MetSim: Integrated Programmatic Access and Pathway Management for Xenobiotic Metabolism Tools

Louis Groff[1], Imran Shah[1], and Grace Patlewicz[1]
[1]U.S. Environmental Protection Agency, Office of Research and Development
Center for Computational Toxicology and Exposure, Research Triangle Park, NC

United States Environmental Protection Agency

62nd Annual Meeting & ToxExpo • Nashville, TN
March 19–23, 2023

ORCID: 0000-0002-6357-0508
Louis Groff | groff.louis@epa.gov | 919-541-4256

Background: *In Silico* Metabolism and its Potential Use in Read-Across Frameworks

Motivation for Metabolism Within Read-Across

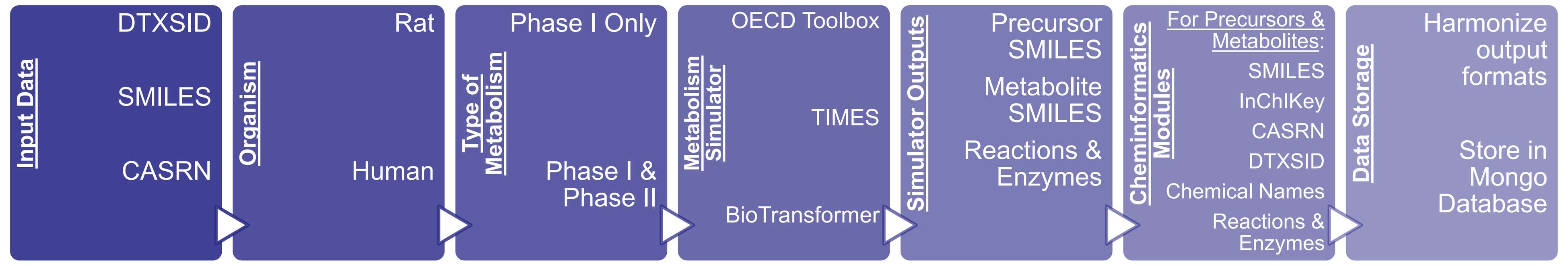
- Analysis of environmental samples via high-throughput screening methods has the potential to uncover new or data-poor chemicals, or contaminants of emerging concern (CECs)
- Read-across methodologies such as EPA's Generalized Read-Across (GenRA) are presently utilized within EPA to select analogue chemicals for which endpoint metrics such as points of departure (POD) or thresholds of toxicological concern (TTC) are known
- For xenobiotic CECs, read-across is often based on structural analogues for chemicals
- We seek to extend this framework to include metabolism similarity between CECs and their analogues as an additional justification for analogue selection within GenRA

Challenges to Address for Successful Implementation into GenRA

- Metabolism is predicted via a variety of *in silico* tools
- Unlike inputs and outputs require harmonization of output data format
- No single standardized database of xenobiotic metabolism pathways of environmental chemicals at present
- Develop database to store prediction data and for developing metabolic graphical networks

Methods: Metabolism Simulation (MetSim) Algorithm

MetSim Algorithm Summary Flow Chart



- All MetSim tools require either an application programming interface (API) or command line interface (CLI) for incorporation into MetSim.
- All tools require SMILES as an input parameter. SMILES strings can be obtained from DSSTox Substance ID (DTXSID) searches if registered, and CAS Registry Number is an alternate input for the OECD Toolbox WebAPI

••Human tissue metabolism is handled with BioTransformer, which can perform phase I and phase II metabolism

••*In Vitro* and *In Vivo* Rat tissue metabolism is handled with the OECD Toolbox and TIMES. Toolbox only performs phase I metabolism, TIMES includes phase II metabolism as well

- Outputs are aggregated into a harmonized output dictionary format, including precursor and metabolite SMILES, and reactions and enzymes from BioTransformer or TIMES outputs.
- Output metadata supplemented via EPA Cheminformatics Modules Standardizer API queries to obtain DTXSID and CASRN for a metabolite from its SMILES, where registered
- Output dictionaries for each parent are stored for later interaction in a Mongo database

