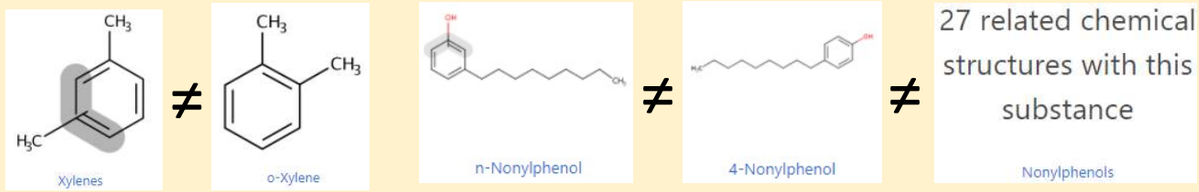
Representation and Enumeration of UVCB Substances to Enable QSAR Predictions for Substances with Structural Uncertainty

Abstract #296

Christopher Grulke[1], Antony J. Williams[1], Brian Meyer[2], Ann Richard[1]
[1]Ctr. for Comput. Toxicol. & Exposure, ORD, U.S. EPA; [2] SEE Program, US EPA
ORCID: 0000-0001-5606-7560

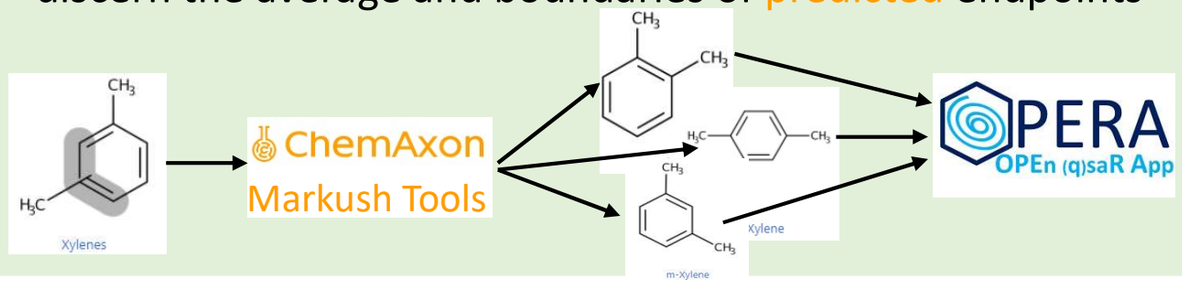
PROBLEM

- UVCB chemicals in public databases are often conflated with well defined structures
- QSAR predictions can be made for well-defined structures
- Using a single well-defined structure to represent a UVCB may lead to faulty application of predictions and poor decisions



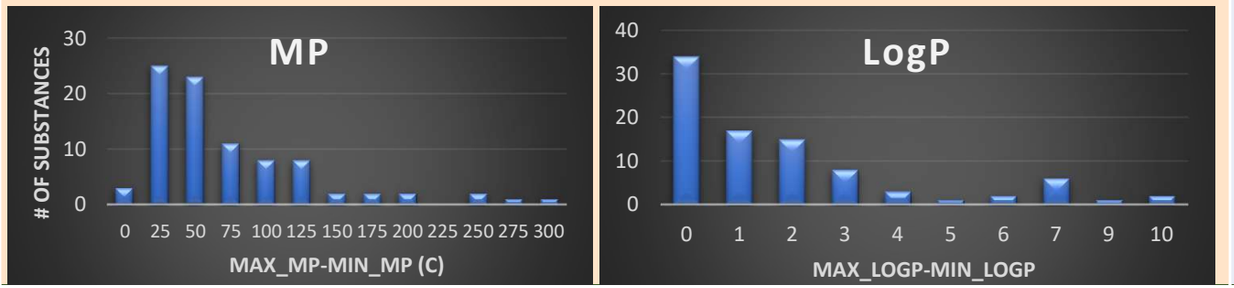
APPROACH

- Represent each UVCB with a Markush structure
- Enumerate or search Markush structures to link to defined structures
- Aggregate the predictions on the defined structures to discern the average and boundaries of predicted endpoints



MAIN RESULTS

- Large differences in predictions between UVCB constituents
- Using only a single representative structure could be problematic and lead to erroneous assessments for UVCBs



IMPACT

- The risk of a UVCB can be informed by programmatically determining the potential constituents and applying well-developed models to improve our understanding of the potential toxicity and exposure profiles for those constituents.

