Funded by the Seventh Framework Programme of the European Union

Project full title:

Hepatic and Cardiac Toxicity Systems modelling

Project acronym:

HeCaTos

Collaborative project
HEALTH.2013.1.3.-1:
Modelling toxic response in case studies for predictive human safety assessment

FP7-HEALTH-2013-INNOVATION-1-602156-HeCaTos

Deliverable Report D1.6: Report on integrated system to predict interactions of compound metabolites with AOPs

Work package 1

Due date of deliverable: M30
Actual submission date: M33

Start date of project: October, 2013 Duration: 60 months

Maastricht University (UM)

Project co-funded by the European Commission within the 7th Framework Programme (2013-2018)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

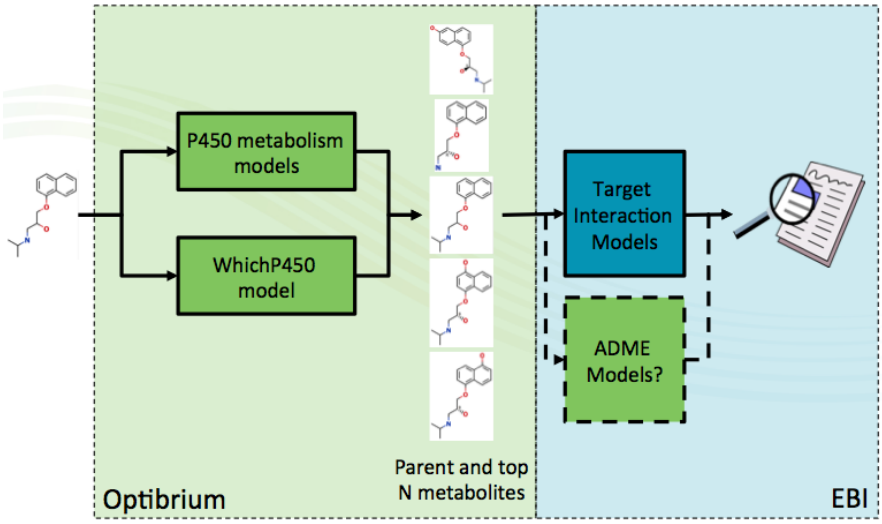
Contributions to deliverable - Internal review procedure

Deliverable produced by:	Date:
Francis Atkinson - Partner EMBL-EBI	April 2016
Peter Hunt - Partner Optibrium	April 2016
Anne Hersey - Partner EMBL-EBI	May 2016
Matt Segall - Partner Optibrium	June 2016
Deliverable internally reviewed by:	**Date:**
Jos Kleinjans - Partner UM	June 2016

Contents

PUBLISHABLE SUMMARY

A system has been created that enumerates metabolites for an input compound and generates predictions of the activities of these species against a panel of toxicity-associated biomolecular targets, activity against these targets representing possible interactions with Adverse Outcome Pathways.



It is intended that the system be extensible, with modules implementing further predictive methods, such as PBPK/ADME models, being added as they become available. The system may be accessed directly from Python or *via* a simple web-service.

OBJECTIVES

The objective of this deliverable is to produce an integrated system to predict interactions of compound metabolites with Adverse Outcome Pathways (AOPs). This will involve coupling a module that enumerates metabolites for an input structure with one that predicts the activities of the parent molecule and its metabolites with biomolecular entities associated with the initiating events of AOPs.

1. INTRODUCTION

The concepts of Adverse Outcome Pathways (AOP) and their associated Molecular Initiating Events (MIE) are becoming well established in the field of toxicology [1, 2]. An MIE is the interaction of a xenobiotic, here taken to be a small molecule, with a biomolecular target, which initiates a well-defined chain of events (the pathway) that result in a toxic phenotype (the adverse outcome). Thus, identifying molecules that might interact with these targets could, in principle, allow us to predict potential toxicities in advance.

It is well known that metabolites of dosed drugs may also be bioactive. Indeed, of the HeCaToS compound set, several are prodrugs that require metabolism for activity [3]. For example, cyclophosphamide undergoes an initial P450-mediated hydroxylation, which is followed by non-enzymatic rearrangements to give the active moieties [4]. Carisprodol is metabolised by CYP2C19 to meprobromate, itself an approved drug [5]; while both species are active, the metabolite has a longer duration of action and is therefore responsible for most of the drug's effects.

Simvastatin is particularly interesting in this context as its effects are likely to be produced by a mixture of species. The dosed drug is a lactone that is hydrolysed *in vivo* to give the active β-hydroxy acid [6]. Both lactone and acid undergo metabolism by CYP450s [7]; various metabolites have sufficient on-target activity that it can be necessary to take account of 'total inhibitor' concentration when monitoring the effects of the drug [8].