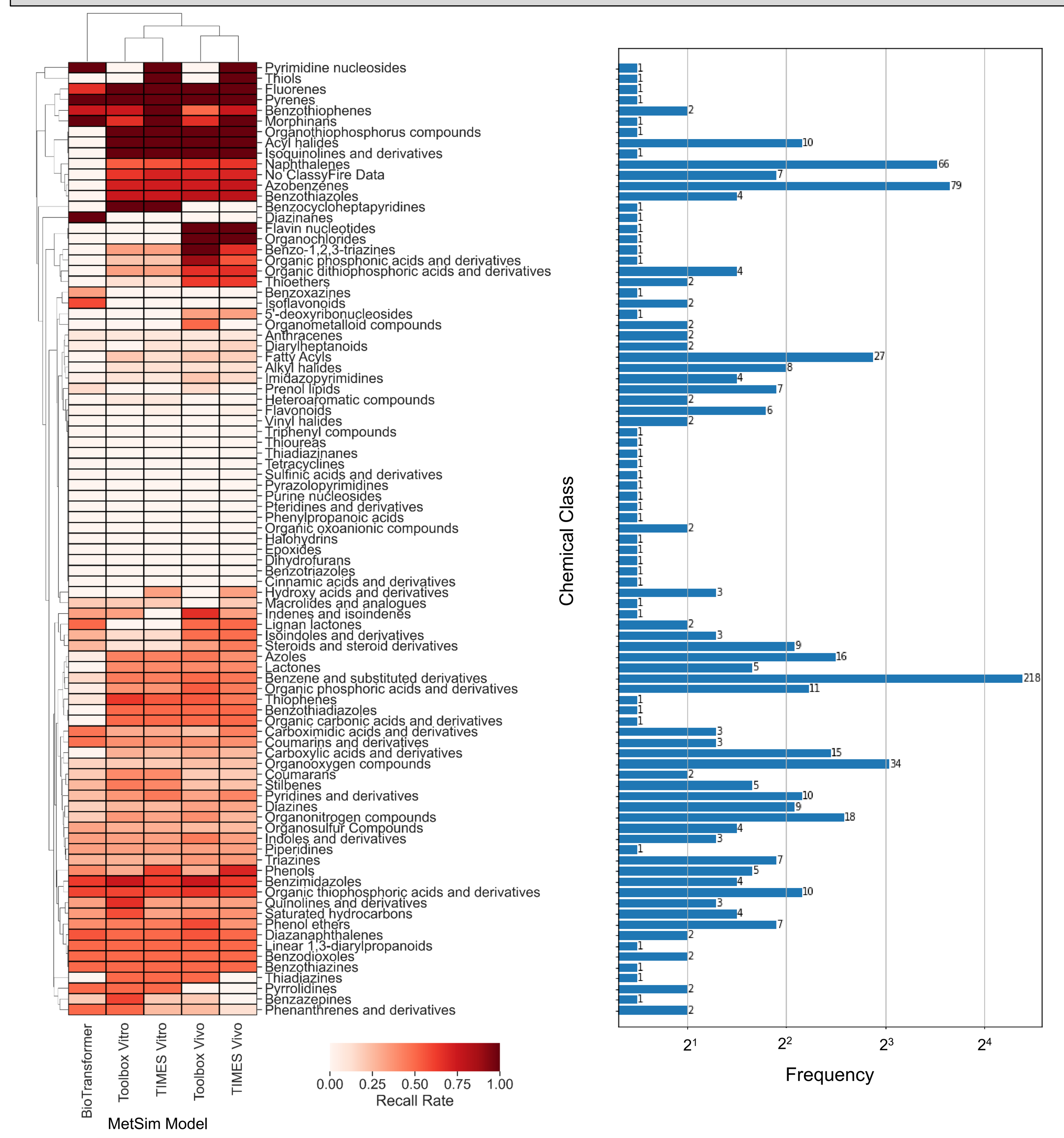
Results: Predictive Performance of MetSim Tools

MetSim Performance Metrics by Choice of Tool/Model on DSSTox-Registered Metabolites

