Pred-Oral™

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

Report

Client:
Altox

Username:
Tiago

Study Number:
Pred-Oral_Compound6_

Date:
2019/06/26 - 17:58:26

Program Version: 1.7

Molecular Query

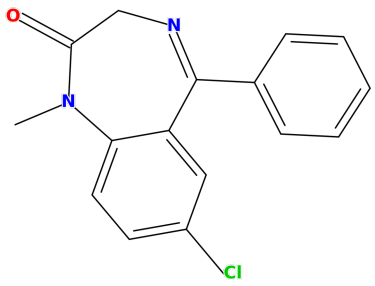
Name:
Compound 6

CAS:
NA

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

logK$_{ow}$:
3.15

logD:
2.75



Model Summary

Pred-Oral™ is a computational tool for prediction of oral absorption through the permeability by statistical and machine learning models. Human Intestinal Absorption (HIA) are able to predict the absorption of drugs by the human intestine and the permeability through human colon adenocarcinom (Caco-2) and Parallel artificial membrane permeability assay (PAMPA). The models were validated following the OECD (Organisation for Economic Co-operation and Development) Principles for the Validation for Regulatory Purposes of (Q)SAR Models. These OECD principles are discussed in each section of this report. The fastest method for evaluating the drug-like properties of a compound is to apply Lipinski's rule of five (RO5). The RO5 is a rule of thumb to evaluate druglikeness for a chemical compound to have a higher probability of being well absorbed by humans after oral dosing.

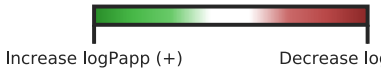
This tool is powered by artificial intelligence models for Human Intestinal Absorption (HIA) evaluation, using 1803 and 709 rigorously curated data of the compounds assayed for Caco-2 Cell Line human and Parallel artificial membrane permeability assay (PAMPA) (defined endpoint - OECD principle 1). The final result is provided in a summarising table with all the individual predictions, the applicability domain (AD), and relevant data for each endpoint/method.

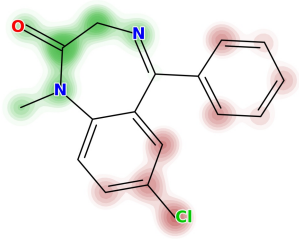
Artificial Intelligence Models

The individual predictions below were obtained for the human colon adenocarcinom (Caco-2) by Artificial Neural Networks (ANN). To ensure transparency in the description of the model (an unambiguous algorithm - OECD Principle 2), more detailed information about each model is presented below.

The QSAR (Quantitative Structure–activity relationship) Probability Mapping indicates the fragments that decrease logPapp (-) (red) or Increase logPapp (+) (green), useful for hypotheses and mechanistic interpretations (OECD Principle 5).

SAR Contribution Map



Predicted endpoint/Method	Predicted Category/Value (Confidence)	SAR Contribution Mapping
Caco-2 (Papp) Deep Learning decision model implemented with hybrid descriptors	High Permeability -4.3 log10 52.2 x 10^{-6} cm/s (96.0%)	

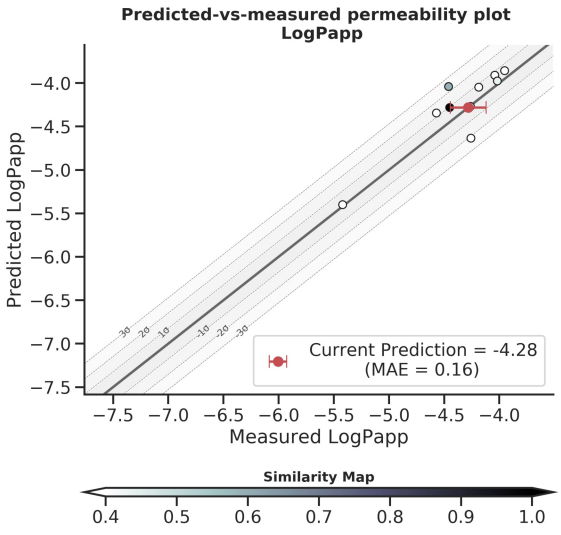
Prediction Confidence (Based on the most similar molecules)

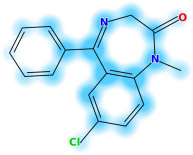
With appropriate measures of goodness-of-fit, robustness, and predictivity (OECD Principle 4). The graph below plots measured versus predicted Caco-2 (Papp) values for each most similar compound and, the mean absolute error (MAE) for each prediction.