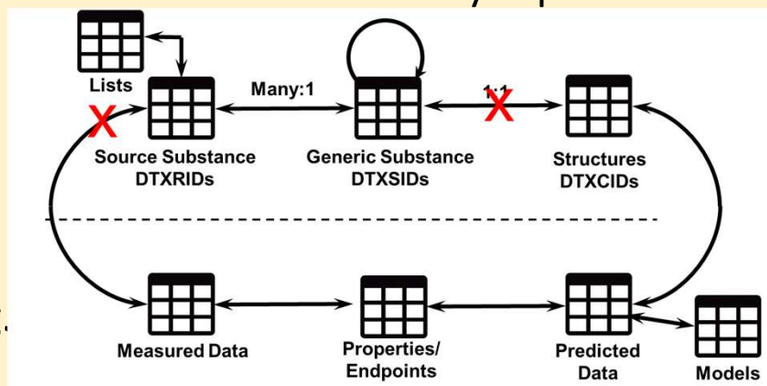
- **For more information, contact:** Christopher Grulke; grulke.chris@epa.gov

Representation and Enumeration of UVCB Substances to Enable QSAR Predictions for Substances with Structural Uncertainty

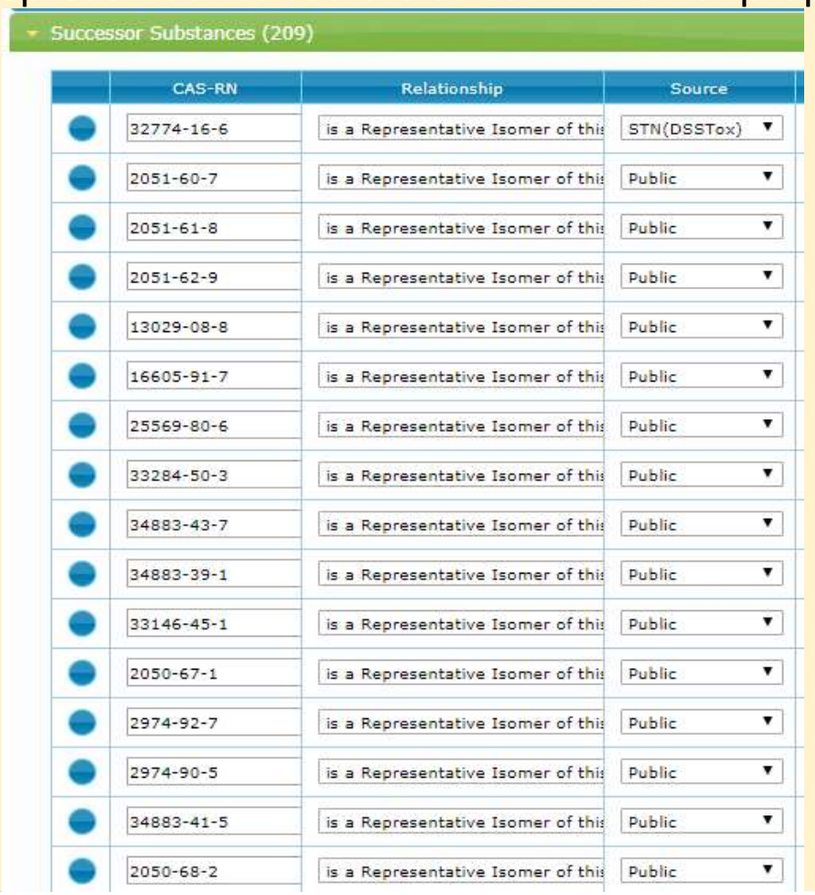
PROBLEM

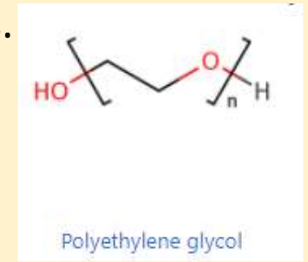
1. The EPA must regulate UVCB chemicals under TSCA

2. However, for most UVCB chemicals, measured data does not exist and a structure cannot fully represent the substance. This broken connection to data inhibits informed decision making.