If metabolites can have on-target, therapeutic activities such as those described above, they may also have undesirable, off-target activities. In principle, if predicting the off-target effects of a dosed compound can contribute to evaluating its risk profile, then including the activities of any metabolites in the assessment should give a better understanding of any potential hazard.

Thus, a tool coupling prediction of the set of potential metabolites of a molecule with a prediction of the possible activities of those metabolites would have value in risk assessment. The aim of this report is to describe work done at the EMBL-EBI and Optibrium that begins to address this issue. It should be noted that both aspects of this problem are very challenging, and some compromises are necessary in order to make the project feasible. These are discussed further below.

One important point is that only P450-mediated Phase I metabolism is currently being considered. The CYP450s are the most important class of enzyme involved in the metabolism of drug-like compounds [9], and Optibrium have a well-established methodology that can provide a quantitative estimate of the reactivity of each site in a molecule for the most important human CYP450 isoforms [10]. A more recent addition to their suite predicts which isoforms are likely to be the major participants in the metabolism of a compound [11].

While CYP450s are dominant, non-CYP enzymes can play an important role [12, 13] and, in principle, it would be desirable to include predictions for them. However, while there are tools that could provide these [14], we do not currently have access to them. Thus, predictions for other enzyme classes could be incorporated into the framework described here if and when suitable tools become available to us.

Similarly, no attempt has been made currently to address the conjugated species produced by Phase II metabolism, such as glucuronides, although these have been known to cause problems [15, 16]. Again, this is something that could be investigated in future iterations.

For the CYP450s, mechanism-based inactivation of the enzyme can be a problem [17], leading, for example, to drug-drug interactions. Furthermore, the modification of proteins or DNA by reactive species produced by CYP450s or other xenobiotic metabolizing enzymes can lead to autoimmune

responses or mutagenicity [18, 19]. The organic chemistry governing these reactions is well understood and may, in principle, be used to derived structural alerts, patterns in a molecule that flag them as being of concern [20, 21]. However, while such alerts can be useful, there are issues with balancing the sensitivity and specifity of the patterns used: for example, it is easy to naively create patterns that generate too many false positives. It has been decided to defer addition of this feature until these issues could be addressed.

Although some AOPs have been described in great detail [22], and many others are being constructed [23], full coverage of the spectrum of human toxicities with AOPs remains a goal rather than a reality. In this work, we use the reference set of toxicology-associated targets described in HeCaToS Deliverable D9.2 [24] as a proxy for the AOPs of interest. In summary, this is a set of 'antitargets', human proteins for which evidence exists that their perturbation by xenobiotics might give rise to a toxicological response [25].

The interaction of xenobiotics with this panel of antitargets is modelled here using a simple, structure-based machine learning approach. This, in summary, involves encoding a chemical structure as a so-called 'fingerprint', which allows it to be statistically related to antitarget activities taken from the ChEMBL database [26, 27]; this work is described in detail in HeCaToS Deliverable D1.5 [28].

Note that in the present case, only a qualitative prediction is attempted: the classifier returns a list of those targets in the model against which a structure is believed likely to be active. The panel of models has been rebuilt as part of this work, using data from the most recent release of the database (CHEMBL_21) [29]; an updated edition of the D1.5 report documenting the model rebuilding process is being prepared.

It should be noted that the modelling, at present, only attempts to handle reversible, non-covalent interactions with the antitargets in the panel. As noted above, covalent modification of biomolecules by reactive metabolites can be important but is not currently handled by this system. This is, however, an objective of future work under HeCaTos Deliverable D1.4.

There are other processes that can initiate adverse outcomes of interest to HeCaToS that are not yet addressed, important examples being redox cycling and mitochondrial uncoupling [30]. Redox cycling is believed to be particularly important in the toxicity of anthracyclines such as Doxorubicin, Daunorubicin and Idarubicin, which are key HeCaToS compounds [31], and mitochondrial toxicity in general is of course of importance to HeCaToS [32].

A tool to assess the ability of compounds to undergo redox cycling and/or uncoupling would be thus useful, and various approaches involving quantum mechanical calculations and/or structural alerts have been described in the literature [33-35]. The feasibility of implementing these approaches will be investigated further.

