OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

Case Study

Metabolism Prediction [**MetaP**]

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 have worked together on making methods and tools available for metabolite and site-of-metabolism (SOM) prediction. For that purpose we integrated and used ligand-based metabolism predictors (e.g. MetPred, enviPath, FAME, SMARTCyp) and we incorporated protein-structure and -dynamics based approaches to predict SOMs by Cytochrome P450 enzymes (P450s). P450s 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. It is expected that this work will be continued after the end of the project to make services available for the prediction of microbial biotransformation pathways by integrating the enviPath data and software developed in part by JGU.

During method development, model calibration and validation we used databases such as 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 benefited of ongoing development in workflow management systems and we made Jupyter Notebooks available to facilitate collection and visualization of results from the different available services. We illustrated the added value of having multiple predictors and our Jupyter notebooks available, in a pilot study on retrospective consensus predictions of known SOMs for drug compounds for which possible metabolite-associated toxicity was previously reported.

DESCRIPTION

Implementation team

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

Case Study objective

The objective of this case study was to enable and facilitate metabolite prediction within the OpenRiskNet infrastructure and to evaluate and demonstrate the added value of it. For that purpose we integrated 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.

Combined use of the tools has been made possible and compared using Jupyter notebooks that gather and visualize results from the available case-study services.

See the “Databases and tools” subsection for more details on the corresponding tools. For our comparisons of predictive (and consensus) performance we used selected compounds from literature for which SOMs and metabolite-associated toxicity have been reported. We anticipate to present our results in an upcoming manuscript on tool integration, which will illustrate how using several tools can have additional value (when compared to individual tools) to (site-of-)metabolism prediction.

Risk assessment framework

Prediction outcomes can 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

Databases and tools

The table below gives an overview of metabolite prediction tools that are integrated and have been used in this case study. During method development, model calibration, and validation, advantage was taken of data from XMetDB (reference 1) and other databases for drugs, xenobiotics and their respective metabolites, as available in ZINC, ChEMBL, DrugBank, EAWAG-BBD and/or the SMARTCyp and FAME suites. Integration of enviPath (envipath.org) is still ongoing, which is a database and prediction system for microbial biotransformation of organic environmental contaminants.[2-4]

Table 1: Currently available MetaP tools.

Tool	Input	Output	Method
MetPred (UU)	2D chemical structure of ligand	SOMs with Reaction Types for Phase I reactions	Preprocess Metabolite reaction database (>100K biotransformations) using MCS. For each query compound, look up similar atom environments based on circular fingerprints and use ReactionSMARTS to identify reaction types. See metpred.service.pharmb.io/draw/
FAME 3 (HHU/ HITeC)	2D chemical structure of ligand	SOMs for Phase I, Phase II, or combined Phase I/II metabolism	Machine learning using 2D-circular-environment based atomic descriptors, see reference 5.
SMARTCyp 2.0 (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. See reference 6 and smartcyp.sund.ku.dk/mol_to_som
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 and corresponding SOM in the active site of CYP isoforms (1A2, 2D6, 3A4). Cf. reference 7.

Technical implementation

As summarised in Table 1, several services have come available in the MetaP case study. The listed services offer their functionality through RESTful APIs that are formalised according to OpenAPI specifications. The APIs are build using the Swagger toolchain and subsequently enable direct user interaction with the API endpoints using a browser-based User Interface (the Swagger UI). In addition, MetPred and SMARTCyp offer a custom

browser-based interface to their service (see links in Table 1). The APIs enable access to the core features of the services as summarised above, and typically accept submissions of chemical structures in common file formats.

API endpoint input and output data exchange is standardised to a machine-readable JSON format. Together with the OpenAPI data type definitions and JSON-LD data annotation it ensures seamless integration of the containerised services in the OpenRiskNet infrastructure and data exchange with other services.

Service API use and interoperability of the listed services is demonstrated using a Jupyter Notebook freely available on Github. Single 3D ligand structures in Tripos MOL2 format are used as input to the various services and the standardised JSON output are aggregated into a Pandas DataFrame demonstrating interoperability. Predicted SOMs are visualized on the 2D ligand depiction using the RDKit package.

