Title page

Evaluation of three state-of-the-art metabolite prediction software packages (Meteor, MetaSite, and StarDrop) through independent and synergistic use

Tjollyn H., Boussery K., Mortishire-Smith R.J., Coe K., De Boeck B., Van Bocxlaer J.F., Mannens G.

Laboratory of Medical Biochemistry and Clinical Analysis, Fac. Pharm. Sci., Ghent University, Harelbekestraat 72, B-9000 Ghent, Belgium (TH; BK; VBJF)

ADME-Tox Department, Janssen Research and Development, Janssen Pharmaceutica, Turnhoutsebaan 30, B-2340 Beerse, Belgium (MSRJ)

ADME-Tox Department, Janssen Research and Development, 3210 Merryfield Row, San Diego, CA, USA (CK)

Department of Medicinal Chemistry, Janssen Research and Development, Turnhoutsebaan 30, B-2340 Beerse, Belgium (DBB)

Drug Metabolism and Pharmacokinetics Department, Janssen Research and Development, Turnhoutsebaan 30, B-2340 Beerse, Belgium (MG)
Running title page

Use of metabolite prediction software: an evaluation

Geert Mannens
Senior Research Fellow, Drug Metabolism & Pharmacokinetics
Janssen Research & Development
Turnhoutsebaan 30
B-2340 Beerse
Belgium
tel. +32/(0)14 603745
cfax. +32/(0)14 605110
e-mail: gmannens@its.jnj.com

Number of:
text pages: 17
tables: 3
figures: 7
words in Abstract: 243
words in Introduction: 761
words in Discussion: 1286
references: 23

Nonstandard abbreviations

AME: absorption, metabolism, excretion; UD: unchanged drug; CYP450: Cytochrome P450; SOM: site of metabolism; ARL: absolute reasoning level; RRL: relative reasoning level
Abstract

The aim of this study was to evaluate three different metabolite prediction software packages (Meteor, MetaSite, and StarDrop) with respect to their ability to predict loci of metabolism and suggest relative proportions of metabolites. A chemically diverse test set of 22 compounds for which in vivo human mass balance studies and metabolic schemes were available, was used as basis for the evaluation. Each software package was provided with structures of the parent compounds, and predicted metabolites were compared with experimentally determined human metabolites. The evaluation consisted of two parts: firstly, different settings within each software package were investigated and the software was evaluated using those settings determined to give the best prediction. Secondly, the three different packages were combined using the optimized settings to see whether a synergistic effect concerning the overall metabolism prediction could be established. The performance of the software was scored for both sensitivity and precision, taking into account the capabilities/limitations of the particular software. Varying results were obtained for the individual packages. Meteor showed a general tendency towards overprediction and this led to a relatively low precision (~35%) but high sensitivity (~70%). MetaSite and StarDrop both exhibited a sensitivity and precision of ~50%. By combining predictions obtained with the different packages, we found that increased precision can be obtained. We conclude that the state-of-the-art individual metabolite prediction software has many advantageous features but needs refinement to obtain acceptable prediction profiles. Synergistic use of different software packages could prove useful.
Introduction

In drug discovery, the metabolism of a drug compound is one of the key parameters to be investigated and optimized in order to obtain acceptable pharmacokinetic and/or safety profiles. In development, early prediction of relevant metabolites prior to introduction of the drug compound into man would substantially help preclinical and clinical development concerning the analysis of (phase I and/or II) metabolism data, and decision making (Afzelius et al., 2007; Van campen, 2009).

A variety of different approaches to predict drug metabolism have been described: rule-, ligand-protein-, and ligand-based methods (Cruciani et al., 2005; Tarcsay and Keseru, 2011). Other approaches include ab initio, classical pharmacophore and 3D-QSAR pharmacophore-based methods (Cruciani et al, 2005; Czodrowski et al., 2009; Testa, 2004). Unfortunately, obstacles such as biological factors and substrate-product selectivity still hinder the widespread use of drug metabolism as a reliable tool (Testa, 2004).

Here, we describe the evaluation of three software tools: Meteor (Lhasa Ltd., Leeds, UK), MetaSite (Molecular Discovery Ltd., Middlesex, UK), and StarDrop (Optibrium Ltd., Cambridge, UK). We have chosen these tools because they are currently being used in both the discovery and development phases of the pharmaceutical industry and each tool is based on a different approach to predict metabolism. A comparison of the software packages is contained in Table 1.