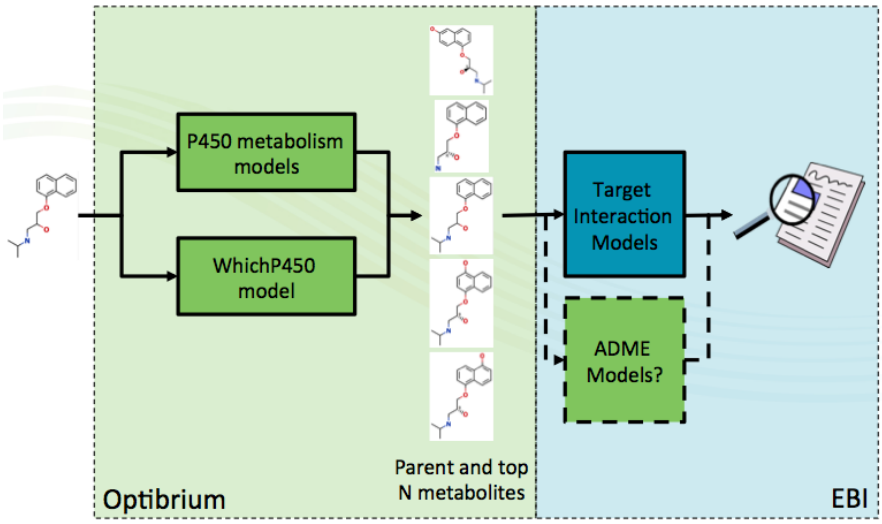
Another improvement to the system would be the incorporation of ADME/PBPK models that would provide an estimate of the concentration-time profile of metabolites in target organs [36]. This could, for example, show that a predicted antitarget interaction need not be of concern as the exposure of the species involved was not sufficient to elicit a toxic response [37].

A closely related area that is yet to be addressed but which can be of great practical importance is that of pharmacogenomics: population variances in xenobiotic metabolism due to polymorphism in genes coding for key enzymes, differences in gene expression with age and diet *etc.* [38]. Tools are available that address this in the ADME/PBPK context [39], and, although we do not currently have access to them, they could in principle be incorporated in future iterations.

2. RESULTS

The design of the system is straightforward. In summary, a chemical structure, which could be an existing or virtual compound, is supplied to the system, for example *via* a web-service call. The Optibrium module enumerates the metabolites, and these are then passed to the target interaction module. The metabolites and their activity predictions are then returned to the caller. A schematic representation of the system is shown in Figure 1.

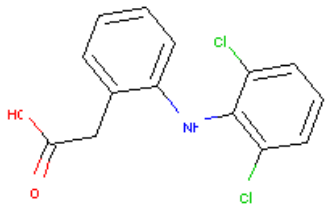
Figure 1



The system has been implemented as a Python package, which may be installed and run locally or installed on a server and accessed *via* a simple web service wrapper. More details are given below.

The Optibrium software produces predictions for a set of human CYP450 isoforms that together account for approximately 90% of observed xenobiotic metabolism: CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and CYP3A4. The data returned for an input compound has various components, including regioselectivity of metabolism predictions and isoform probability estimates, diagnostic information *etc*. Three key pieces of information have been used in the current project, and these are described below. Unless otherwise stated, Diclofenac is used in the examples (Figure 2).

Figure 2



First is the 'WhichP450' information, which lists the probabilities that the structure is metabolised by each isoform [11]. At what probability level an isoform becomes relevant in the metabolism of a compound is obviously somewhat arbitrary. Here, an isoform is flagged as 'likely' to metabolise a compound if the probability is at least one third of that of the most probable isoform. Only data for these likely isoforms are returned in the final report. An example of the WhichP450 data for Diclofenac is shown in Table 1.

Table 1

isoform	probability	likely
CYP2C9	0.34	true
CYP3A4	0.25	true
CYP1A2	0.15	true
CYP2C19	0.13	true
CYP2D6	0.06	false
CYP2E1	0.04	false
CYP2C8	0.04	false

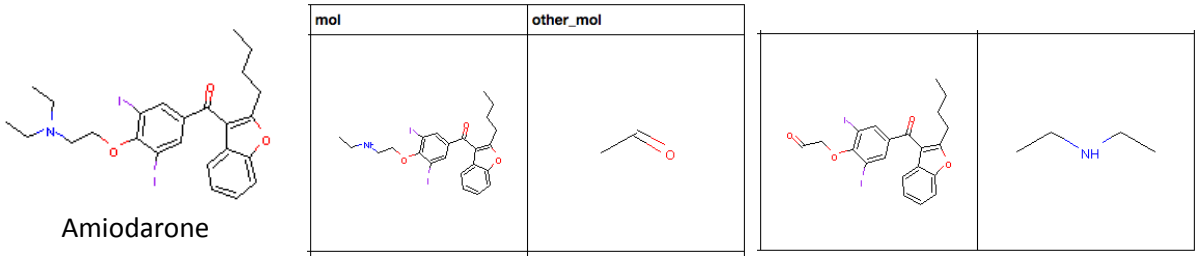
An enumeration of the possible metabolites, based on the identification of plausible sites of CYP-mediated hydroxylation, is also generated [10]. A table listing the estimated fraction of metabolism at each possible site on the parent molecule produced by each isoform is provided (the so-called regioselectivity or 'R' values).

