AI-Fate™

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

Report

Client:
Altox Ltda

Username:
Carlos Eduardo Matos dos Santos

Study Number:
AltoxLtdaCompound1AIFate190926201343631

Date:
2019/09/26 - 17:13:43

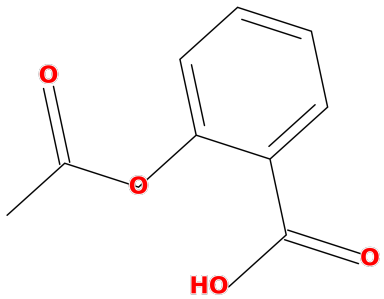
Program Version: 2.1

Molecular Query

Name:
Compound 1

CAS:

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



Model Summary

AI-Fate™ is a computational tool for the prediction of the environmental fate and behavior for estimating the ready biodegradability in an anaerobic aqueous medium (OECD 301C) by statistical and machine learning models. The method permits the estimation of the biochemical oxygen demand (BOD) at 28 days. The models were validated following the Organization for Economic Co-operation and Development (OECD) Principles for the Validation for Regulatory Purposes of (Q)SAR Models. These OECD principles are discussed in each section of this report.

This tool is powered by artificial intelligence models developed with a rigorously curated biochemical oxygen demand (BOD) dataset containing 1,458 chemical structures (defined endpoint, OECD principle 1)[1].

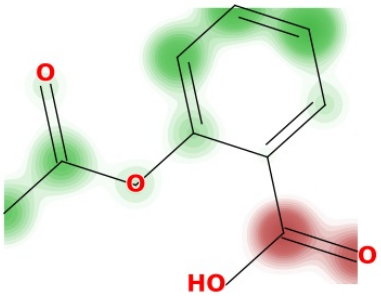
The final result is provided in a table summarizing all of the individual predictions, the applicability domain (AD) confidence levels, with a decision for the outcome.

Artificial Intelligence Models

The individual predictions below were obtained for the biochemical oxygen demand (BOD) by artificial neural networks (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.

STR Contribution Map

Not ready biodegradable (-) Ready biodegradable (+)

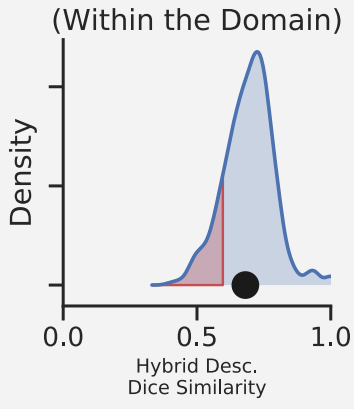
Assay	Prediction class (Confidence)	STR Contribution Mapping
Biochemical oxygen demand (BOD) (OECD 301C) Deep Learning categorical model implemented with hybrid descriptors (ECFP4 fingerprint and physicochemical properties: MW, TPSA, $logK_{ow}$, logD)	**Ready biodegradable (+) 81.1%**	

To assess the confidence of the AI-Fate™ predictions, after ascertaining the applicability domain of the model using the Visual AD Inspection®, assessing the model performance by the predicted-vs-measured toxicity plot and confidence level for the most similar molecules is recommended.

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 AD 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 valid prediction cannot be accepted.

Biochemical oxygen demand (BOD) (OECD 301C)



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 three different types of fingerprints: MACCS, ECFP4, and 3D toxicophoric. At the Visual AD Inspection®, the black circle represents the evaluated compound, the highlighted red area denotes the forbidden similarity region, and the blue region is the allowed similarity chemical space to predict new compounds.

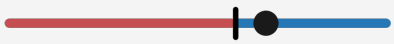
Although a well-designed AD helps the user to assess the reliability of predictions made by the model, it

should not automatically be assumed that all predictions within the defined AD are necessarily reliable.

