iS-Ocular™

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

Report

Client:
Altox

Username:
Tiago

Study Number:
iS-Ocular_Compound4_

Date:
2019/06/19 - 15:23:52

Program Version: 1.7

Molecular Query

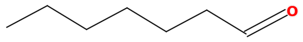
Name:
Compound 4

CAS:
NA

SMILES:
CCCCCCC=O

logK$_{ow}$:
2.16

logD:
3.95



Model Summary

iS-Ocular^{TM} is a computational tool for prediction of eye irritation developed by Altox, using knowledge derived from a combination of mechanisms and experimental data related to eye irritation and corrosion.

The Alert-based model is powered by three types of protein-binding alerts: (i) 122 protein-binding alerts developed by OASIS (general mechanistic), (ii) 184 protein-binding alerts developed OASIS (general mechanistic), and (iii) 300 Glutathione-binding potency alerts obtained from *in chemico* Glutathione depletion assay (GSH).

The Model 2, an artificial intelligence algorithm, predicts molecules that can met the BfR (German Federal Institute for Risk Assessment) rules, for both physicochemical exclusion rules and structure-based inclusion rules, developed and validated by a dataset containing over 2 millions of compounds.

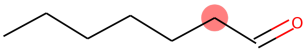
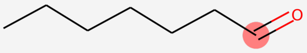
In the Model 3, an artificial intelligence algorithm, were used 5,231 structures with eye irritation and corrosion (human and animal) data rigorously curated. The final result is provided in a summarising table with all the individual predictions, the applicability domain (AD), and relevant data for each endpoint/method.

Model 1 - Structural Alert Analysis

The result below is based on an analysis of fragments assigned to have Protein-binding and Glutathione-binding potency alerts, providing a mechanistic basis for the potential for eye irritation and corrosion (OECD Principle 5).

Result: (+) Positive

Alerts were found in the molecule. The results are in the table below and a description is provided at the end of the report.

Category	Alert ID	Alert
Protein binding by OASIS	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	
Protein binding by OECD	Schiff Base Formers >> Direct Acting Schiff Base Formers >> Mono-carbonyls	
Protein Binding by GHS	-	**Negative (-)**

Model 2 - Eye irritation and corrosion predictions for met BfR Rules

The individual predictions below were obtained by Artificial intelligence model based on the BfR (German Federal Institute for Risk Assessment) rule, for identifying chemical relevant patterns for corrosion/irritation.

Two models which encode both the “inclusion rules” (structural alerts predicting corrosion/irritation potential) and the “exclusion rules” (“IF…THEN NOT…” rules predicting the absence of irritation/corrosion potential) due to certain physicochemical properties.

For classification and labeling, the BfR rule-base provides information that is the closest to the regulatory goal, since the system was designed to predict former EU Risk Phrases for eye irritation (R38) and corrosion (R34, R35) under the EU Dangerous Substance Directive (EU DSD). However, in borderline cases, the prediction may not fully reflect the correct classification under EU CLP (ECHA-17-G-18-EN, 2017).

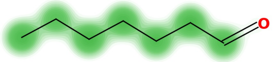
Exclusion rules by BfR	Inclusion rules by BfR	Conclusion
Predicted as not irritating or corrosive to eye based in rules (-)	**Predicted as irritating or corrosive to eye based in rules (+)**	**Inconclusive**

Model 3 - Eye irritation and corrosion predictions for human and animal eye

The individual result in the table below was obtained from an artificial neural network (ANN). The eye Irritation/Corrosion (Human and Animal) ANN model assigns a category positive (+) or negative (-) for eye irritation.

To ensure transparency in the description of the deep learning model (an unambiguous algorithm - OECD Principle 2), more detailed information about each model is presented below. The STR (Structure-Toxicity Relationship) Probability Mapping indicates the fragments more related to the absence/decrease of irritation/corrosion (green) or presence/increase (red) as well, useful for hypotheses and mechanistic interpretations (OECD Principle 5).

STR Contribution Map

Assay/Event	Predicted class (Confidence)	STR Contribution Mapping
Eye Irritation/Corrosion (Human and Animal) Deep Learning categorical model implemented with hybrid descriptors (ECFP6 fingerprint and physicochemical properties: MW, TPSA, logK$_{ow}$, logD)	Negative (-) (68.9%)	

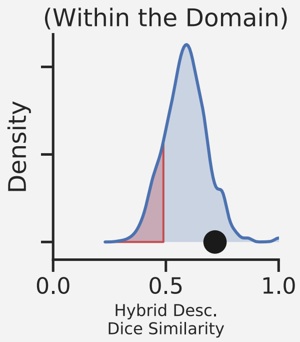
Visual AD Inspection®

The applicability domain (AD) is the chemical and toxicological space encoded by the developed model, in which they operate t o 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 ECFP6 fingerprint 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 represents 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.