OUTCOMES

In addition to the service integration of the metabolite prediction tools listed above, we have evaluated the added value of having multiple tools and their combined use available (via Jupyter Notebooks). The different predictors give complementary types of output, *cf.* Table 1. MetPred, FAME 3, and the VU and SMARTCyp tools predict SOMs related to Phase I, Phase I/II, and Cytochrome P450 isoform specific conversion, respectively. Per (heavy) atom, normalized propensities are written out to indicate the likelihood of the atom to be a SOM. In addition, MetPred also gives back most probable reaction types at predicted SOMs. Facilitated by the Jupyter Notebook that supplies and visualizes output from the different predictors (Appendix 1), the MetaP tools can thus aid experts in guiding decision making on metabolite formation and/or in obtaining input for subsequent case studies.

The added value of having the multiple complementary tools available for metabolite prediction is illustrated by the Jupyter-notebook output presented in Appendix 1, which collects SOM predictions and MetPred predictions of Phase I reaction types (and which color-highlights atoms as predicted SOM if propensities are larger than a preset cutoff) for the three compounds in Figure 1. These compounds were selected because possible toxicological effects have been related with their metabolites, and their metabolism is extensively studied in literature (see Figure 1 for the experimentally determined SOMs).[8-10]

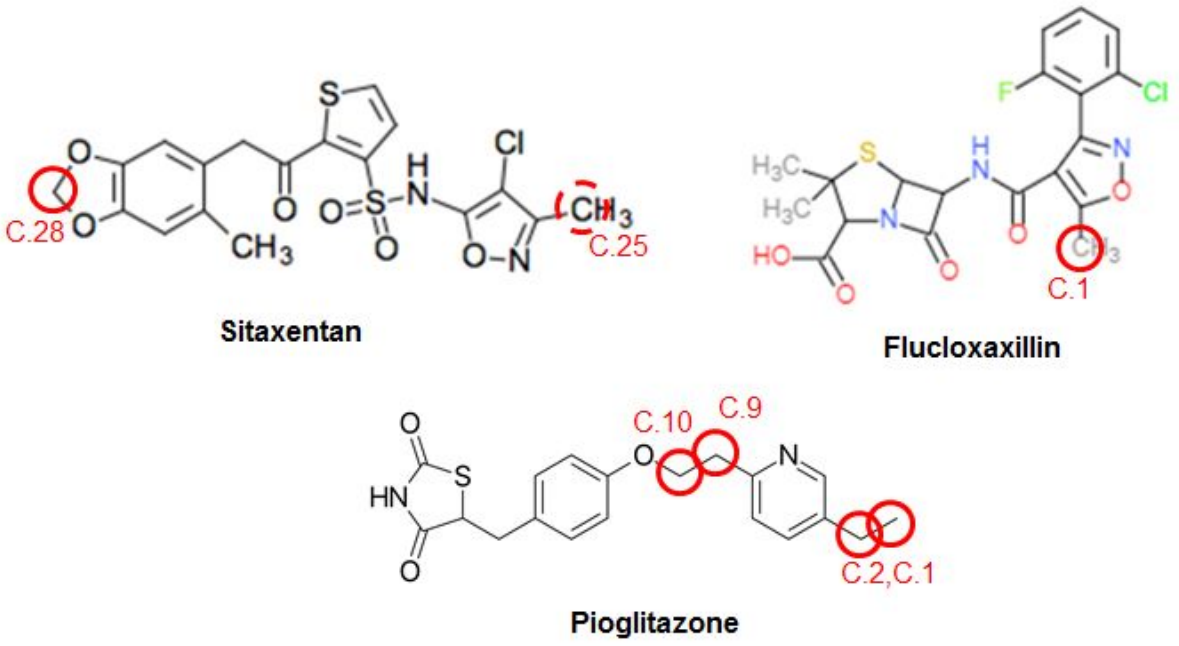


Figure 1. Molecular structures of Sitaxentan, Pioglitazone and Flucloxacillin, together with their experimentally determined sites-of-metabolism[8-10] as indicated by circles and atom indices.

Appendix 1 demonstrates that for Sitaxentan, consensus is obtained with the different tools in (correctly) predicting C.28 as SOM: FAME 3 and SMARTCyp 2.0 assign it as the most probable metabolic site, while docking confirms a possible reactive binding orientation in CYP (3A4). SMARTCyp also appoints C.25 as reactive and a possible SOM, which was identified in metabolism studies with dog liver microsomes.[8] In addition, the more mechanistic based SMARTCyp and docking tools can help in identifying the SOM-assignment by MetPred of O.17/18 as a false positive. Similarly, the zero scores of SMARTCyp and docking identify the predictions of MetPred and FAME 3 for C.23 of Pioglitazone to be a false positive as well, Appendix 1. For this compound, consensus is reached that C.9 and C.10 are possible sites of metabolism, while the current unavailability of a docking model for P450 2C8 may well partly explain why not all tools identify C.1 and C.2 both as possible SOM (in that case 2C8 is known to be the major involved P450 isoform[9]). It should also be noted that based on the significant scores for three out of four predictors, experts may (wrongly) assign Pioglitazone's S.25 as a possible SOM as well. As a third example, Appendix 1 illustrates the obtained consensus in correctly assigning C.1 of Flucloxacillin as the most probable metabolic site.

