Pred-Ecotox™

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

Report

Client:
Altox

Username:
Tiago

Study Number:
Pred-Ecotox_Compound10_

Date:
2019/06/19 - 15:31:19

Program Version: 1.7

Molecular Query

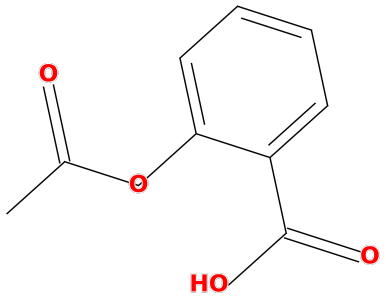
Name:
Compound 10

CAS:
NA

SMILES:
CC(=O)Oc1ccccc1C(=O)O

logK$_{ow}$:
1.31

logD:
-1.35



Model Summary

Pred-Ecotox™ is a computational tool for prediction of aquatic acute toxicity of chemicals for fish (96 hours Letal Concentration - LC_{50}), Daphnia magna (48 hours Letal Concentration - LC_{50}) and Tetrahymena pyriformis (40 hours Inhibition Growth Concentration - IGC_{50}) by Mode of Action (MOA) classification, statistical and machine learning-models, validated following the OECD[1] (Organisation for Economic Co-operation and Development) Principles for the Validation for Regulatory Purposes of (Q)SAR Models. These OECD[1] principles are discussed in each section of this report.

These different models use *in vivo* datasets with with fish 96 hour LC_{50}, Daphnia magna 48 hour LC_{50} and Tetrahymena pyriformis 40 hour IGC_{50}, containing 1,549, 649 and 1,129 structures rigorously curated, respectively, for each acute ecotoxicological prediction (defined OECD Principle 1)[1]. Additionally, based on the compound fragments the Mode of Action (MOA) classification is used and provides a useful mechanistic understanding. The final results are provided in graphical plot summarising all the individual predictions, the applicability domain (AD), and relevant data for each endpoint in a visual way.

Mode of Action (MOA) Class Prediction

The result below is an in silico classification for early assessment of the mode of toxic action (MOA) for the evaluated compound. MOA is important in assessing chemicals because it represents an intermediate level of complexity between molecular mechanisms and physiological or organismal outcomes and provides an organizing scheme for chemical classification.[1]

Our Deep Learning model uses a hybrid descriptor combining physical-chemical properties and structural fingerprints for the biggest dataset available for MOA classification.

Five broad groups of the Verhaar scheme and a specific MOA based on pharmacological/toxicological mechanism are used employing the most frequently used in classification schemes.

The possible MOA classifications are MOA-1 for inert chemicals with narcotic properties, MOA-2 for inert chemicals showing polar narcosis, MOA-3 for reactive chemicals which reactive unselectively with biomolecules, and MOA-4 chemicals acting by a specific mechanism. MOA-5 is an unspecific class with unknown mode of action.

This classification provides toxicological mechanistic insights based on the available knowledge (OECD Principle 5).[1]

Predicted MOA Class (Probability %)	MOA Broad (Probability %)	MOA specific categories
MOA-3 (100%)	**Narcosis (100%)**	**General**
		Ester
	Reactivity (0%)	**Alkylation**
		General
		Acrylate
		Di/trinitroaromatic
		Carbonyl
		Other
		Cyanate/nitrile
		Chromate
		Hydrazine
		Phosphide
		Ester

**It is recommendable to assess if the MOA or protein target is relevant to fishes, arthropod and/or protozoans.*

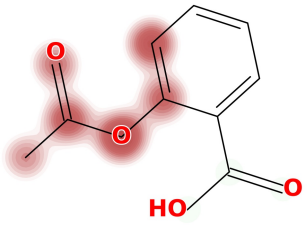
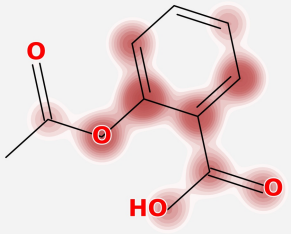
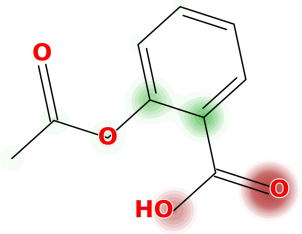
Artificial Intelligence Models