3. Manually assigning representative structures is time consuming. Generating the full compendium of constituents for a complex UVCB is labor intensive. Selecting a single or small number of representatives could bias estimation of properties.





represented by this

Ethylene glycol

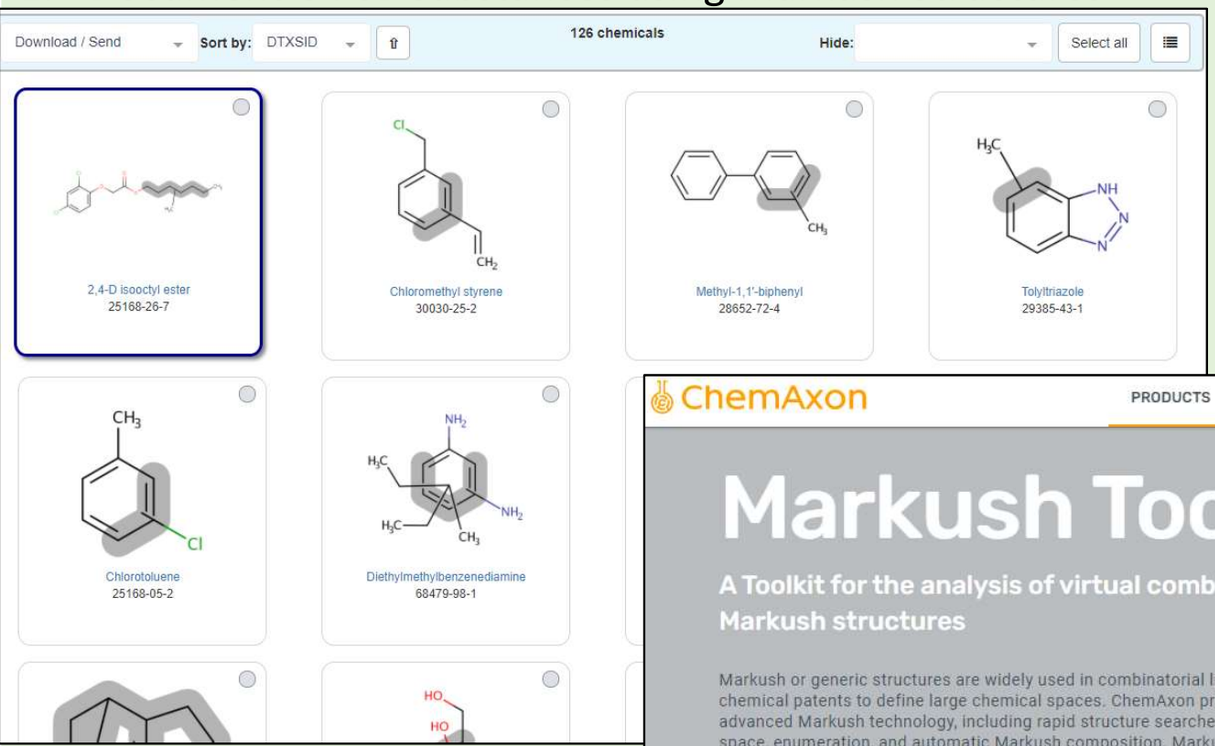
or this?