DSSTox 700 Phase I & Phase II Metabolites	BioTransformer 2x Phase I & 1x Phase II	TIMES In Vitro	TIMES In Vivo	OECD Toolbox In Vitro	OECD Toolbox In Vivo
True Positives	266	595	648	595	679
False Positives	58357	4416	6018	4485	6937
False Negatives	1629	1305	1249	1304	1222
Total Predictions	58623	5410	6666	5080	7616
Total Precision	0.004	0.11	0.10	0.12	0.09
Total Recall	0.14	0.31	0.34	0.31	0.36
Total Reported Metabolites	1895	1895	1895	1895	1895

- Dataset consists of 700 EPA-relevant drugs and environmental chemicals
- Regardless of tool choice, recall rate was low
- BioTransformer generates many predictions, but yielded lowest recall for this dataset
- Toolbox/TIMES generate fewer predictions, yielding both higher precision and higher recall than BioTransformer for this dataset
- Toolbox *In Vivo* Rat Simulator yielded the highest recall of all tools and models

Recall Rate Hierarchically Clustered on Chemical Class With Histogram of Class Frequency

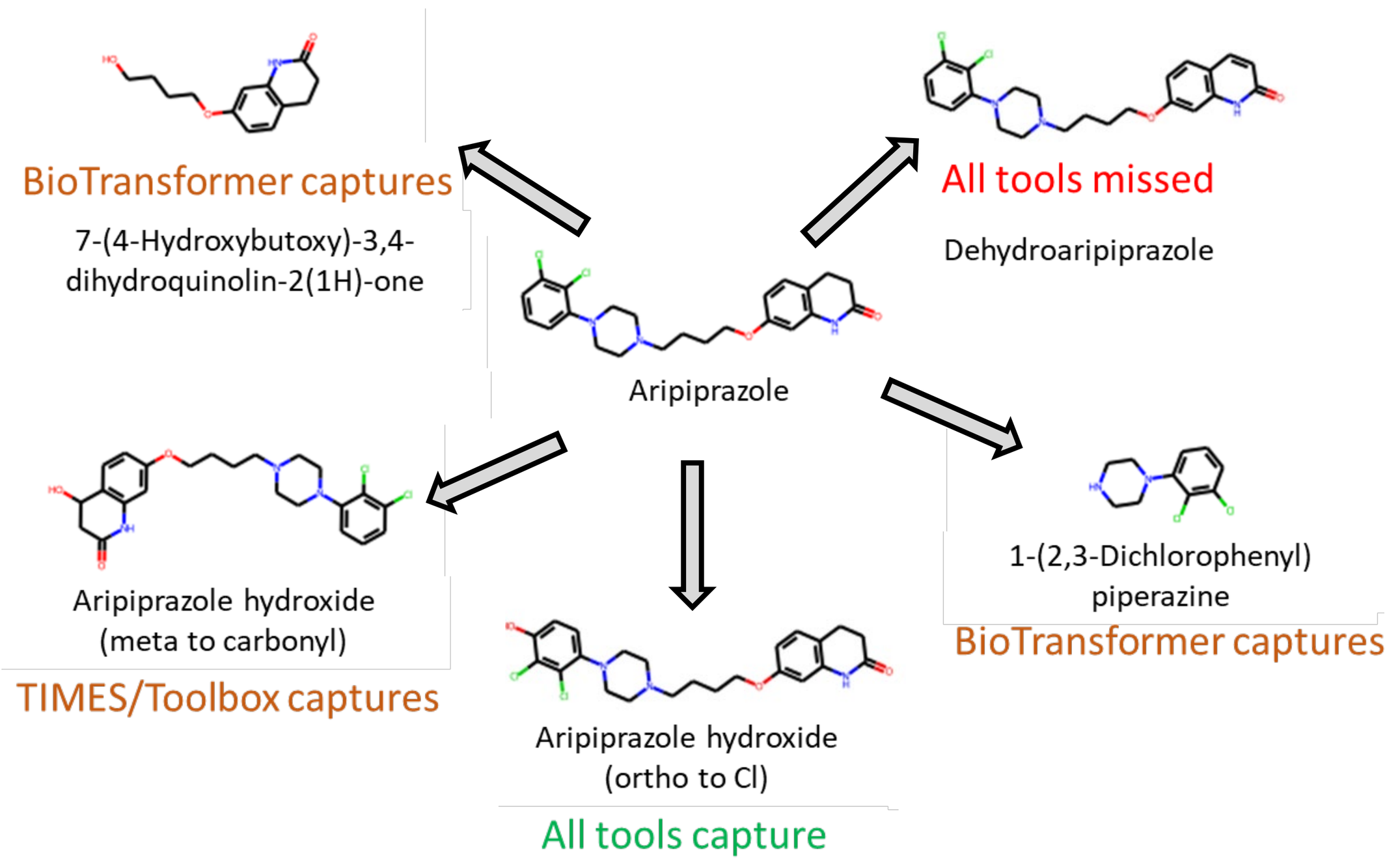


- Hierarchically clustered heatmap on left shows mean recall rate per chemical class given scaled white to red from 0 to 1, respectively
- Histogram on right shows number of compounds in the dataset that comprise each chemical class, on a log 2 scale
- Chemical classes assigned using Wishart Lab ClassyFire API
- Highest mean recall rates observed across all tools for classes including pyrimidine nucleosides, fluorenes, pyrenes, and benzimidazoles
- Lowest mean recall rates observed across all tools for classes including vinyl halides, cinnamic acids, organic oxoanionic compounds, and flavonoids
- Highest frequency chemical classes include benzenes and substituted derivatives, azobenzenes, naphthalenes, and organooxygen compounds

Discussion: Diversity of Tools Can Enhance Metabolite Recall

Example: Tool Specific Performance on Aripiprazole and its Phase I Metabolites

- From literature dataset of pharmaceuticals and non-steroidal anti-inflammatory drug (NSAID) chemicals including their phase I metabolites
- Individual recall per tool is 0.6 for BioTransformer, 0.4 for TIMES/Toolbox, and combined recall is 0.8