The individual predictions below were obtained for three acute ecotoxicological endpoints by Artificial Neural Networks (ANN). The Neural Network Predictor outputs fish 96 hours LC_{50}, Daphnia magna 48 hours LC_{50} and Tetrahymena pyriformis 40 hours IGC_{50} predictions using deep learning (ANN).

To ensure transparency in the description of the model (an unambiguous algorithm - OECD Principle 2),[1] more detailed information about each model is presented below. The STR (Structure–toxicity relationship) Probability Mapping indicates the fragments more related to the absence of toxicity (green) or presence (red) as well, useful for hypotheses and mechanistic interpretations (OECD Principle 5)[1].

STR Contribution Map

Non-Toxic Toxic

Predicted endpoint/Method	Predicted Value (Confidence)	STR Contribution Mapping
IGC_{50} (*Tetrahymena p.*, 40hrs) Deep Learning decision model implemented with hybrid descriptors	3.2 mg/L 17.6 µM (94.0%)	
LC_{50} (*Daphnia magna*, 48hrs) Deep Learning decision model implemented with hybrid descriptors	1.9 mg/L 10.7 µM (90.0%)	
LC_{50} (Fish, 96hrs) Deep Learning decision model implemented with hybrid descriptors	662.0 mg/L 3674.4 µM (89.0%)	

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 of this tool is a machine learning algorithm based on artificial neural network (ANN) or artificial intelligence to calculate fish 96 hours LC_{50}, Daphnia magna 48 hours LC_{50} and Tetrahymena pyriformis 40 hours IGC_{50}.

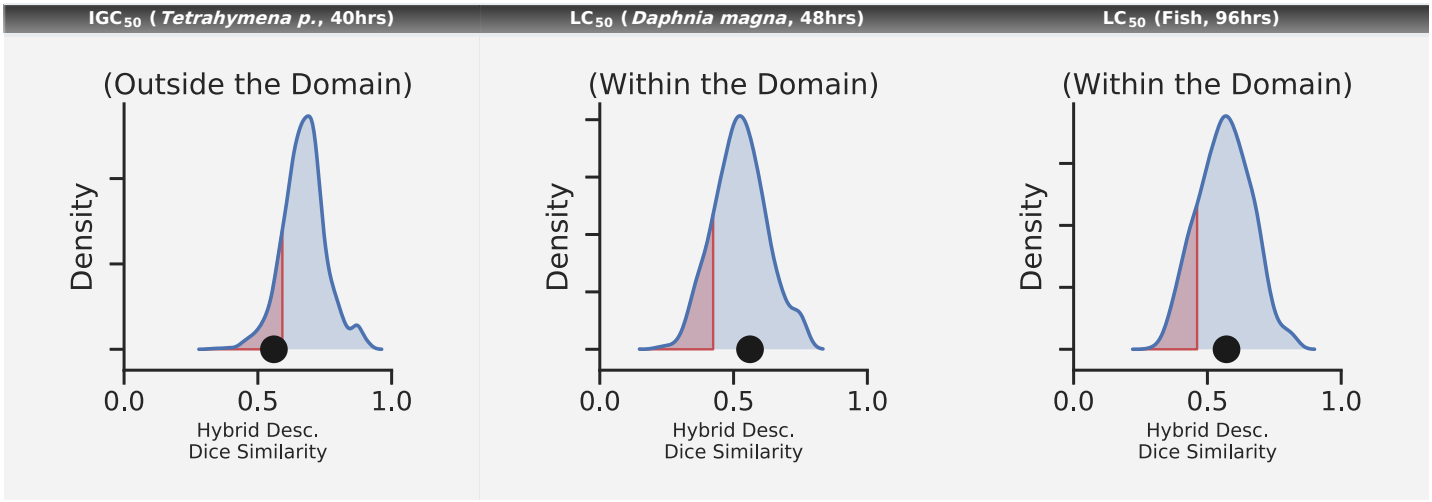
The model has components with multiple hidden layers that can learn increasingly abstract representations of the chemical hybrid descriptors. The chemical hybrid descriptor is composed by ECFP6 fingerprint and physicochemical properties [molecular weight (MW), the topological polar surface area (TPSA), octanol-water partition coefficient for neutral compounds ($logK_{ow}$) or at different pH states (logD)]. Then, every 4 subsequent layers learn more complex representations. Finally, the last layer can evaluate acute toxicity values of the given compound for fish and aquatic invertebrates.

