Comparing performance of *in silico* metabolism tools using data derived from literature and non-targeted analysis

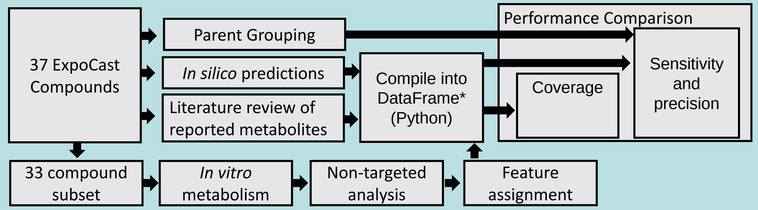
Matthew W. Boyce[1,a,b], Brian Meyer[b], Chris Grulke[b], Grace Patlewicz[b]
ORCID: [1]https://orcid.org/0000-0002-3794-1678
[a]Oak Ridge Associated University, Oak Ridge, Tennessee, USA
[b]Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

United States Environmental Protection Agency

Abstract

Understanding the metabolic fate of a chemical substance is important for evaluating its toxicity. Changes in the regulatory landscape of chemical safety assessment provide opportunities to use *in silico* tools for metabolism prediction. In this study, a set of 37 structurally diverse chemicals were compiled from the EPA ExpoCast inventory to compare and contrast a selection of *in silico* tools, in terms of their coverage and performance. The tools were Systematic Generation of Metabolites (SyGMa), Meteor Nexus, BioTransformer, Tissue Metabolism Simulator (TIMES), OECD Toolbox, and Chemical Transformation Simulator (CTS). Performance, as characterized by sensitivity and precision, were determined by comparing predictions against metabolites reported in literature. Reported metabolites (438 in total) were extracted from 49 papers. Coverage was calculated to provide a relative comparison between tools. Meteor, TIMES, Toolbox, and CTS predictions were run in batches, using default settings. SyGMa and BioTransfomer were run with user-defined settings, (two passes of phase I and one pass of phase II). Hierarchical clustering revealed high similarity between TIMES and Toolbox. SyGMa had the highest coverage, matching an average of 41.2% of predictions generated by the other tools. SyGMa was also prone to significant overpredicting, generating a total of 5,125 predictions or 67% of total predictions. Precision and sensitivity values ranged from 4.7-23.7% and 15-27.5% respectively. TIMES had the highest performance overall. Current efforts are focused on evaluating the concordance of *in vitro* data, newly generated, relative to the literature data and *in silico* predictions.

Analysis Workflow



*Data were compiled using InChI Keys and DTXSIDs, which were sourced from the EPA CompTox Chemicals Dashboard or generated via the DSSTox ChemReg internal application.

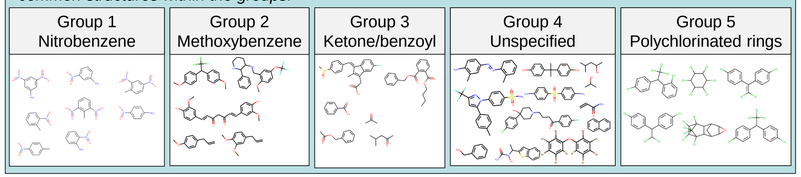
ExpoCast compounds and Parent Grouping

37 Starting Compounds – Grouped via ClassyFire classifications

ClassyFire (http://classyfire.wishartlab.com/) is a freely available web-based application that uses a rule-based approach to relate chemical structure to taxonomic groupings (in the form of textual descriptors).

- Taxonomic descriptors were converted into bit vectors for each parent.
- Parents were grouped using Jaccard distance between bit vectors and hierarchical clustering.

By generating parental groupings, we can evaluate local differences in model performance relative to common structures within the groups.



Generating *In Silico* Predictions

Predictions were generated using six *in silico* models:

Model	Availability	Module/Species	Prediction Settings	Number of Predicted Metabolites
BioTransformer	Free (http://biotransformer.ca)	Human	Phase I: 2 steps Phase II: 1 step	827
Chemical Transformation Simulator	Free (https://qed.epacdx.net/cts/)	Human	Phase I: 3 steps	887
Meteor	Commercial (https://www.lhasalimited.org/)	Mammal*	Default	459
Toolbox	Free (https://qsartoolbox.org/)	Rat (S9, *in vitro*), Rat (*in vivo*)	Default	312
TIMES‡	Commercial (http://oasis-lmc.org)	Rat (S9, *in vitro*), Rat (*in vivo*)	Default	211 (*in vitro*), 459 (*in vivo*)
SyGMa	Free (https://sygma.readthedocs.io)	Human	Phase I: 2 steps Phase II: 1 step	5215

*No differences in metabolite predictions were identified when comparing the 'mammal' and 'rat' modules
‡Both TIMES modules (*in vivo* and *in vitro*) were run separately to allow for intra-model comparison

Baseline for generating predictions

Predictions were generated using specific conditions across all models to ensure performance comparisons were as uniform as possible:

- All models used a max depth of 3 biotransformations from the parent with exception to the Rat (*in vivo*) module of Toolbox, which has a predefined depth limit of 999.
- Phase I and Phase II predictions are included in all models except Toolbox and Chemical Transformation Simulator, which exclude Phase II predictions.

Literature Review

Literature search approach

1. Biotransformations for each parent compound were queried in multiple databases:
   - EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard)
   - Human Metabolome Database (https://hmdb.ca/)
   - DrugBank (https://go.drugbank.com/)
2. AbstractSifter (v4) (Baker et al, 2017: doi:10.12688/f1000research.12865.1) was used to query the PubChem database using the chemical name, DTXSID, and a collection of additional terms (metabolite, metabolism, pharmacokinetics, clearance, and excretion).
3. In cases where no suitable articles were identified within the databases or AbstractSifter, manual searches were performed in Web of Science and Google Scholar using chemical names and similar search terms outlined for the AbstractSifter.

Metabolites extracted from literature

- Metabolites were identified for all 37 ExpoCast compounds across 49 journal articles, which encompassed *in vitro* and *in vivo* data generated from dog, mouse, rat, and human studies.
- 438 metabolites (reported and theoretical) were extracted to compare against *in silico* predictions
- If structural isomers (Markush structures) were reported, discrete children structures were generated for comparison.

Performance Comparison - Coverage

Coverage: $\frac{Predictions\ A\ \cap Predictions\ B}{Predictions\ B}$

Coverage indicates the overlap of predictions between two models e.g., Toolbox's predictions cover 15.13% of predictions generated by Meteor, whereas the inverse direction shows that Meteor covers 34.39% of Toolbox's predictions.

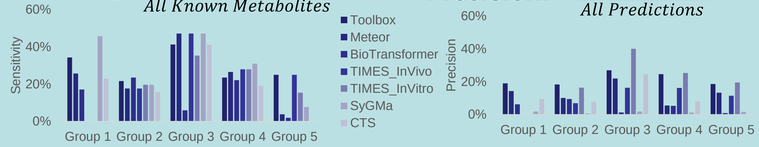
Model	%Coverage of other software						
	Toolbox	Meteor	BioTransformer	TIMES (*in vivo*)	TIMES (*in vitro*)	SyGMa	CTS
Toolbox	-	15.13	7.62	40.74	74.41	2.86	27.33
Meteor	34.39	-	9.07	25.93	33.65	4.60	19.92
BioTransformer	20.06	10.50	-	10.24	23.70	4.31	17.16
TIMES (*in vivo*)	59.55	16.67	5.68	-	61.61	3.99	14.41
TIMES (*in vitro*)	50.00	9.94	6.05	28.32	-	2.13	16.74
SyGMa	47.45	33.61	27.21	45.32	52.61	-	32.42
CTS	41.08	13.17	9.79	14.81	37.44	2.93	-

- SyGMa consistently exhibited high coverage when compared to other tools. These results indicate that SyGMa is the least conservative in its predictions, which is further supported by the number of predictions generated.
- Toolbox and TIMES have high reciprocal coverage, suggesting these simulators have relatively high similarity.

Performance Comparison - Sensitivity/Precision

Precision and sensitivity provide empirical metrics for evaluating a model's performance
Both metrics were determined for each of the five groupings generated using ClassyFire descriptors.

Sensitivity: $\frac{True\ Predictions}{All\ Known\ Metabolites}$ **Precision:** $\frac{True\ Predictions}{All\ Predictions}$



With exception to Group 1 (nitrobenzenes), TIMES (in vitro) was the best overall performing software with and average sensitivity of 25% and precision of 25%. In cases where sensitivity is prioritized, SyGMa showed the greatest score (30%), which can be attributed to the extensive number of predictions it generated.

Next steps - Incorporate non-targeted analysis

Non-targeted analysis (NTA) is an analytical technique that relies on high-resolution mass spectrometry (HRMS) to identify compounds in complex mixtures. The ability to measure a wide range of unknowns within a single sample makes NTA well suited for characterizing metabolites.

NTA will be used to analyze *in vitro* data collected from a subset of the original ExpoCast compounds. Metabolites identified through this workflow will be compared against the predictions to further refine comparisons between the software.

In vitro metabolism	Non-targeted analysis	Metabolite identification
Primary human hepatocytes metabolized 33 parent compounds over 4 h time course.	LC- and GC-MS used to collect high-resolution single and tandem MS data.	Tentative structures identified by comparing predicted fragmentation patterns to tandem MS data.