To assess the chemical space of the prediction, this section presents the model performance for the most similar molecules (Dice) during 5-fold external validation and the six most similar molecules (Dice) with the confidence level side-by-side to each biochemical oxygen demand (BOD) prediction. If the predicted value matches the experimental values for the test set chemicals, the model has a low error and greater confidence.

Final Result

The results are presented below: not ready biodegradable (-) (red) or ready biodegradable (+) (green) predictions are presented with the applicability domain (AD) and confidence level (robustness, OECD principle 4)[1] for the physical-chemical properties and advanced statistical systems.

Assay	Prediction class (Confidence)	Applicability Domain
Advanced Statistical System		
Biochemical oxygen demand (BOD) (OECD 301C) Deep Learning categorical model implemented with hybrid descriptors (ECFP4 fingerprint and physicochemical properties: MW, TPSA, $logK_{ow}$, logD)	**Ready biodegradable (+)** **81.1%**	Within 

Additional Information

Prediction Confidence

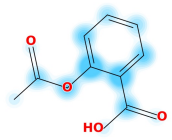
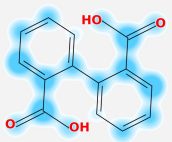
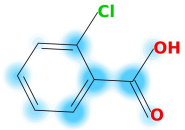
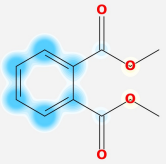
Based on the most similar molecules

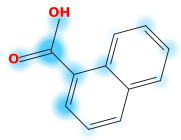
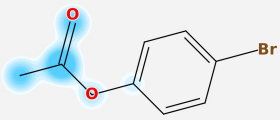
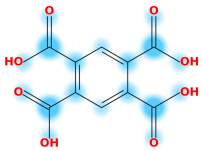
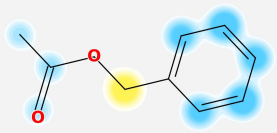
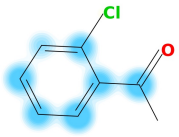
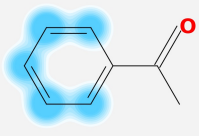
With appropriate measures of goodness-of-fit, robustness, and predictivity (OECD Principle 4)[1], our model used different strategies to establish the performance of the model, which consisted of internal model performance (goodness-of-fit and robustness) and external model performance (predictivity).

To assess the confidence of the AI-Fate ™ predictions, after considering the applicability domain of the model (Visual AD Inspection®), the statistics of the ability to detect known ready biodegradable compounds (sensitivity), not ready biodegradable compounds (specificity), and all molecules in general (concordance) based on the most similar substances are provided below. A map with the similarity level for the 10 most similar molecules is provided with the confidence level for each statistical and artificial intelligence model.

Hybrid descriptor Dice similarity was used to improve the deep learning confidence by interpolating the confidence equalized by the compound similarity criteria obtained from the dataset chemical space. This helps improve the *in silico* toxicological model to reduce the false positive and negative errors.

Performance for the 10-most similar molecules

Molecule (Similarity)	Experimental Data	AI_Fate Prediction (Confidence)
 (1.0)	Not ready biodegradable (-)	Ready biodegradable (+) 81.11%
 (0.69)	Not ready biodegradable (-)	Ready biodegradable (+) 80.96%
 (0.68)	Ready biodegradable (+)	Ready biodegradable (+) 81.42%
 (0.64)	Not ready biodegradable (-)	Ready biodegradable (+) 80.79%

Molecule (Similarity)	Experimental Data	AI_Fate Prediction (Confidence)
 (0.64)	Ready biodegradable (+)	Ready biodegradable (+) 81.7%
 (0.64)	Ready biodegradable (+)	Ready biodegradable (+) 82.37%
 (0.64)	Not ready biodegradable (-)	Ready biodegradable (+) 82.29%
 (0.63)	Not ready biodegradable (-)	Ready biodegradable (+) 81.62%
 (0.63)	Not ready biodegradable (-)	Ready biodegradable (+) 81.26%
 (0.62)	Not ready biodegradable (-)	Ready biodegradable (+) 79.17%

References

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