Visual AD Inspection®

The applicability domain (AD) is defined by the chemical structure space and the toxicological response encoded by the developed model, to make new predictions with a given reliability (a defined domain of applicability - OECD Principle 3).[1] 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.

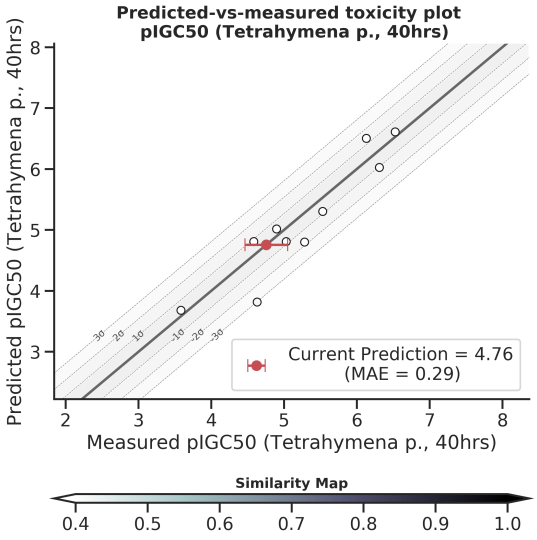


Prediction Confidence (Based on the most similar molecules)

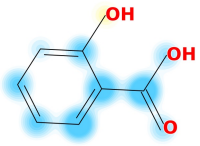
With appropriate measures of goodness-of-fit, robustness, and predictivity - OECD Principle 4, our model involves different strategies to establish the performance of the model. The graph below plots measured versus predicted ecotoxicity values (pIG_{50} and pLC_{50}) for each species and the mean absolute error (MAE) for each prediction.

Experimental vs predicted toxicity values

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

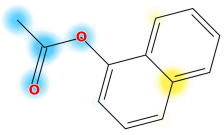


Similarity: 0.42



Exp. value: 4.6
Ext. Pred. value: 3.8
(Confidence: 82%)

Similarity: 0.39

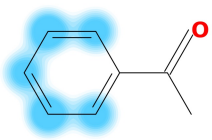


Exp. value: 6.3
Ext. Pred. value: 6.0
(Confidence: 95%)

Similarity: 0.37

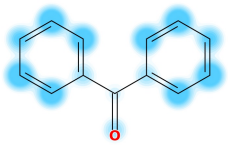
Exp. value: 3.6
Ext. Pred. value: 3.7
(Confidence: 97%)

Similarity: 0.35



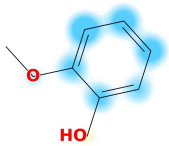
Exp. value: 5.0
Ext. Pred. value: 4.8
(Confidence: 96%)

