

OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

Case Study – Metabolism Prediction [MetaP]

SUMMARY	2
DESCRIPTION	3
Implementation team	3
Case Study objective	3
Risk assessment framework	3
DEVELOPMENT	4
Use Cases Associated	4
Databases and tools	4
Technical implementation	5
OUTCOMES	6
REFERENCES	7

SUMMARY

Metabolites may well play an important role in adverse effects of parent drug (or other xenobiotic) compounds. In this case study VU (CS leader), HITEC/HHU (associate partner and implementation challenge winner), JGU and UU work together on making methods and tools available for metabolite and site-of-metabolism (SOM) prediction. For that purpose we use and integrate ligand-based metabolite predictors (e.g. MetPred, enviPath, FAME, SMARTCyp) and we will incorporate protein-structure and -dynamics based approaches to predict the site of metabolism (SOM) by Cytochrome P450 enzymes (CYP450s). CYP450s metabolise ~75% of the currently marketed drugs and their active-site shape and plasticity often play an important role in determining the substrate's SOM. We will also make services available for the prediction of microbial biotransformation pathways. During method development, model calibration and validation we will use data from XMetDB and other open-access databases for drugs, xenobiotics and their respective metabolites. To facilitate the combined use of the metabolite prediction approaches and their outcomes, we will benefit of ongoing development in workflow management systems (Nextflow, Squonk, MDStudio) and we will explore integration into and application of these platforms. Once integrated the added value of multiple predictors will be subject of a pilot study on consensus metabolite prediction.

DESCRIPTION

Implementation team

CS leader	Team
Daan Geerke (VU)	VU, UU, JGU, HITeC/HHU

Case Study objective

The objective of this case study is to enable metabolite prediction within the OpenRiskNet infrastructure, and to evaluate and demonstrate the added value of it. For that purpose we will integrate, compare and (ultimately) combine different tools for metabolism prediction, including tools for:

- Ligand-based Site-Of-Metabolism (SOM) prediction using reaction SMARTs, circular fingerprints and/or atomic reactivities;
- QSBR (quantitative-structure biotransformation relationship) modeling of microbial biotransformation;
- Protein-structure and -dynamics based prediction of CYP450 isoform specific binding and SOMs;
- Predicting probabilities for specific reaction type events.

See the “Databases and Tools” section for more details on the corresponding tools. For our comparisons of tool performance we will use selected compounds from literature for which metabolite formation is well described. We anticipate to present our results in one (or more) manuscript(s) on tool integration to illustrate that using several tools has additional value compared to individual tools.

Risk assessment framework

Prediction outcomes will serve as input for other molecular structure-based AO predictors, which relates to Tier 0 (Step 1: identification of molecular structure) and Tier 1 (Step 6: mechanism of action).

DEVELOPMENT

Use Cases Associated

MetaP is associated with [UC2 - Building and using a prediction model](#). For improved metabolite prediction (and thereby improved input for other predictors) we will explore the potential of combining different approaches.

Databases and tools

The table below gives an overview of metabolite prediction tools that are or will be integrated and used in this case study. During method development, model calibration and validation we will take advantage of data from XMetDB and other open-access databases for drugs, xenobiotics and their respective metabolites, as available in ZINC, ChEMBL, DrugBank, EAWAG-BBD and/or the SMARTCyp and FAME suites.

	input	output	Method
Metaprint2D (UU)	2D chemical structure of ligand	Rank atoms (SOMs) for Phase 1 reactions	Preprocess Metabolite reaction database (>100K biotransformations) using MCS. For each query compound, look up similar atom environments based on circular fingerprints https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-362
SMARTcyp (external)	2D chemical structure of ligand	Rank atoms (SOMs) for P450-isoform specific reactions	Combining reactivity (from database on QM calculated transition state energies) with simple 2D molecular accessibility descriptors for SOM prediction http://www.farma.ku.dk/smartcyp/
Plasticity tools (VU)	3D Chemical structure of ligand	Prediction of most probable SOMs for P450-isoform specific reactions	Protein-structure and dynamics based prediction of substrate binding orientations in active site of CYP isoforms (1A2, 2D6, 3A4)
UM-PPS (JGU)			
enviPath (JGU)			
MetPred (UU)	2D chemical structure of ligand	SOMs with Reaction Types for Phase 1 reactions	Similar to Metaprint2D but use ReactionSMARTS to identify reaction types. <i>Manuscript submitted.</i>
MetVap (UU)	2D chemical structure of ligand	Probabilities for Reaction Types	Extract datasets for each reaction type from database, based on a set of ReactionSMARTS. Predict reaction types using Venn-Abers predictors http://proceedings.mlr.press/v60/arvidsson17a.html
FAME (HITeC/HHU)	Chemical structure of ligand	Predicting most probable SOMs for isoform specific reactions	Machine learning using few quantum and circular-environment based atomic descriptors, see https://pubs.acs.org/doi/pdf/10.1021/ci400503