Experimental vs predicted permeability values

for the six most similar molecules (Dice) with the prediction confidence level

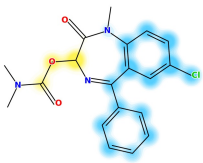
Similarity Map





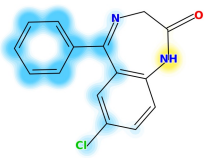
Similarity: 1.00

Exp. value: -4.4
Ext. Pred. value: -4.3
(Confidence: 96%)



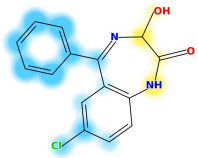
Similarity: 0.59

Exp. value: -4.5
Ext. Pred. value: -4.0
(Confidence: 90%)



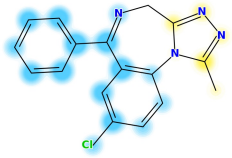
Similarity: 0.44

Exp. value: -4.0
Ext. Pred. value: -4.0
(Confidence: 99%)



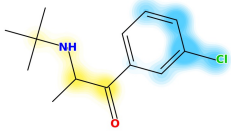
Similarity: 0.38

Exp. value: -4.0
Ext. Pred. value: -3.9
(Confidence: 97%)



Similarity: 0.37

Exp. value: -4.3
Ext. Pred. value: -4.6
(Confidence: 91%)



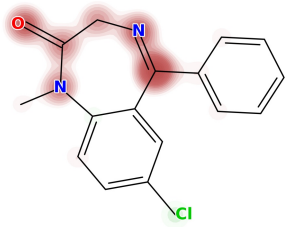
Similarity: 0.33

Exp. value: -4.0
Ext. Pred. value: -3.9
(Confidence: 98%)

The individual predictions below were obtained for the human colon adenocarcinom (Caco-2) by Artificial Neural Networks (ANN). To ensure transparency in the description of the model (an unambiguous algorithm - OECD Principle 2), more detailed information about each model is presented below.

The QSAR (Quantitative Structure–activity relationship) Probability Mapping indicates the fragments that decrease logPAMPA (+) (green) or Increase logPAMPA (-) (red), useful for hypotheses and mechanistic interpretations (OECD Principle 5).

SAR Contribution Map

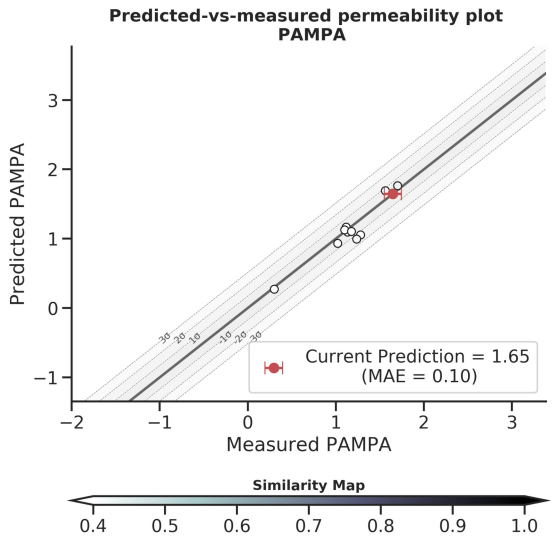
Predicted endpoint/Method	Predicted Category/Value (Confidence)	SAR Contribution Mapping
Parallel Artificial Membrane Permeability (PAMPA) Deep Learning decision model implemented with hybrid descriptors	High Permeability 1.6 log10 4.4 x 10^{-6} cm/s (92.0%)	

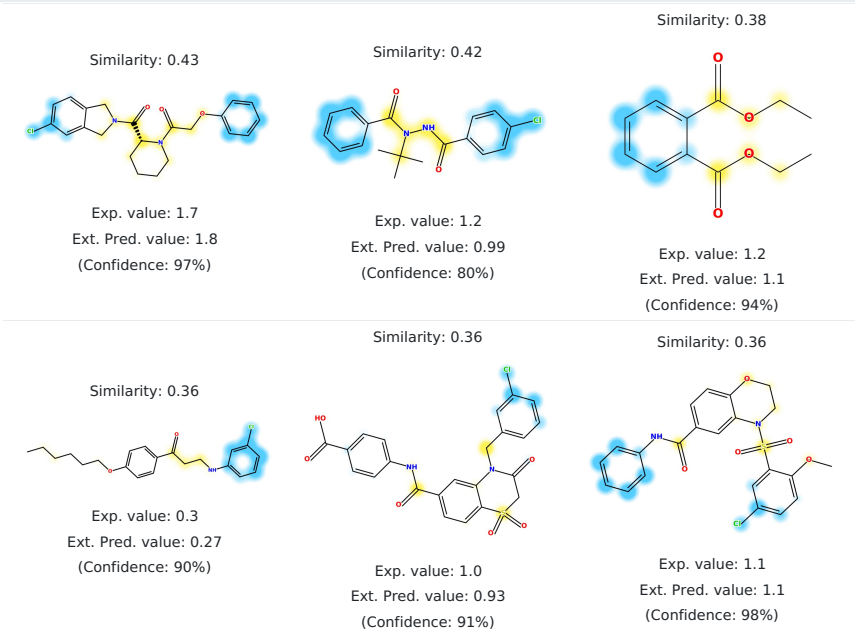
Prediction Confidence (Based on the most similar molecules)

With appropriate measures of goodness-of-fit, robustness, and predictivity (OECD Principle 4). The graph below plots measured versus predicted PAMPA values for each most similar compound and, the mean absolute error (MAE) for each prediction.