Similarity: 0.35

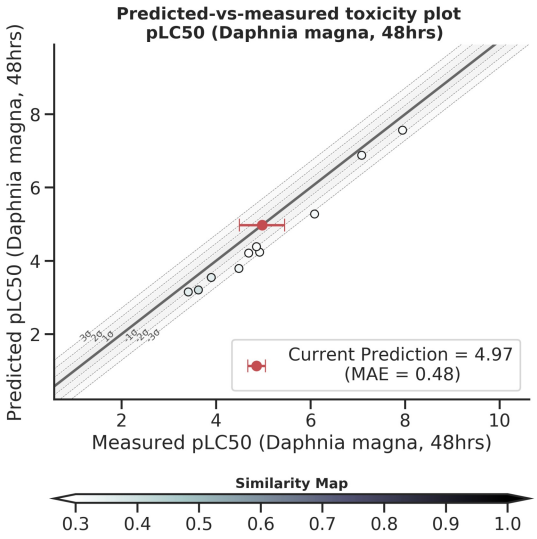


Exp. value: 6.1
Ext. Pred. value: 6.5
(Confidence: 94%)

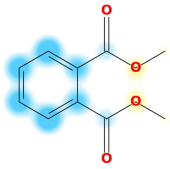
Similarity: 0.33



Exp. value: 4.6
Ext. Pred. value: 4.8
(Confidence: 95%)

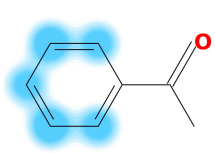


Similarity: 0.39



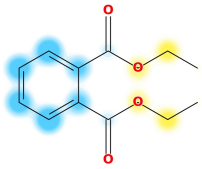
Exp. value: 3.6
Ext. Pred. value: 3.2
(Confidence: 88%)

Similarity: 0.35



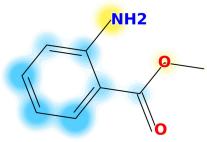
Exp. value: 3.9
Ext. Pred. value: 3.5
(Confidence: 91%)

Similarity: 0.35



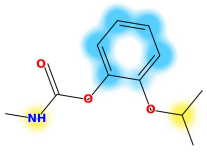
Exp. value: 3.4
Ext. Pred. value: 3.1
(Confidence: 92%)

Similarity: 0.33



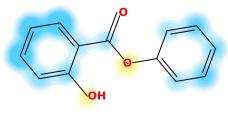
Exp. value: 4.5
Ext. Pred. value: 3.8
(Confidence: 85%)