The final component is a categorical prediction of the 'lability' of each site of metabolism, an assessment of the relative efficiency of metabolite formation at that site relative to the decoupling pathways for the CYP [10]. Ideally, this would be done for all isoforms; however, in practice, accurate literature data for the decoupling pathway is only available to parameterise the method for CYP3A4. It is estimated that the reaction rates for other isoforms will be fairly similar to that for CYP3A4, and so the CYP3A4 prediction is used as a proxy for the complete panel. There are four categories: Stable, Moderately Stable, Moderately Labile and Labile. Currently, only those metabolites produced by sites that are categorised as Labile or Moderately Labile are considered.

The complete set of data for all isoforms for Diclofenac is shown in Table 2; the subset taken forward after taking only the likely isoforms from the 'WhichP450' prediction and only the metabolites from the lability prediction is shown in Table 3. In the latter, some extra information such as the identity of the hydroxylated atom is included; note that the 'sym_class' column shows whether the affected atom is one of a set of equivalent, symmetry-related atoms.

Also included in Table 3 is the column 'alert', which simply indicates whether an aldehyde functional group is present in the metabolite. Aldehydes can be produced by decomposition of a hemiacetal or hemiaminal produced by hydroxylation of a methylene group adjacent to an alkoxy or alkylamine moiety. Where this leads to a fragmentation of the molecule, the largest fragment is returned as the metabolite (*i.e.* 'mol') and any smaller fragment(s) returned in a column named 'other_mol'. This does not occur for Diclofenac, but does with Amiodarone; see Figure 3 for an illustration of this.

Figure 3



In some cases, no ‘Labile’ or ‘Mod labile’ sites are identified. Where this occurs, if there are any ‘Mod stable’ sites where the R-value is above 50%, then the single metabolite with the highest R-value is returned; If no metabolites meet this criterion then none are returned.

This step is performed to be certain that nothing is missed: however, the criterion is somewhat arbitrary and this step could be made optional. As noted above, not all drugs are metabolised by CYP450s [9], so the lack of a metabolite here does not necessarily imply an error has been made.

Table 2

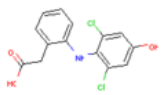
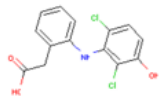
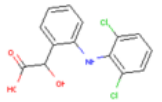
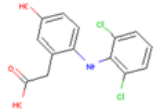
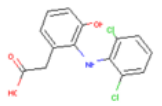
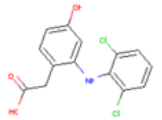
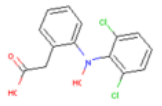
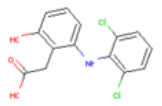
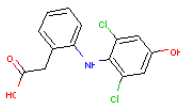
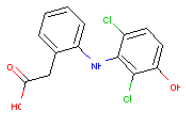
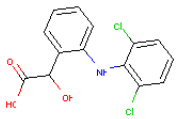
	mol	lability	CYP2C9_R	CYP3A4_R	CYP1A2_R	CYP2C19_R	CYP2D6_R	CYP2E1_R	CYP2C8_R
0		Mod Labile	66.12	38.02	40.89	32.30	8.15	60.21	87.71
1		Mod Labile	12.39	13.54	4.20	6.55	2.35	6.16	5.01
2		Mod Labile	5.19	30.87	47.43	50.15	83.36	24.52	1.13
3		Mod Stable	3.84	3.45	3.24	4.35	3.57	2.92	1.13
4		Stable	0.05	0.29	0.02	0.06	0.14	0.02	0.00
5		Stable	0.02	0.01	0.01	0.01	0.01	0.01	0.00
6		Stable	0.01	0.27	0.00	0.01	0.06	0.00	0.00
7		Stable	0.00	0.01	0.00	0.00	0.01	0.00	0.00

Table 3

	mol	element	atom_id	sym_class	alert	lability	CYP2C9_R	CYP3A4_R	CYP1A2_R	CYP2C19_R
0		C	16	16	False	Mod Labile	66.12	38.02	40.89	32.30
1		C	15	15,17	False	Mod Labile	12.39	13.54	4.20	6.55
2		C	4	4	False	Mod Labile	5.19	30.87	47.43	50.15

Once the metabolites of interest have been obtained, they are passed to the target interaction module. Here they are standardised, so that *e.g.* functional group representations are the same as those used in the model building [40].

These normalised structures are then passed to the classifier, which makes qualitative prediction of activity for the panel of toxicology associated targets [24, 28]. A list of the HUGO gene symbols for the targets at which the structure is predicted to be active is returned [41]. Predicted activities for the parent compound are also returned.

As well as the activity predictions, ancillary data such as the canonical SMILES for the normalised structures used in the calculation is returned. Exactly what metadata are required remains to be determined, and could easily be altered in future iterations.

As noted, the system may be used locally as a Python module or *via* a simple JSON-based web-service. Advantages of the former approach might include its bundling as part of a larger system, while those of the latter include the ability to interface with workflow tools such as KNIME [42].

An example session using the module directly in a Jupyter Notebook is show in Figure 4 [43]; note that verbose logging from the module has been enabled and is rendered with a pink background by Jupyter. Such output can be particularly useful when extending or debugging the module.