			S
--	--	--	-------------------

Technical implementation

As summarised in the table above, the following tools (may) come available as service:

- MetPred (UU) for SOM prediction (OpenAPI implementation ready, available as webservice); MetPred predicts phase I metabolites by ranking the most probable SOMs and reaction type(s) for a given molecule based on similar atom environments and ReactionSMARTS in annotated datasets;
- SMARTCyp predictor for CYP450 SOMs based on reactivity (freeware and webserver developed by University of Copenhagen); VU is close to finishing its implementation as OpenRiskNet service for combination with information from docking;
- Docking/structure-based CYP450-isoform specific SOM accessibility predictor, using protein template structures obtained from Molecular Dynamics based plasticity/ensemble docking studies (currently being implemented as combined ORN service with SMARTCyp by VU);
- enviPath: prediction of microbial transformation biotransformation pathways and products (following rules represented by SMIRKS);
- UU tool for probabilistic modeling of reaction types;
- FAME: enzyme-specific SOM prediction from machine learning using few quantum and circular-environment based atomic descriptors.

For deployment in the reference environment, an OpenRiskNet API is/will be typically developed for the different tools, and where possible, services will be containerized to facilitate their deployment. The APIs will support and provide access to the core features of the services as summarised above, and will typically accept submissions of chemical structures in common file formats. The API endpoints will offer prediction results in various formats, including the machine-readable JSON format that should ensure seamless integration with the OpenRiskNet infrastructure.

OUTCOMES

In addition to service integration as described above, we will evaluate the added value of integration of metabolite prediction tools in ORN. The different predictors give different types of output (cf. table above); for example, FAME, the VU tool and SMARTcyp will give back SOMs related to P450- or other isoform specific conversions, whereas UU tools give predictions for phase I prediction in general. Hence, ORN can/should in first instance (stage (i) below) aid experts in guiding decision making on metabolite formation, by directly supplying output from the different available predictors. For that purpose the different partners first integrate the different tools in ORN, and we may subsequently consider to make a wrapper available to present the different results of the tools. This tool integration into ORN is intended to be described in a first manuscript along with a few showcases for selected compounds to demonstrate that having the different tools directly available helps in SOM prediction (by experts), for a few examples on metabolite formation as known from literature. VU will take the lead in example selection and may thereby benefit from some previous in-house efforts on annotation of databases for CYP-isoform specific conversions.

In a possible second stage we could think about combining tools (e.g. protein-structure based and reactivity based tools) for metabolite prediction (by non-experts?).

This could be summarized in the following sub-objectives:

i) Integration and evaluation for existing tools for (site-of-)metabolism prediction, to guide experts in the field in decision making on possible metabolites of substrates to be included in risk assessment:

1. Integration of tools for site-of-metabolism prediction
2. Run individual tools separately (baseline)
3. Demonstrate the integrated report has value
 - a. Combine the tools into an integrated report
 - b. Demonstrate that several tools has more value than individual tools.
[Manuscript?]
4. POSSIBLE FUTURE WORK: Contribute to SOM-predictor improvement
 - a. Create datasets for SOM benchmarking (expert annotates a few hundred reactions)
 - b. Make it possible to assess benchmark dataset using ORN infrastructure
 - c. Add additional (external) tools?

ii) Integration and combination of tools for metabolite prediction for expert and non-expert use within the risk assessment framework

REFERENCES

UM-PPS:

<https://academic.oup.com/bioinformatics/article/26/6/814/244321>

enviPath:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702869/>

<https://envipath.org/>

<http://pubs.rsc.org/en/Content/ArticleLanding/2017/EM/C6EM00697C#!divAbstract>

MetPred:

<https://metpred.service.pharmb.io/draw/>

<https://metpred.service.pharmb.io/>

XmetDB:

<https://jcheminf.springeropen.com/articles/10.1186/s13321-016-0161-3>

MetaPrint2D as a service in ORN:

<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-362>

Probabilistic predictions of Site-of-Metabolism:

<http://proceedings.mlr.press/v60/arvidsson17a.html>

FAME:

<https://www.ncbi.nlm.nih.gov/pubmed/24219364>

SMARTCyp:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4055970/>

Plasticity/docking (to P450 2D6):

<https://pubs.acs.org/doi/10.1021/jm801005m>