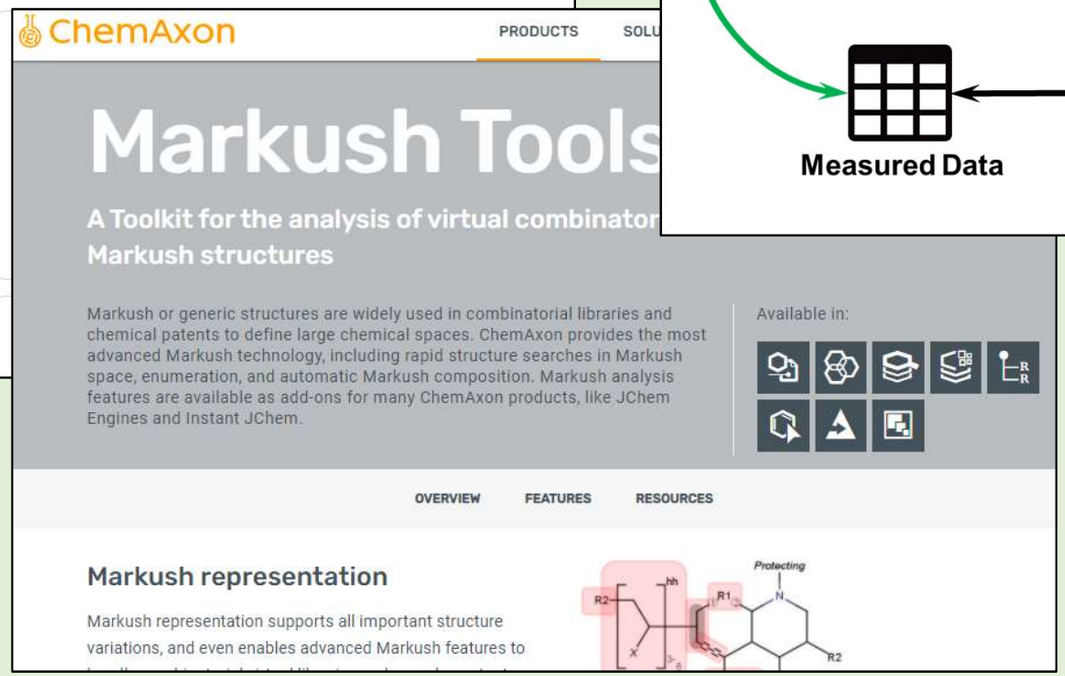
Octaethylene glycol

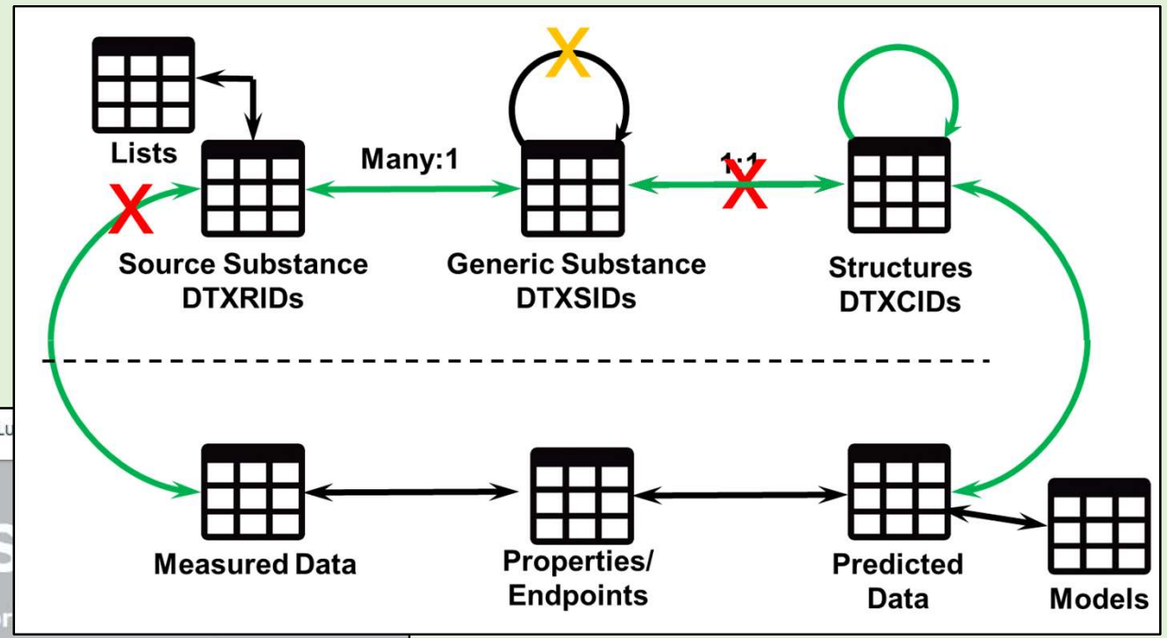
Representation and Enumeration of UVCB Substances to Enable QSAR Predictions for Substances with Structural Uncertainty