The JSON data structure returned is designed to make it simple to create a Pandas dataframe, a tabular format ideal for visualisation and further processing [44]. A rendering of such a dataframe may be seen at the bottom of the figure: the activity predictions for the metabolites are contained in the highlighted column ‘actives’.

An example session using the module *via* the web service is shown in Figure 5, this time using Amiodarone as the input compound. The data structure returned by the module is passed back *via* the web service, again allowing easy recreation of the dataframe. The activity predictions for the parent may be seen at the bottom of the figure in this example.

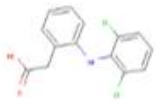
Figure 4

diclofenac

```
In [40]: import prediction
         prediction.logger.setLevel('DEBUG')

In [41]: smiles = 'O=C(O)Cc1ccccc1Nc1c(Cl)cccc1Cl'
         Chem.MolFromSmiles(smiles)
```

Out[41]:



```
In [42]: result = prediction.run(smiles)
```

```
(05/05/16 20:05:17 prediction DEBUG] Input SMILES: 'O=C(O)Cc1ccccc1Nc1c(Cl)cccc1Cl'
(05/05/16 20:05:17 prediction DEBUG] Parent mol SMILES: 'O=C(O)Cc1ccccc1Nc1c(Cl)cccc1Cl', activities: ''
(05/05/16 20:05:17 prediction DEBUG] Mol level data...
```

	param	value
0	MF	-53.5
1	IP	8.6
2	A2	0.0

```
(05/05/16 20:05:17 prediction INFO] WhichP450 prediction...
```

	isoform	probability	likely
3	CYP2C9	0.3	True
6	CYP3A4	0.2	True
0	CYP1A2	0.2	True
1	CYP2C19	0.1	True
4	CYP2D6	0.1	False
5	CYP2E1	0.0	False
2	CYP2C8	0.0	False

```
(05/05/16 20:05:17 prediction DEBUG] CNL data...
```

	param	value
0	Value	0.7
1	Uncertainty	0.3

```
(05/05/16 20:05:17 prediction DEBUG] All metabolites, sorted on degree of lability and decreasing probability of products for the most likely isoform ...
```

	smiles	element	atom_id	sym_class	alert	lability	CYP2C9_R	CYP3A4_R	CYP1A2_R	CYP2C19_R	CYP2D6_R	CYP2E1_R	CYP2C8_R
0	OC(=O)CC1=CC=CC=C1NC2=C(Cl)C=CC=C2Cl)O	C	16	16	False	Mod Labile	66.1	38.0	40.9	32.3	8.2	60.2	87.7
1	OC(=O)CC1=CC=CC=C1NC2=C(Cl)C(=CC=C2Cl)O	C	15	15,17	False	Mod Labile	12.4	13.5	4.2	6.6	2.3	6.2	5.0
2	OC(=O)C(Cl=CC=CC=C1NC2=C(Cl)C=CC=C2Cl)O	C	4	4	False	Mod Labile	5.2	30.9	47.4	50.1	83.4	24.5	1.1
3	OC(=O)CC1=CC(=CC=C1NC2=C(Cl)C=CC=C2Cl)O	C	7	7	False	Mod Stable	3.8	3.4	3.2	4.4	3.6	2.9	1.1
4	OC(=O)CC1=CC=CC(=C1NC2=C(Cl)C=CC=C2Cl)O	C	9	9	False	Stable	0.0	0.3	0.0	0.1	0.1	0.0	0.0
5	OC(=O)CC1=CC=C(C=C1NC2=C(Cl)C=CC=C2Cl)O	C	8	8	False	Stable	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	OC(=O)CC1=CC=CC=C1N(C2=C(Cl)C=CC=C2Cl)O	N	11	11	False	Stable	0.0	0.3	0.0	0.0	0.1	0.0	0.0
7	OC(=O)CC1=C(C=CC=C1NC2=C(Cl)C=CC=C2Cl)O	C	6	6	False	Stable	0.0	0.0	0.0	0.0	0.0	0.0	0.0

```
(05/05/16 20:05:17 prediction DEBUG] After filtering sites and isoforms...
```

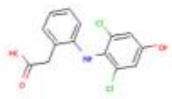
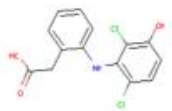
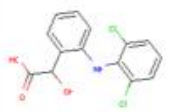
	smiles	element	atom_id	sym_class	alert	lability	CYP2C9_R	CYP3A4_R	CYP1A2_R	CYP2C19_R
0	OC(=O)CC1=CC=CC=C1NC2=C(Cl)C=CC=C2Cl)O	C	16	16	False	Mod Labile	66.1	38.0	40.9	32.3
1	OC(=O)CC1=CC=CC=C1NC2=C(Cl)C(=CC=C2Cl)O	C	15	15,17	False	Mod Labile	12.4	13.5	4.2	6.6
2	OC(=O)C(Cl=CC=CC=C1NC2=C(Cl)C=CC=C2Cl)O	C	4	4	False	Mod Labile	5.2	30.9	47.4	50.1

```
(05/05/16 20:05:17 prediction DEBUG] Metabolite data table to return...
```