Experimental vs predicted permeability values

for the six most similar molecules (Dice) with the prediction confidence level





To assess the confidence of the Pred-Oral ™ predictions, after of to assess the applicability domain of the

model by the Visual applicability domain (AD) Inspection®, is recommended assess the model performance by the predicted-vs-measured permeability plot and confidence level for the most similar molecules.

To assess the chemical space of the prediction and its mean absolute error (MAE), this section presents the model performance for the most similar molecules (Dice) during 5-Fold external validation, also, for the six most similar molecules (Dice) with the confidence level side-by-side to each Human Intestinal Absorption (HIA) prediction. If the predicted value matches the experimental values for the test set chemicals, the model has a low MAE and greater confidence.

Detailed data about the dataset of chemicals, including endpoint and descriptor values; derivation of the descriptors; test and training sets; removed outliers; statistical parameters and others are available in the QMRF (QSAR Model Reporting Format) report under a confidentiality agreement. Altox Ltda assures scientific integrity of the data.

The Deep learning model has components with multiple hidden layers that can learn increasingly abstract representations of the chemical hybrid descriptors. The chemical hybrid descriptor. Then, every 4 subsequent layers learn more complex representations. Finally, the last layer can evaluate Human Intestinal Absorption (HIA) values of the given compound.

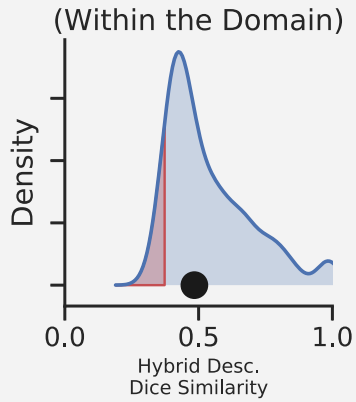
Visual AD Inspection®

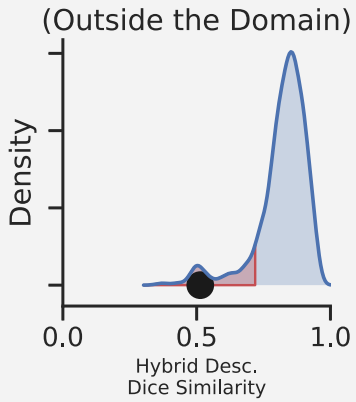
The applicability domain (AD) is defined by the chemical structure space and the permeability response encoded by the developed model, to make new predictions with a given reliability (a defined domain of applicability - OECD Principle 3). Our visual AS Inspection® is used to establish the scope and limitations of the models. Basically, new chemicals must be reasonably similar to training set compounds or a prediction cannot be accepted.

Our visual AD inspection is represented by a density plot 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 ECFP4 fingerprints and physicochemical measurements: molecular weight (MW), topological polar surface area (TPSA), octanol-water partition coefficient for neutral compounds (log K_{ow}) 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.

Even though a well-designed AD helps the user to assess the reliability of predictions made by the model, it should not automatically be assumed that all predictions within the defined AD are necessarily reliable.

Caco-2 (Papp)	Parallel Artificial Membrane Permeability (PAMPA)





Final Result

The results are presented below; low permeability (red), or moderate permeability (yellow), or high permeability (green) predictions are presented with applicability domain (AD) and confidence level (robustness, OECD principle 4) for the physical-chemical properties and advanced statistical systems.

Physical-chemical Properties	Value
LogK$_{ow}$	3.15
LogD	2.75

Molecular Properties	Result	Rules Failed
Rule of Five	**Passed**	-

Advanced Statistical System (Predicted endpoint/Method)	**Category/Value (Confidence)**	**Applicability Domain**
Caco-2 (Papp) Deep Learning decision model implemented with hybrid descriptors	High Permeability -4.3 log10 52.2 x 10^{-6} cm/s (96.0%)	Within 
Parallel Artificial Membrane Permeability (PAMPA) Deep Learning decision model implemented with hybrid descriptors	High Permeability 1.6 log10 4.4 x 10^{-6} cm/s (92.0%)	Outside