APPROACH

1. Document UVCB substances using Markush Structures





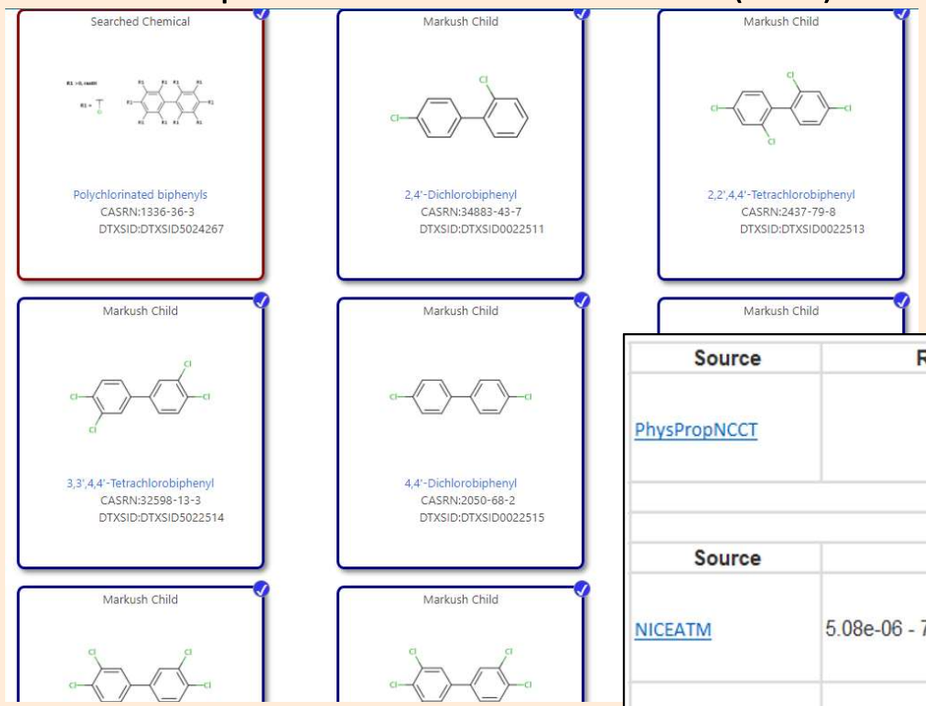


2. Apply ChemAxon's Markush Tool to enumerate and link to defined structures stored in DSSTox database.

3. Enumeration and linkage to compounds (in addition to manual annotation of substance relationships) provides access to a new set of predicted and measured data for the constituents.

Representation and Enumeration of UVCB Substances to Enable QSAR Predictions for Substances with Structural Uncertainty

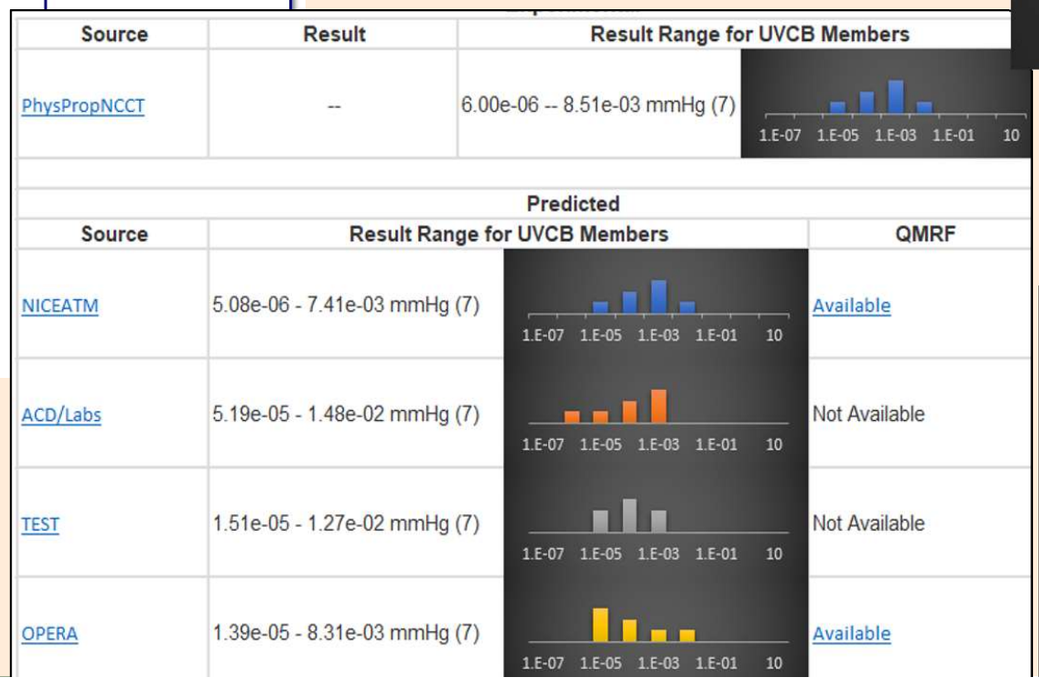
- Constituent substances can be accessed in the CompTox Chemicals Dashboard (CCD)