	alert	lability	CYP2C9_R	CYP3A4_R	CYP1A2_R	CYP2C19_R	smiles	other_smiles	activities
0	False	Mod Labile	66.1	38.0	40.9	32.3	O=C(O)CC1=CC=CC=C1NC2=C(Cl)CC(O)=CC2=C1Cl		
1	False	Mod Labile	12.4	13.5	4.2	6.6	O=C(O)CC1=CC=CC=C1NC2=C(Cl)C=C(O)C=C2=C1Cl		NR1H3;PTGS2
2	False	Mod Labile	5.2	30.9	47.4	50.1	O=C(O)C(Cl)=C1C=CC=C1NC2=C(Cl)C=CC=C2=C1Cl		

```
In [43]: df = pandas.DataFrame(result['metabolite_data'])
         for col in ['smiles', 'other_smiles']: df[col.replace('smiles', 'mol')] = df[col].apply(lambda x: Chem.MolFromSmiles(x) if x else None)
         df
```

Out[43]:

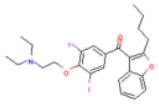
	CYP1A2_R	CYP2C19_R	CYP2C9_R	CYP3A4_R	activities	alert	lability	other_smiles	smiles	mol	other_mol
0	40.89350	32.30360	66.11870	38.0247		False	Mod Labile	None	O=C(O)Cc1ccccc1Nc1c(Cl)cc(O)cc1Cl		None
1	4.20435	6.55345	12.38950	13.5398	NR1H3;PTGS2	False	Mod Labile	None	O=C(O)Cc1ccccc1Nc1c(Cl)ccc(O)c1Cl		None
2	47.42900	50.14990	5.18583	30.8658		False	Mod Labile	None	O=C(O)C(Cl)c1ccccc1Nc1c(Cl)cccc1Cl		None

More examples of the use of the module and web service, including topics such as error handling and the use of different HTTP request types, are provided as a set of Jupyter Notebooks included with the package.

Figure 5

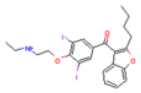
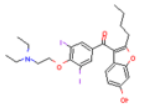
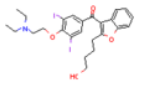
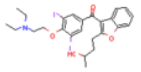
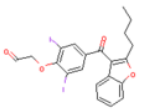
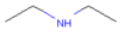
amiodarone

```
smiles = 'CCCCc1oc2ccccc2c1C(=O)c1cc(I)c(OCCN(CC)CC)c(I)c1'
Chem.MolFromSmiles(smiles)
```



```
import requests
from urllib.parse import quote as urlencode
URL = 'http://neo4j-vm.windows.ebi.ac.uk:8002'
```

```
response = requests.get(URL + '/smiles/' + urlencode(smiles))
result = response.json()
df = pandas.DataFrame(result['metabolite_data'])
for col in ['smiles', 'other_smiles']: df[col.replace('smiles', 'mol')] = df[col].apply(lambda x: Chem.MolFromSmiles(x) if x else None)
df
```

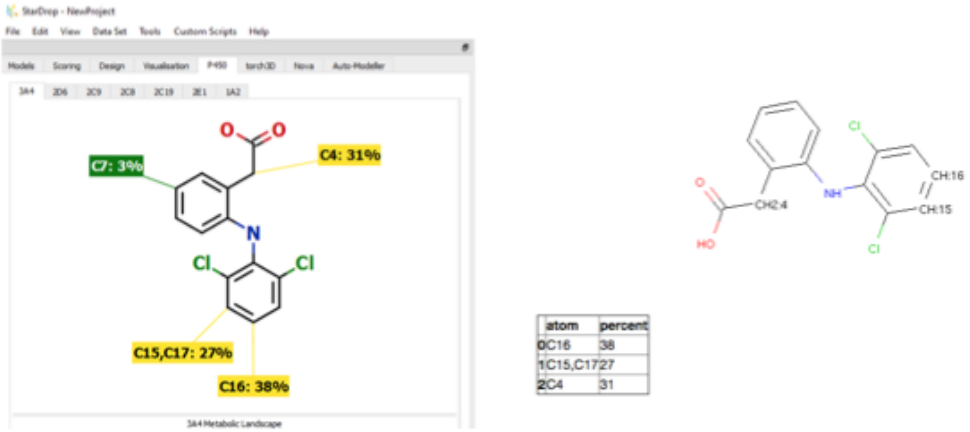
	CYP2D6_R	CYP3A4_R	activities	alert	lability	other_smiles	smiles	mol	other_mol
0	45.804700	49.389600		True	Labile	CC=O	CCCCc1oc2ccccc2c1C(=O)c1cc(I)c(OCCNCC)c(I)c1		
1	4.038930	0.158629	ACHE;CA2;CYP2C8;ESR1;ESR2	False	Mod Labile	None	CCCCc1oc2cc(O)ccc2c1C(=O)c1cc(I)c(OCCN(CC)CC)c(I)c1		None
2	1.049790	0.294910	CA2	False	Mod Labile	None	CCN(CC)CCOc1c(I)cc(C(=O)c2c(CCCCO)oc3ccccc23)cc1I		None
3	0.281715	0.218902	ADRA2B	False	Mod Labile	None	CCN(CC)CCOc1c(I)cc(C(=O)c2c(CCC(C)O)oc3ccccc23)cc1I		None
4	0.089519	0.342148		True	Mod Labile	CCNCC	CCCCc1oc2ccccc2c1C(=O)c1cc(I)c(OCC=O)c(I)c1		

```
result['parent_data']['activities'].split(';')
```

```
['ACHE', 'ADRA2B', 'CYP2C8']
```