- Of the five metabolites, all three tools captured one hydroxylated aripiprazole metabolite.
- All tools missed the oxidized metabolite dehydroaripiprazole



- TIMES and OECD Toolbox capture the second of two aripiprazole hydroxides but miss the other three metabolites
- BioTransformer captures the 7-(4-hydroxybutoxy)-3,4-dihydroquinolin-2(1H)-one and 1-(2,3-Dichlorophenyl) piperazine fragment metabolites but misses the remaining two metabolites

Usefulness of Tools Applied in this Study

- Low precision of BioTransformer is offset by the fact that expansive sets of predictions can be helpful for high-throughput screening studies aimed at identifying CECs, such as Non-Targeted Analysis (NTA), where CECs could be metabolites in an environmental sample, and later correlated with their respective parents
- Toolbox/TIMES training datasets contain more presently known environmentally relevant chemicals

Conclusions and Future Directions

- Our MetSim algorithm was successfully utilized to harmonize inputs and outputs between dissimilar xenobiotic metabolism tools, and for storage and interaction with prediction data via Mongo database
- Algorithm set up such that new tools are readily implemented if they possess an API or CLI
- Individual tools can be utilized for metabolism in human tissues or rat tissues, but show some initial synergy with combining tools
- All tools yielded low recall on EPA-relevant dataset of 700 drugs and environmental chemicals,
- Inspection of false negative predictions on chemical class can provide insight into improvements on transformation rule sets for each tool
- Future work in progress using graphing of MetSim predictions (MetGraph) can be utilized to compare metabolic similarity for read-across analogue selection

References

Upcoming Publication: Groff, L. C., et al. *Chem. Res. Tox. 2023,* In Final Preparation.
GenRA Tool: https://comptox.epa.gov/genra/
Cheminformatics Modules Standardizer: https://hazard-dev.sciencedataexperts.com/#/stdizer
ClassyFire API: http://classyfire.wishartlab.com/

Innovative Research for a Sustainable Future

Disclaimer: The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA.