National Institute of Environmental Health Sciences

Prioritization of the Tox21 10k Library for Xenobiotic Metabolism and Toxicity Studies Using In Silico Metabolism Models

Stephen S. Ferguson[1], Neepa Choksi[2], Nicole Kleinstreuer[3], Michael Lawless[4], Raymond Tice[1]

[1]Biomolecular Screening Branch, DNTP, NIEHS, RTP, NC, USA; [2]ILS, RTP, NC; [3]ILS/NICEATM, RTP, NC; [4]Simulations Plus, Inc.

Abstract

One of the current challenges of the U.S. Tox21 high throughput screening effort is the lack of approaches incorporating xenobiotic metabolism. Unfortunately, with the large numbers of chemicals being examined (~8,300 uniques in Tox21 Phase II 10k library), it is cost prohibitive to assess all of these chemicals in metabolically competent assay systems (e.g., primary hepatocytes, HepaRG). To address this issue, in silico metabolism & toxicity models (ADMET Predictor) have been employed to prioritize chemicals for testing in metabolically-competent assay systems. Three metabolism prediction types were used to rank the Tox21 10k library.

1) **Substrate predictions** for 341 established substrates for CYP1A2, 2C9, 2C19, 2D6, and 3A4 showed 89.4% correct calls, and the approach was extended to 8,193 Tox21 chemicals for 9 P450 and 9 UGT enzymes. **48,320** substrate 'hits' were identified. Chemicals were ranked via 'hit' frequencies. However, substrate predictions alone do not account for the extent of metabolism and metabolite structure properties (toxicities).

2) **Extent of metabolism (metabolic clearance) predictions** were generated for the 10k library (5 individual human P450 enzymes). A 74 chemical set with published in vitro clearance data (human liver microsomes) was evaluated. Summed/weighted CL_{int} predictions to relative expression in human liver yielded a marginal Pearson correlation (0.51) with this limited set of chemicals. The approach was used to rank the 10k library via predicted extents of metabolism.

3) **Metabolite structure predictions** (3 rounds of metabolism) were generated with MedChem Studio on 8,193 chemicals. Putative metabolite structures were generated for 126,186 unique metabolite structures. To assess the accuracy/applicability of metabolite structure predictions, a database of 211 known metabolite structures (41 drugs and 170 environmental chemicals) was constructed. 87 of the 211 metabolites had 'predictable' structures limited by the lack of non-P450 enzyme predictions models (e.g. organophosphates, amides, halogens, inorganics). Of the 87 'predictable' structures, 69 (79.3%) were correctly predicted as true positive metabolite structures.

Predicted metabolite structures (and parent chemicals) were further analyzed with 27 mammalian toxicity prediction models within ADMET Predictor, and chemicals were ranked/prioritized via predicted toxicities.

Introduction

One of the current challenges of the U.S. Tox21 high throughput screening effort (i.e. screening large numbers of chemicals in human cell-based assays using robotics) is the lack of approaches incorporating robust xenobiotic metabolism to generate species appropriate metabolites. In vitro liver models and assay systems continue to evolve to support more physiologically-relevant phenotypes and cell types, and hold the promise of addressing these challenge while improving our understanding of chemical toxicity pathways.

Pharmaceutical research has had reasonable success in using various in vitro and in silico ADME models to predict xenobiotic metabolism (e.g., compound exposure to in vitro liver models) and scale to in vivo pharmacokinetics. Therefore, these types of model systems have the potential to prioritize chemicals for testing in more physiologically-relevant model systems where xenobiotic metabolism may be most important while supporting efforts for in vitro to in vivo extrapolation of pathway perturbations in a more quantitative manner.

Unfortunately, with the large numbers of chemicals being examined (~8,300 in Phase II Tox21 library), it is impractical and cost prohibitive to assess all chemicals in more physiologically-relevant metabolically competent assay systems (e.g. primary hepatocytes, HepaRG). In addition, we continue to evaluate the best assay approaches with these evolving models to provide data rich results that identify and explore toxicity pathways. Therefore, moving forward with smaller, focused sets of chemicals will be required in the near term.

To address this issue, in silico methods for predicting xenobiotic metabolism and toxicity have been evaluated and employed to rank the Tox21 10k library. In this report, we evaluated the ADMET Predictor and MedChem Studio software in collaboration with Simulations Plus. Using both metabolism and toxicity prediction models on parent chemical structures, we assessed P450 and UGT substrates, predicted extents of metabolism for 5 major human drug metabolism models (P450s), and predicted putative metabolite structures over 9 major drug metabolizing enzyme pathways (P450s). Using these data, we generated ranked lists of chemicals for further analysis with in vitro liver models competent for xenobiotic metabolism.

Materials and Methods

The Tox21 10k library used for these analyses contained 8,307 chemical structures generated in Oct 2012 (TOX21S_v3a_8307_02Oct2012.sdf).

ADMET Predictor and MedChem Studio software (Simulations Plus) were used to make predictions for: 1) Substrate predictions (yes/no for 18 human xenobiotic transformation enzymes (9 P450s, 9 UGTs), 2) metabolic clearance predictions (µL/min/mg recombinant protein) of 5 individual human P450s (CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4), 3) sites of metabolism for 9 human P450 enzymes (CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, and 3A4 and metabolite structures from sites of metabolism predictions, 4) toxicity predictions on parent chemical and putative metabolite structures over 27 mammalian predictions.

Substrate predictions evaluation: A database of 375 compounds identified as substrates for the 5 P450 drug metabolizing enzymes with quantitative models was constructed from literature and databases to evaluate the utility of the substrate predictions within ADMET Predictor (Figure 1).

Metabolic clearance predictions evaluation: The quantitative models for metabolic clearance were constructed from recombinant enzyme data for 5 drug metabolizing enzymes (CYP1A2, CYP2C9, CYP2C19, CYP2D6, & CYP3A4). Since predictions across all enzymes are more suited to our needs in ranking chemicals based on overall extent of metabolism, individual CL_{int} values were summed (weighted by relative expression levels in human liver). A database of 74 compounds was constructed from literature using human liver microsomes data (Figure 2).

Metabolite structure predictions evaluation: A database of 211 metabolite structures was constructed representing 41 drug structures and 170 environmental chemicals (Figure 3).

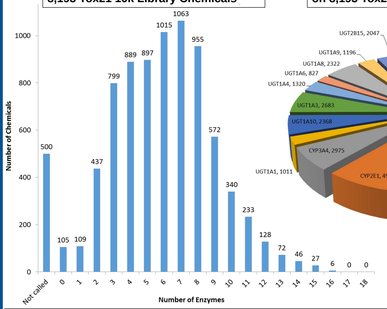
For toxicity predictions, ADMET-Predictor generated 27 predictive models for mammalian toxicity for chemicals and putative metabolite structures including: chromosomal aberrations, 10 genotoxicity models (Ames), elevated human alkaline phosphatase levels, elevated human GGT enzyme levels, elevated LDH enzyme levels, maximum tolerated dose <3.16mg/kg/day, phospholipidosis, reproductive toxicity, respiratory sensitization, elevated human AST enzyme levels, elevated human ALT enzyme levels, skin sensitization, TD_{50} mouse carcinogenicity, TD_{50} rat carcinogenicity, LD_{50} acute rat exposure, human hERG cardiac ion channel, estrogen receptor (ER-rat) relative binding affinity, androgen receptor (AR-rat) relative binding affinity.[2]

Figure 1: Substrate Predictions

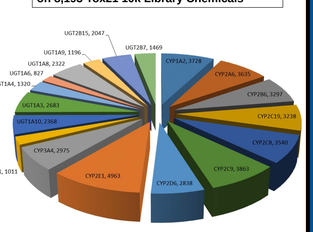
Analysis of database of 375 established P450 substrates to evaluate substrate prediction models for enzymes with quantitative CL_{INT} models

	CYP3A4	CYP1A2	CYP2C9	CYP2C19	CYP2D6
No. Established Substrates Database	105	77	59	51	83
No. Chemicals Within Model Chemical Space	92	73	52	47	77
No. Correct Predictions	86	64	47	36	72
% Correct Prediction	93.5%	87.7%	90.4%	76.6%	93.2%

Histogram of Substrate Predictions for 8,193 Tox21 10k Library Chemicals

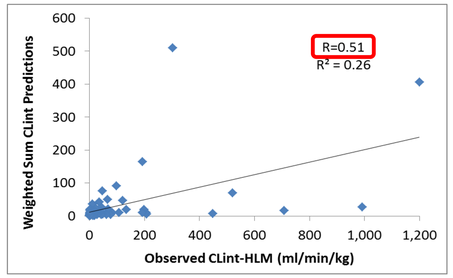


Distribution of 47,320 Substrate Predictions on 8,193 Tox21 10k Library Chemicals



Substrate predictions with 5 enzymes (with respective CL_{int} models) were 76.6%-93.5% accurate for 375 known substrates. However, substrate predictions alone are insufficient for chemical prioritization (e.g. testing in metabolically-robust systems) as they do not account for extents of metabolism nor predict metabolite structures for structural alert evaluations.

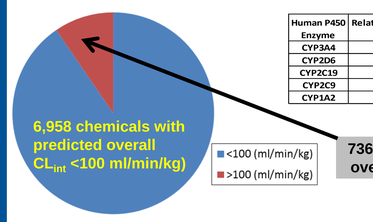
Figure 2: Metabolic Clearance Predictions



- Constructed database from literature of metabolic clearance data with human liver microsomes for 74 compounds in the Tox21 10k library.
- ADMET-Predictor was used to predictions metabolic clearance for 5 human P450 enzymes with quantitative CL_{INT} models.
- CL_{int} predictions were summed/weighted to relative expression levels in human liver to calculated a weighted sum CL_{int} to rank chemicals.

Equation for Combined CLint Over 5 P450s with Quantitative Models

$$CL_{int} = CLi_{int\ 3A4} * 0.514 + CL_{int\ 2D6} * 0.026 + CL_{int\ 2C19} * 0.046 + CL_{int\ 2C9} * 0.242 + CL_{int\ 1A2} * 0.173$$



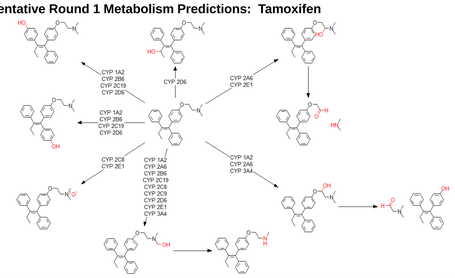
Human P450 Enzyme	Relative Expression Level (%) [1]	Fraction of 5 Enzymes with Quantitative Models
CYP3A4	35.7	0.514
CYP2D6	1.8	0.026
CYP2C19	3.2	0.046
CYP2C9	16.8	0.242
CYP1A2	12	0.173

- Identified CL_{int} cutoff of 100 ml/min/kg to 'bin' the Tox21 10k library for further evaluations in metabolically-robust in vitro liver models and assay systems. For the chemicals in the established substrates database, the 100 ml/min/kg cutoff gives an 85.7% correct true positive prediction, with true total at 63.9%.

Figure 3: Metabolite Structure Predictions

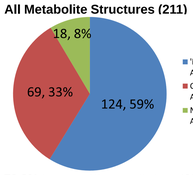
From 8,193 parent chemicals in the Tox21 10k library, ~126k unique metabolite structures were generated (3 levels) with MedChem Studio v. 3.5.0.21 (AP 6.5) across 9 cytochrome P450 enzymes (CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4)

Representative Round 1 Metabolism Predictions: Tamoxifen

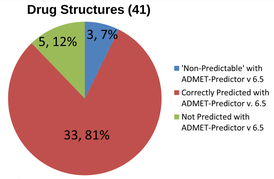


Evaluation of Metabolite Structure Predictions with a Database of 211 Metabolite Structures (~40 drugs, ~170 environmental chemicals)

All Metabolite Structures (211)

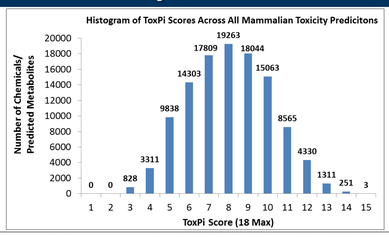


Drug Structures (41)

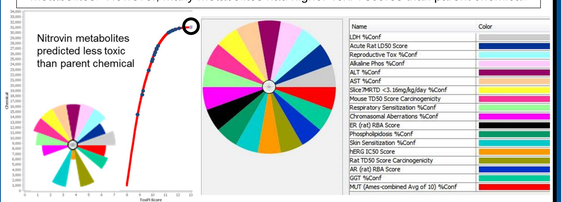


- **79.3% true positive** predictions (69/87) for 'predictable' structures within compiled metabolite database
- Only 87 of 211 metabolite structures compiled were 'predictable' within the transformation space of ADMET-Predictor 6.5
- Missing transformation types included: esterase, dehalogenation reactions, epoxide formation, metalloid reactions, amidase, environmental degradation products, amino acid conjugation, Phase II metabolism, organophosphate transformations

Figure 4: Toxicity Predictions with AP 6.5



- 133k unique structures were analyzed with the toxicity models within ADMET-Predictor 6.5 including: genotoxicity, human enzymes, reproductive toxicity, skin/respiratory sensitization, carcinogenicity (TD_{50}: mouse, rat), rat ER & AR, acute rate LD_{50}
- Toxicity predictions were scaled to generated a cumulative ToxPi score across predicted mammalian toxicity pathways to generate a single prioritized list of chemicals
- Change in ToxPi values between parent & predicted metabolite structures were calculated/ranked, and in general parent chemicals had higher ToxPi scores than metabolites. However, many metabolites had higher ToxPi scores than parent chemical.



- A focused analysis of genotoxicity predictions with the 10 Ames test QSAR models (five +/- S9) was performed to prioritize chemicals for testing in metabolically-robust in vitro liver models
- Genotoxicity predictions were summed across models with S9 and subtracted from the sum from models without S9 for all chemicals and putative metabolite structures. Chemicals were ranked based and the top 30 are shown below.

Name	Combined Ames: (With S9) - (Without S9)
2-Amino-4-methoxybenzothiazole	412
2-Amino-5,6-dimethylbenzothiazole	409
2-Amino-6-ethoxybenzothiazole	401
2-Amino-6-methoxybenzothiazole	400
9,10-Dibromoanthracene	389
Tioxidazole - M2	384
2,4,6-Trinitro-1,3-dimethyl-5-tert-butylbenzene	380
2,7,8,9-Tricyclazole - M4 - M2	371
Methisazone	370
2-Aminobenzothiazole	370
2-Amino-4-methylbenzothiazole	369
2-Amino-5-azotoluene	353
4,4'-Diaminoazobenzene	353
Phenazopyridine hydrochloride	353
3'-Methyl-4-dimethylaminoazobenzene - M4 - M3	350

Name	Combined Ames: (With S9) - (Without S9)
2-Amino-5,6-dimethylbenzothiazole - M3 - M1	348
7,12-Dimethylbenz(a)anthracene	346
Pipofezine - M2 - M5	345
2-Naphthylamine	343
4-H-Cyclopenta(d,e,f)phenanthrene	343
Pyrene	343
Amiphenazole	343
Benzo(a)pyrene	343
Dequalinium dichloride	343
2-Amino-5,6-dimethylbenzothiazole - M2 - M2	342
Benzo(e)pyrene	340
N-Nitrosodibutylamine	339
N-Nitrosodipropylamine	339
Nitrosomethylaniline	338
4-(Dimethylamino)azobenzene - M3	335

Conclusions & Future Direction

- **Substrate predictions were effective, but are insufficient to prioritize chemicals for further testing in metabolically competent models**
- **CL_{int} predictions were modestly correlative when combining across enzymes to predict HLM turnover. Better models may be needed.**
- **Using a CL_{int} prediction cutoff threshold of 100 ml/min/kg allowed the 10k library to be binned into a 'manageable' number of chemicals with higher likelihood to undergo extensive xenobiotic metabolism (736)**
- **Metabolite structure predictions were made for 8,193 chemicals yielding ~126k unique metabolites using MedChem Studio, with the software effectively predicting 79% for known metabolites (within current models)**
- **For environmental chemicals the range of transformations in AP 6.5 was insufficient (P450 only), and future model development should focus on developing epoxide metabolite predictions for toxicity research.**
- **Simple combined ToxPi visualizations across all available (mostly qualitative) mammalian toxicity models may bias results towards pan toxicants while not accounting for potencies**
- **Specific toxicity predictions are being further evaluated. Initial evaluation of Ames test predictions for genotoxicity yielded 21 out of the top 30 (70%) ranked chemicals that were reported to be associated with genotoxicity.**
- **CL_{int} predictions and toxicity predictions (e.g., genotoxicity) will be combined to further refine chemical rankings**

References

1. Yeo KR, Rostami-Hodjegan A, Tucker GT. **Abundance of cytochromes P450 in human liver: a meta-analysis.** *Proceedings of the British Pharmacological Society.* **2004, v.1(4), p.225. (Abstract)**

2. ADMET-Predictor Internet Site: http://www.simulations-plus.com/Products.aspx?pID=13.