The Optibrium StarDrop interface provides graphical tools for visualizing the results of calculations [45]. Replicating sophisticated visualizations such as these in this system is not practical at present. However, as a demonstration of what might be possible, a basic depiction of the key sites of metabolism and their proportions by isoform has been implemented. An example for Diclofenac is shown in Figure 6.

Figure 6



In summary, a system has been created that enumerates metabolites and generates predictions of the activities of these species against a panel of toxicity-associated biomolecular targets. It may be accessed directly from Python or *via* a simple web-service. It is intended that the current functionality be extended with modules implementing further predictive methods, such as PBPK models, as these become available to us.

DIFFICULTIES

As stated in the introduction, the objectives for this deliverable are ambitious, bringing together several fields that are the subject of active research in academia and the pharmaceutical industry. Some key areas are:

- Quantitative prediction of metabolites, especially for non CYP450-enzymes;
- Population variation in rates of metabolism *etc*;
- Incorporation of temporal and local concentration effects;
- Elucidation of all relevant AOPs and characterization of the biomolecular targets of the MIEs;
- Accurate prediction of xenobiotic interactions with and effects on these targets;
- Incorporation of non target-based toxicities (redox cycling, mitochondrial uncoupling *etc.*).

As noted, all these areas are under investigation in many academic and industrial laboratories. The system we have outlined here provides a framework into which modules handling these various facets could be incorporated as tools to address them become available.

REFERENCES

1. Vinken, M., *The adverse outcome pathway concept: a pragmatic tool in toxicology.* Toxicology, 2013. **312**: p. 158-65.
2. Allen, T.E., et al., *Defining molecular initiating events in the adverse outcome pathway framework for risk assessment.* Chem Res Toxicol, 2014. **27**(12): p. 2100-12.
3. Rautio, J., et al., *Prodrugs: design and clinical applications.* Nat Rev Drug Discov, 2008. **7**(3): p. 255-70.
4. FDA, *Cyclophosphamide: Full Prescribing Information from Drugs@FDA*. 2013.
5. FDA, *Carisoprodol: Full Prescribing Information from DailyMed.* 2013.
6. FDA, *Simvastatin: Full Prescribing Information from DailyMed.* 2010.
7. *Simvastatin*, in *Therapeutic Drugs*, C. Dollery, Editor. 1999, Churchill Livingstone. p. S37-S42.
8. Cheng, H., et al., *Influence of age and gender on the plasma profiles of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitory activity following multiple doses of lovastatin and simvastatin.* Pharm Res, 1992. **9**(12): p. 1629-33.
9. Cerny, M.A., *Prevalence of non-cytochrome P450-mediated metabolism in FDA approved oral and intravenous drugs: 2006 - 2015.* Drug Metab Dispos, 2016.
10. Tyzack, J.D.H., Peter A.; Segall, Matthew D., *Predicting Regioselectivity and Lability of Cytochrome P450 Metabolism using Quantum Mechanical Simulations.* Submitted for publication., 2016.
11. Tyzack, J.D.H., Peter A.; Segall, Matthew D., *Which P450.* Manuscript in preparation, 2016.
12. *Appendix: Drug Metabolizing Enzymes and Biotransformation Reactions*, in *ADME-Enabling Technologies in Drug Design and Development*. 2012, John Wiley & Sons, Inc. p. 545-565.
13. Strolin Benedetti, M., R. Whomsley, and E. Baltes, *Involvement of enzymes other than CYPs in the oxidative metabolism of xenobiotics.* Expert Opin Drug Metab Toxicol, 2006. **2**(6): p. 895-921.
14. Lhasa. *Meteor Nexus Product Page*. 2016; Available from: http://www.lhasalimited.org/products/meteor-nexus.htm.
15. Potter, T., et al., *In silico prediction of acyl glucuronide reactivity.* Journal of Computer-Aided Molecular Design, 2011. **25**(11): p. 997-1005.
16. Bode, C., *The nasty surprise of a complex drug-drug interaction.* Drug Discov Today, 2010. **15**(9-10): p. 391-5.
17. Fontana, E., P.M. Dansette, and S.M. Poli, *Cytochrome p450 enzymes mechanism based inhibitors: common sub-structures and reactivity.* Curr Drug Metab, 2005. **6**(5): p. 413-54.
18. Kalgutkar, A.S., et al., *A comprehensive listing of bioactivation pathways of organic functional groups.* Curr Drug Metab, 2005. **6**(3): p. 161-225.
19. Park, B.K., et al., *Managing the challenge of chemically reactive metabolites in drug development.* Nat Rev Drug Discov, 2011. **10**(4): p. 292-306.
20. Enoch, S.J., et al., *A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity.* Crit Rev Toxicol, 2011. **41**(9): p. 783-802.
21. Enoch, S.J. and M.T. Cronin, *A review of the electrophilic reaction chemistry involved in covalent DNA binding.* Crit Rev Toxicol, 2010. **40**(8): p. 728-48.
22. Vinken, M., et al., *Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury.* Toxicol Sci, 2013. **136**(1): p. 97-106.
23. *Adverse Outcome Pathway Wiki*. 2016; Available from: https://aopwiki.org/wiki/index.php/Main_Page.
24. Atkinson, F.L., *HeCaToS Deliverable Report D9.2:Report on reference set of toxicology associated targets*. 2015.
25. *Antitargets and Drug Safety*. Wiley's Methods and Principles in Medicinal Chemistry, ed. R. Mannhold, Kubinyi, H., Folkers, G. Vol. 66. 2015: Wiley-VCH.
26. Bento, A.P., et al., *The ChEMBL bioactivity database: an update.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1083-90.