In conclusion, these examples illustrate how combining and comparing output from the different tools available in MetaP (and how collecting and visualizing their output in a Jupyter Notebook) can aid in and increase the value of SOM prediction when compared to having individual tools available alone.

REFERENCES

1. O. Spjuth, P. Rydberg, E.L. Willighagen, C.T. Evelo, N. Jeliazkova, *J. Cheminf.* **2016**, *8*, 47.
2. J. Wicker, K. Fenner, L. Ellis, L. Wackett, S. Kramer, *Bioinformatics* **2010**, *26*, 814-821.
3. J. Wicker, T. Lorsbach, M. Gütlein, E. Schmid, D.A.R.S. Latino, S. Kramer, K. Fenner, *Nucleic Acids Res.* **2016**, *44*, D502-D508.
4. D.A.R.S. Latino, J. Wicker, M. Gütlein, E. Schmid, S. Kramer, K. Fennerad, *Environ. Sci.: Processes Impacts* **2017**, *19*, 449-464
5. M. Sicho, C. Stork, A. Mazzolari, C. de Bruyn Kops, A. Pedretti, B. Testa, G. Vistoli, D. Svozil, J. Kirchmair, *J. Chem. Inf. Model.* **2019**, *59*, 3400-3412.
6. P. Rydberg, D.E. Gloriam, J. Zaretzki, C. Breneman, L. Olsen, *ACS Med. Chem. Lett.* **2010**, *1*, 96-100.
7. J. Hritz, A. de Ruiter, C. Oostenbrink, *J. Med. Chem.* **2008**, *51*, 7469-7477.
8. J.C.L. Erve, S. Gauby, J.W. Maynard Jr., M.A. Svensson, G. Tonn, K.P. Quinn, *Chem. Res. Toxicol.* **2013**, *26*, 926-936.
9. T. Jaakkola, J. Laitila, P.J. Neuvonen, J.T. Backman, *Basic Clin. Pharmacol. Toxicol.* **2006**, *99*, 44-51.
10. R.E. Jenkins, X. Meng, V.L. Elliott, N.R. Kitteringham, M. Pirmohamed, B.K. Park, *Proteomics Clinic. Applic.* **2009**, *3*, 720-729.

APPENDIX

Example Jupyter notebook and predicted SOMs

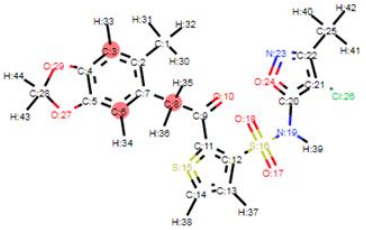
In [1]:
```
%%javascript
IPython.OutputArea.prototype._should_scroll = function(lines) {
    return false;
}
```

In [2]:
```
import os
import requests
import pandas

from rdkit import Chem
from rdkit.Chem.Draw import rdMolDraw2D
from IPython.display import HTML
from example_helpers import *

MDSTUDIO_URL = 'http://mdstudio-smartcyp.dev.openrisknet.org/'
FAME_URL = 'http://fame3.dev.openrisknet.org'
METPRED_URL = 'http://metpred.prod.openrisknet.org/v2'

data = get_dataset()
```

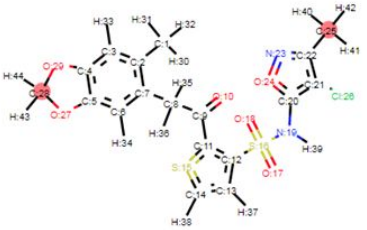
SOM prediction comparing Docking, SMARTCyp, FAME and MetPred