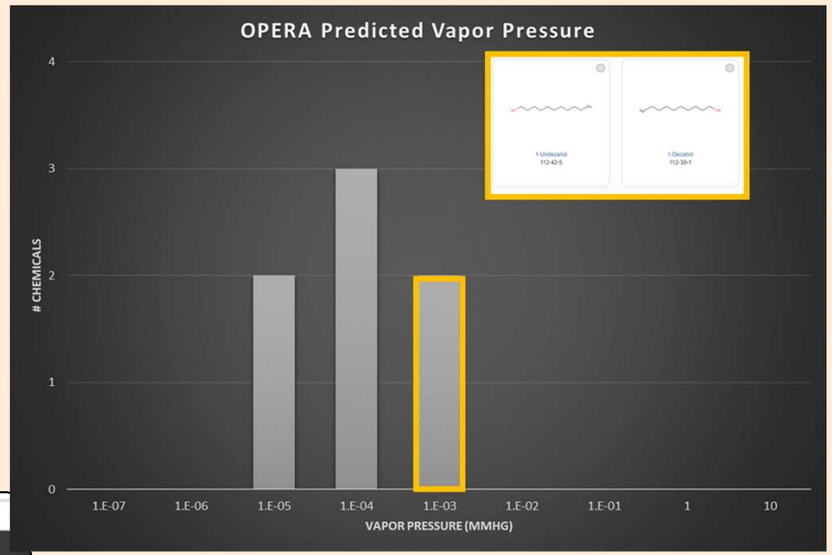


- Distributions of measured and predicted property values can be collected and provided to enable assessment of all constituents (not yet in CCD)

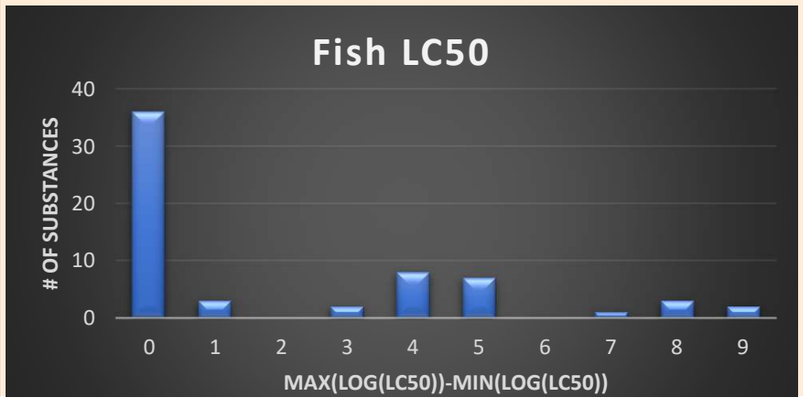
MAIN RESULTS

- When examining the distributions of constituent properties, a user could assess the constituents existing in any bin to evaluate which may be those of greatest concern





- Comparing the limits of predicted fish toxicity within a UVCB's constituents indicate that using a single representative could yield misleading results



Representation and Enumeration of UVCB Substances to Enable QSAR Predictions for Substances with Structural Uncertainty

Summary

- Markush and query structures provide a mechanism for documenting chemical concepts that are otherwise not representable with defined structure formats
- Documenting such structures allows for programmatic workflows to generate predictions and aggregate stored content of constituent substances which could better support risk assessments.

Future Plans

- Continue expanding DSSTox Markush content for representation of UVCB chemicals
- Advance tools for canonicalization of Markush and Query structure representations
- Determine where repeating group enumeration can be ended due to addition of units not causing significant property changes
- Add functionality into the CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) for accessing the property values linked by enumerate structures

References

1. Grulke, C. M. et al. EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. Computational Toxicology 12, 100096 (2019)
2. Williams, A. J. et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. Journal of Cheminformatics 9, 61 (2017)
3. Mansouri, K. et al. OPERA models for predicting physicochemical properties and environmental fate endpoints. Journal of Cheminformatics 10, 10 (2018)
4. Martin, T. M. et al. Prediction of the Acute Toxicity (96-h LC50) of Organic Compounds in the Fathead Minnow (Pimephales Promelas) Using a Group Contribution Method. Chemical Research in Toxicology, 14, 10 (2001)