Similarity: 0.32

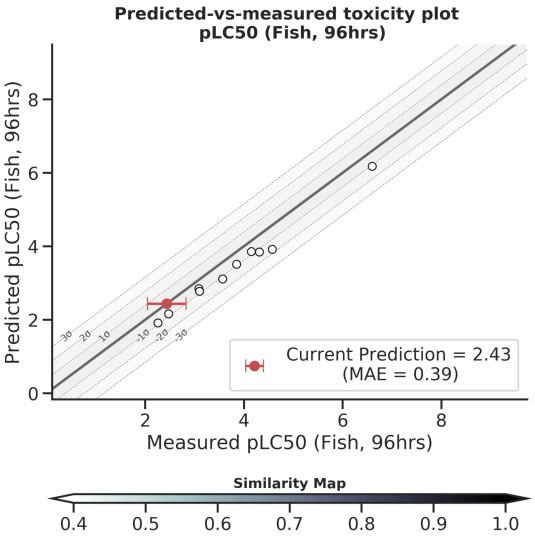


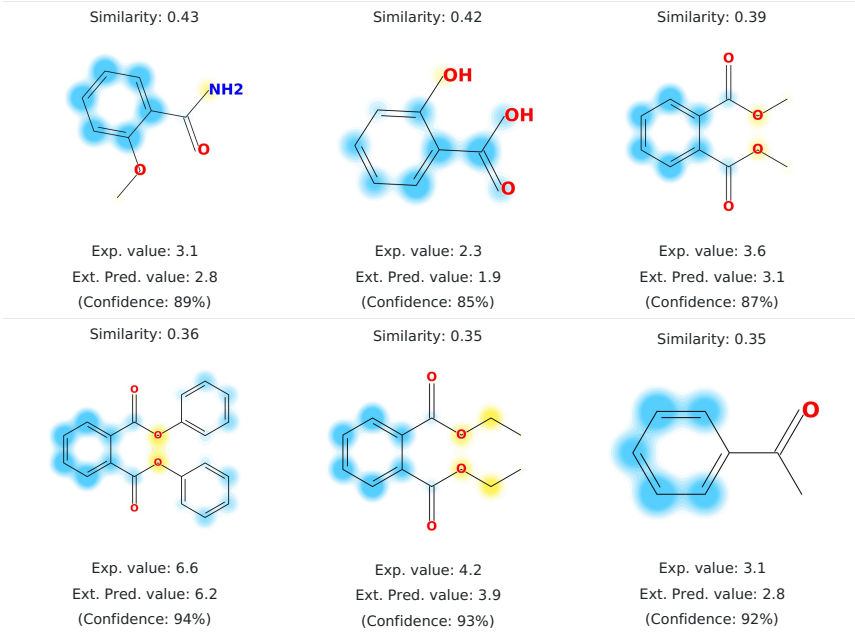
Exp. value: 7.1
Ext. Pred. value: 6.9
(Confidence: 97%)

Similarity: 0.32



Exp. value: 6.1
Ext. Pred. value: 5.3
(Confidence: 87%)





To assess the confidence of the Pred-Ecotox TM 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 toxicity 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 ecotoxicological prediction. If the predicted value matches the experimental values for the test set chemicals, the model has a low MAE and greater confidence.

Results

The results are presented below, individual predictions and applicability domain (AD) for each endpoint.

Physicochemical Properties

Predicted logK_{ow}: 1.31

Predicted logD: -1.35

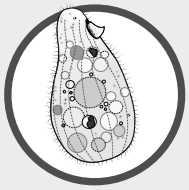
Predicted MOA Class (Probability %)

MOA-3 (100%)

Narcosis (100%)

Predicted aquatic toxicity values

Protozoan



IGC_{50}(40 hrs):	3.2 mg/L 17.6 µM
$pIGC_{50}$(40 hrs):	4.8
Confidence:	94.0%
AD:	Outside

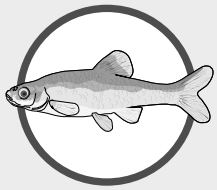
Tetrahymena pyriformis

Crustacean

LC_{50}(48 hrs):	1.9 mg/L 10.7 µM
pLC_{50}(48 hrs):	5.0
Confidence:	90.0%
AD:	Within

Daphnia magna

Fish



LC_{50}(96 hrs):	662.0 mg/L 3674.4 µM
pLC_{50}(96 hrs):	2.4
Confidence:	89.0%
AD:	Within

Danio rerio
Poecilia reticulata
Oryzias latipes
Lepomis macrochirus
Oncorhynchus mykiss
Pimephales promelas

IGC: Inhibition Growth Concentration
LC: Lethal Concentration
p: -log([Molar Concentration])
AD: Applicability Domain

References

1. OECD; Guidance Document On The Validation Of (Quantitative)Structure-Activity 2007. (https://goo.gl/GgSwrj).

2. TR 102 – Intelligent Testing Strategies In Ecotoxicology: Mode Of Action Approach For Specifically Acting Chemicals (https://goo.gl/wgo6Cz)

3. Kienzler, A.; Barron, M. G.; Belanger, S. E.; Beasley, A.; Embry, M. R. Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches. Environ. Sci. Technol. 2017, 51 (17), 10203. DOI:10.1021/acs.est.7b02337

4. Kienzler, A.; Barron, M. G.; Belanger, S. E.; Beasley, A.; Embry, M. R. Response to “Comment on ‘Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches.’” Environ. Sci. Technol. 2017, 51 (22), 13511. DOI:10.1021/acs.est.7b05413