In [3]:
```
selection = ('sitaxentan', 'pioglitazone', 'flucloxacillin')
for case_name in selection:

    case_data = data[case_name]

    # Get SMARTCyp prediction
    response = requests.post('{0}/som_prediction'.format(MDSTUDIO_URL),
                             files={'ligand_file': case_data['mol2']}, data={'cyp': '3A4'})
    response_df = pandas.DataFrame.from_dict(response.json(), orient='index')

    # Get FAME3 prediction
    response = requests.post('{0}/predictFromFiles'.format(FAME_URL), files={'files': case_data['mol']},
                             data={'includeMDL': False}, headers={"accept" : "application/json"})
    response_df, famecutoff = process_fame_results(response.json(), response_df)

    # Get MetPred prediction
    response = requests.get('{0}/prediction'.format(METPRED_URL), params={'compound': case_data['mol']})
    response_df, metpredcutoff = process_metpred_results(response.json(), response_df)

    # Draw molecule
    response_df_prob = response_df[['Docking', 'SMARTCyp', 'FAME', 'MetPred', 'MetPred reaction']].fillna(0)
    rdmol = Chem.rdmolfiles.MolFromMol2File('{0}.mol2'.format(case_name), sanitize=False)
    dock_svg = show_predictions(rdmol, response_df_prob['Docking'])
    smart_svg = show_predictions(rdmol, response_df_prob['SMARTCyp'])
    fame_svg = show_predictions(rdmol, response_df_prob['FAME'], cutoff=famecutoff)
    metpred_svg = show_predictions(rdmol, response_df_prob['MetPred'], cutoff=metpredcutoff)

    # Display Pandas Dataframe and 2D depictions
    display(HTML('<h3 align="center">{0}</h3>'.format(case_name)))
    display(HTML(no_wrap_div.format(dock_svg, smart_svg, fame_svg, metpred_svg)))
    display(style_dataframe(response_df_prob, fame_cutoff=famecutoff))
```

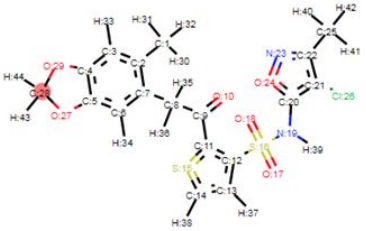
sitaxentan



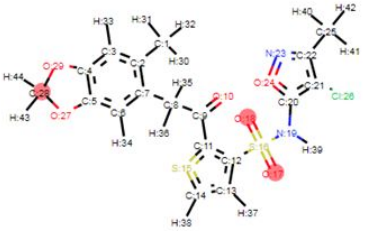
Docking



SMARTCyp



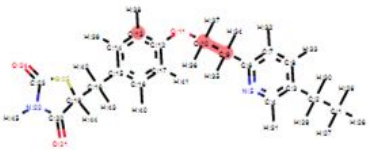
FAME 3



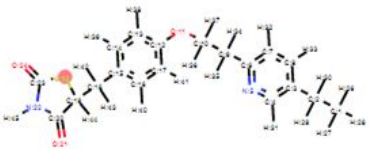
MetPred

	Docking	SMARTCyp	FAME	MetPred	MetPred reaction
C.1	0.714	0.73	0.224	0	0
C.11	0	0	0.0596	0	0
C.12	0	0	0.052	0	0
C.13	0.214	0.621	0.107	0	0
C.14	0.214	0.699	0.176	0	0
C.2	0	0	0.004	0	0
C.20	0	0	0.056	0	0
C.21	0	0	0.08	0	0
C.22	0	0	0.184	0	0
C.25	0	0.81	0.304	0	0
C.28	0.214	1	0.892	0.179	O-dealkylation; alkyl hydroxylation
C.3	1	0.6	0.0276	0	0
C.4	0	0	0.004	0	0
C.5	0	0	0.008	0	0
C.6	1	0.6	0.02	0	0
C.7	0	0	0.004	0	0
C.8	0.786	0.73	0.148	0	0
C.9	0	0	0.176	0	0
CL.26	0.0714	0	0.048	0	0
N.19	0.0714	0.514	0.036	0	0
N.23	0	0.527	0.251	0	0
O.10	0	0	0.0516	0	0
O.17	0	0	0.0836	1	oxidative desulfuration
O.18	0	0	0.0836	1	oxidative desulfuration
O.24	0	0	0.152	0	0
O.27	0	0	0.004	0	0
O.29	0	0	0	0	0
S.15	0.357	0.693	0.167	0	0
S.16	0.643	0	0.052	0	0