27. Papadatos, G. *Ligand-based target predictions in ChEMBL*. 2014; Available from: http://chembl.blogspot.co.uk/2014/04/ligand-based-target-predictions-in.html.
28. Atkinson, F.L., *HeCaToS Deliverable Report D1.5:Package of Predictive Models.* 2015.
29. ChEMBL. *CHEMBL_21*. 2016; Available from: ftp://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/chembl_21/.
30. Kovacic, P., et al., *Mechanism of mitochondrial uncouplers, inhibitors, and toxins: focus on electron transfer, free radicals, and structure-activity relationships.* Curr Med Chem, 2005. **12**(22): p. 2601-23.
31. Volkova, M. and R. Russell, 3rd, *Anthracycline cardiotoxicity: prevalence, pathogenesis and treatment.* Curr Cardiol Rev, 2011. **7**(4): p. 214-20.
32. *HeCaToS Annex 1 - Description of Work*. 2013, Seventh Framework Programme.
33. Rana, P., et al., *Chemical motifs that redox cycle and their associated toxicity.* MedChemComm, 2013. **4**(8): p. 1175-1180.
34. Nelms, M.D., et al., *Development of an in Silico Profiler for Mitochondrial Toxicity.* Chem Res Toxicol, 2015. **28**(10): p. 1891-902.
35. Naven, R.T., et al., *The development of structure-activity relationships for mitochondrial dysfunction: uncoupling of oxidative phosphorylation.* Toxicol Sci, 2013. **131**(1): p. 271-8.
36. Espie, P., et al., *Physiologically based pharmacokinetics (PBPK).* Drug Metab Rev, 2009. **41**(3): p. 391-407.
37. Phillips, M.B., et al., *A Workflow to Investigate Exposure and Pharmacokinetic Influences on High-Throughput in Vitro Chemical Screening Based on Adverse Outcome Pathways.* Environ Health Perspect, 2016. **124**(1): p. 53-60.
38. Bosch, T.M., *Pharmacogenomics of Drug-Metabolizing Enzymes and Drug Transporters in Chemotherapy*, in *Pharmacogenomics in Drug Discovery and Development: From Bench to Bedside*, Q. Yan, Editor. 2008, Humana Press: Totowa, NJ. p. 63-76.
39. Certara, *Simcyp Simulator.* 2016.
40. Atkinson, F.L. *standardiser*. 2014; Available from: https://wwwdev.ebi.ac.uk/chembl/extra/francis/standardiser/.
41. HUGO. *HGNC: The HUGO Gene Nomeclature Committee*. 2015; Available from: http://www.genenames.org/.
42. Berthold, M.R., et al., *KNIME : The Konstanz Information Miner.* 2007.
43. *Jupyter Notebook*. 2016; Available from: http://jupyter.org/.
44. PyData. *PyData: Python Data Tools*. 2015; Available from: http://pydata.org/downloads/.
45. Optibrium. *StarDrop*. 2016; Available from: http://www.optibrium.com/stardrop/stardrop-features.php.