Eye Irritation/Corrosion (Human and Animal) - Domain of Applicability



Final Result

The final predictions for the potential for eye irritation by Structural Alerts, Artificial Intelligence (AI) algorithms based in animal and human data and BfR rules (Models 2 and 3) are presented in the table below.

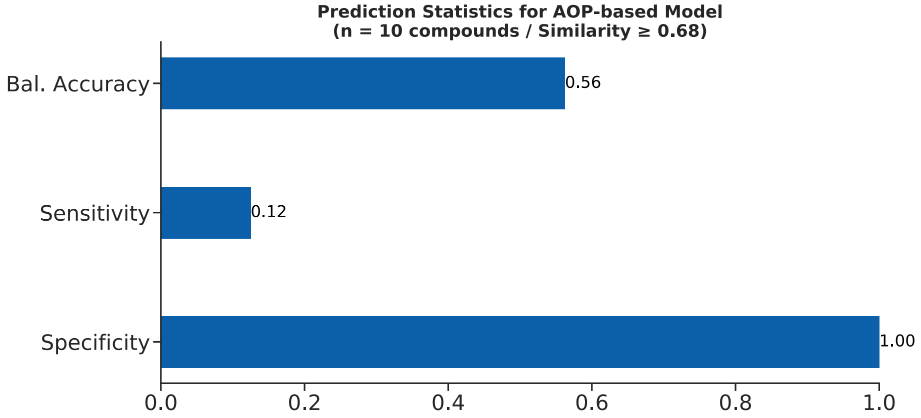
Predicted endpoint/Method	Predicted Class (Confidence)	Applicability Domain
Rule-based Expert System		
Structural Alerts (Protein binding by OASIS, GHS and OECD)	**Positive (+)**	-
Eye irritation/corrosion rules by BfR	**Inconclusive**	-
Advanced Statistical System		
Eye Irritation/Corrosion (Human and Animal) Deep Learning categorical model implemented with hybrid descriptors (ECFP6 fingerprint and physicochemical properties: MW, TPSA, logK_{ow}, logD)	**Negative (-) (68.9%)**	Within 

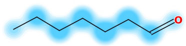
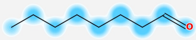
The deep learning model has components with multiple hidden layers that can learn increasingly abstract representations of the chemical hybrid descriptors. The chemical hybrid descriptor is composed by ECFP6 fingerprint and physicochemical properties [molecular weight (MW), the topological polar surface 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 the eye irritation and corrosion of the given compound.

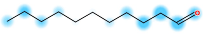
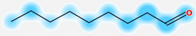
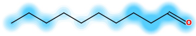
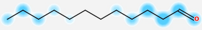
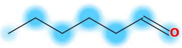
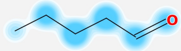
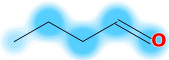
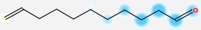
Additional Information - Prediction Confidence based on the most similar molecules

With appropriate measures of goodness-of-fit, robustness, and predictivity (OECD Principle 4), 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 predictions, after of to take into account the applicability domain of the model, additionally, the statistics of the ability to detect known irritating compounds (sensitivity), non-irritating compounds (specificity), and all molecules in general (concordance) based in 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 to improve the in silico toxicological model to reduce the false positive and negative error.



Molecule (Similarity)	Experimental data	Hazard Category	Alerts (Model 1)	Rules (Model 2)	Prediction (Model 3) and confidence level
 (1.0)	**Positive (+)**	Category 2 (ECHA) (+)	Positive (+)	Inconclusive	**Negative (-)** **68.9%**
 (0.94)	**Positive (+)**	Category 2 (ECHA) (+)	Positive (+)	Inconclusive	**Negative (-)** **69.3%**

Molecule (Similarity)	Experimental data	Hazard Category	Alerts (Model 1)	Rules (Model 2)	Prediction (Model 3) and confidence level
 (0.92)	**Negative (-)**	Not classified (ECHA) (-)	Positive (+)	Inconclusive	**Negative (-)** **70.5%**
 (0.92)	**Positive (+)**	Category 2 (ECHA) (+)	Positive (+)	Inconclusive	**Negative (-)** **69.8%**
 (0.92)	**Negative (-)**	Not classified (ECHA) (-)	Positive (+)	Inconclusive	**Negative (-)** **70.2%**
 (0.92)	**Positive (+)**	Category 2 (ECHA) (+)	Positive (+)	Inconclusive	**Negative (-)** **70.8%**
 (0.87)	**Positive (+)**	Category 2 (ECHA) (+)	Positive (+)	Inconclusive	**Negative (-)** **69.3%**
 (0.79)	**Positive (+)**	Category 1 (ECHA) (++)	Positive (+)	Inconclusive	**Positive (+)** **87.0%**
 (0.7)	**Positive (+)**	Category 2 (ECHA) (+)	Positive (+)	Inconclusive	**Negative (-)** **60.2%**
 (0.68)	**Positive (+)**	Category 2 (ECHA) (+)	Positive (+)	Inconclusive	**Negative (-)** **82.3%**