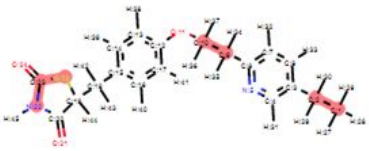
pioglitazone



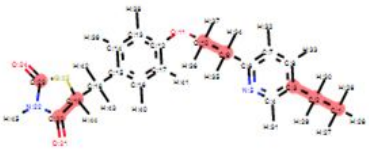
Docking



SMARTCyp



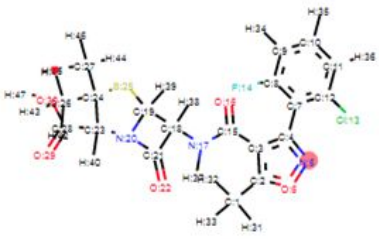
FAME 3



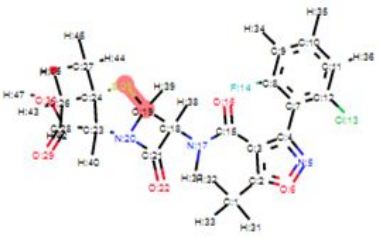
MetPred

	Docking	SMARTCyp	FAME	MetPred	MetPred reaction
C.1	0	0.385	0.788	0.626	carboxylation; alkyl hydroxylation
C.10	1	0.555	0.728	0.933	alkyl hydroxylation; O-dealkylation
C.12	0	0	0	0	0
C.13	0.833	0.447	0.008	0	0
C.14	0.722	0.4	0	0	0
C.15	0	0	0	0	0
C.16	0.611	0.4	0	0	0
C.17	0.556	0.447	0.008	0	0
C.18	0.611	0.52	0.004	0	0
C.19	0.667	0.598	0.204	0.616	sulfide oxidation; alkyl hydroxylation
C.2	0	0.52	0.876	1	alkyl hydroxylation
C.20	0	0	0	0.423	alkyl hydroxylation; secondary amide hydrolysis (keep CO)
C.23	0	0	0.86	0.959	alkyl hydroxylation; sulfide oxidation; amide hydrolysis (keep amine)
C.3	0	0	0	0.207	aromatic hydroxylation
C.4	0.222	0.375	0.004	0	0
C.6	0	0	0.012	0	0
C.7	0.556	0.4	0	0	0
C.8	0.389	0.4	0.008	0	0
C.9	0.944	0.52	0.728	0.778	alkyl hydroxylation
N.22	0	0.385	0.844	0	0
N.5	0.333	0.456	0.032	0	0
O.11	0	0	0.016	0	0
O.21	0	0	0	0	0
O.24	0	0	0	0	0
S.25	0.5	1	0.996	0	0

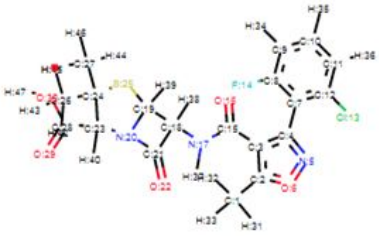
flucloxacillin



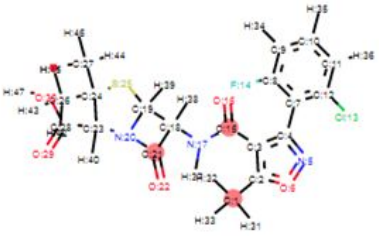
Docking



SMARTCyp



FAME 3



MetPred

	Docking	SMARTCyp	FAME	MetPred	MetPred reaction
C.1	0.571	0.741	0.216	0.703	alkyl hydroxylation
C.10	0	0.55	0.152	0	0
C.11	0	0.528	0.04	0	0
C.12	0	0	0	0	0
C.15	0	0	0.112	0.346	amide hydrolysis (keep amine); amine dehydrogenation
C.19	0.429	0.769	0.032	0	0
C.2	0	0	0.06	0	0
C.21	0	0	0.156	1	amide hydrolysis (keep amine); tertiary amide hydrolysis (keep CO); alkyl hydroxylation
C.24	0	0	0.016	0	0
C.26	0.571	0.496	0.04	0	0
C.27	0.143	0.496	0.04	0	0
C.28	0	0	0.012	0	0
C.3	0	0	0.02	0	0
C.4	0	0	0.0589	0	0
C.7	0	0	0	0	0
C.8	0.643	0	0.004	0	0
C.9	0	0.528	0.04	0	0
CA.18	0.286	0.695	0.008	0	0
CA.23	0.286	0.669	0.012	0	0
CL.13	0.357	0	0	0	0
F.14	0.429	0	0	0	0
N.17	0.714	0.496	0.016	0	0
N.20	0.5	0.496	0.02	0	0
N.5	1	0.482	0.196	0	0
O.16	0	0	0.0276	0	0
O.22	0	0	0.012	0	0
O.29	0	0	0	0	0
O.6	0	0	0.0636	0	0
OXT.30	0	0	0.052	0	0
S.25	0.286	1	0.048	0	0