed model, in which they operate to make new predictions with a given reliability (a defined domain of applicability - OECD Principle 3).

The visual AD inspection is represented by a bar graph of the average fingerprint-dice similarity for the k-nearest neighbors of each compound during the 5-Fold external validation. The chemical structure is represented by a hybrid descriptor composed by ECFP6 fingerprint and physicochemical measurements: MW, TPSA, octanol-water partition coefficient for neutral compounds (logK) or at different pH states (logD). At the visual AD inspection, the black circle represents the evaluated compound, the highlighted red area means the restricted similarity region, and the blue region is the allowed similarity of the chemical space to predict new compounds.

New predictions must be reasonably similar to training set compounds or a prediction cannot be accepted.

Results - Model 3 (Metabolic Inhibition)

The predictions for inhibition of specific cytochrome P450 enzymes, with an analysis for the molecule as being a potential inhibitor or not for relevant CYPs, are presented in the table below.

The individual predictions were obtained for the 8 fractions of Cytochrome P450 by Artificial Neural Networks (ANN). The ANN assign a category non-inhibitor (-) or inhibitor (+) with confidence levels (OECD Principle 2) and Applicability Domain (OECD Principle 2).

CYP	Predicted category	Confidence	Applicability Domain
CYP2B6	Inhibitor (+)	82.7%	Within 
CYP2C19	Non-inhibitor (-)	73.0%	Within 
CYP2E1	Non-inhibitor (-)	75.6%	Within 
CYP2D6	Non-inhibitor (-)	81.5%	Within 
CYP2C8	Inhibitor (+)	82.8%	Outside 
CYP1A2	Non-inhibitor (-)	77.7%	Within 
CYP3A4	Non-inhibitor (-)	91.2%	Within 
CYP2C9	Inhibitor (+)	78.8%	Within 

The deep learning model has components with multiple hidden layers that can learn increasingly abstract representations of the chemical hybrid descriptors. Then, every 4 subsequent layers learn more complex representations. Finally, the last layer can evaluate the Cytochrome P450 (CYPs) inhibition of the given compound.

The applicability domain (AD) is the chemical and toxicological space encoded by the developed model, in which they operate to make new predictions with a given reliability (a defined domain of applicability - OECD Principle 3).

The visual AD inspection is represented by a bar graph of the average fingerprint-dice similarity for the k-nearest neighbors of each compound during the 5-Fold external validation. The chemical structure is represented by a hybrid descriptor composed by ECFP6 fingerprint and physicochemical measurements: MW, TPSA, octanol-water partition coefficient for neutral compounds (logK) or at different pH states (logD). At the visual AD inspection, the black circle represents the evaluated compound, the highlighted red area means the restricted similarity region, and the blue region is the allowed similarity of the chemical space to predict new compounds.

New predictions must be reasonably similar to training set compounds or a prediction cannot be accepted.