Meteor is a rule-based (empirical) software tool (Langowski and Long, 2002). Its algorithm involves three successive steps. Firstly, Meteor checks whether the query structure contains substructures which are labile towards any of the biotransformations contained in its knowledge base. Secondly, absolute reasoning rules evaluate the likelihood of a biotransformation taking place based on five levels: Probable, Plausible, Equivocal, Doubted and Improbable (Button et al., 2003). This classification depends on the logP of the query structure, whether the query structure is the unchanged drug (UD) or a metabolite, and in
which species the metabolism is to be predicted. Thirdly, relative reasoning then is used to rank those biotransformations that can occur concomitantly on the same compound, based on a set of relative precedences (e.g. primary alcohols are oxidized in preference to secondary alcohols). Relative reasoning can be set at 0, 1, 2, 3 and can only be used when 2 or more biotransformations apply for the same compound: RRL0 means that no relative reasoning will be applied in the upcoming analysis; RRL1 means that only metabolites will be displayed for which there is no metabolite more likely (i.e. most likely); RRL2 means that most likely metabolites, and metabolites for which there is only 1 level of metabolites more likely, will be displayed. In essence, the reasoning engine uses further rules to avoid a combinatorial explosion of output resulting from unconstrained analyses of query structures (Ekins et al., 2007).

MetaSite is an automated docking model with reactivity correction (this correction considers the reactivity components of an atom related to the heme) and is designed to predict phase I CYP450 metabolism. Based on GRID descriptors for the CYP450 enzymes and the potential substrate (Cruciani et al, 2005) metabolism is evaluated at all possible sites on the molecular structure, assigning every atom a likelihood of metabolism. Reactivity correction can be put in 3 different modes: (i) ‘off’; (ii) ‘on for substrate’; (iii) ‘on for substrate & CYP’ (Cruciani et al, 2005; Vaz et al., 2010; Zhou et al., 2006)

StarDrop uses a quantum mechanical approach for the prediction of the relative involvement of CYP3A4, 2D6 and 2C9 of the query compound. Its mechanism is based on calculation of the energy barrier to the electron removal, which is considered to be the rate limiting step in product formation. All of the modeled CYP450 isoforms use the same model for the calculation of electronic lability but each isoform has a different model for steric accessibility and orientation effects (different regioselectivity) (Earnshaw, 2010).

Metabolism prediction is scored using two performance parameters: sensitivity and precision, with the in vivo metabolic profile as a reference point. Sensitivity is a measure of
how many *in vivo* metabolites are captured by the software; precision is a measure of how many predicted metabolites are actually observed *in vivo* (see paragraph 3.5 for a definition). Ideally sensitivity as well as precision should be as high as possible for any given prediction. However a prioritization of either sensitivity or precision will be necessary, depending on the scope of the intended research. Figure 1 correlates the location of a prediction result within each quadrant in the sensitivity/prediction plot with the value of the prediction result. In the discussion we will further comment on the usefulness of the prediction result in drug discovery and development.
Materials and Methods

The in vivo reference set

The goal of most in silico models is to predict what cannot directly be measured. In this project, the output of the metabolism prediction software packages was compared to the known metabolite profile obtained following administration in man. To understand the most important metabolic clearance pathways of a compound, the mass balance of the parent and its metabolites was determined in excreta after a single oral dose. A test set of 22 compounds (Figure 2; Table 2) was chosen with divergent chemical structures, therapeutic doses, therapeutic indications, extent of metabolism and number of metabolites, and representing a range of metabolic lability from barely to extensively metabolized compounds.

For every compound in the set of test compounds, the report of the human AME trial was abstracted to provide a table of observed metabolites. Finally, the primary, main metabolic pathways were collated, highlighting metabolites that were expected to be formed directly from the parent compound (illustrated in the Supplemental data).

The reference set contained 68 (phase I CYP450, phase I non-CYP450 and phase II) primary biotransformation pathways. A list of these biotransformations can be found per compound in Table 2. MetaSite and StarDrop are only able to predict phase I CYP450 metabolism. Consequently, only 50 (instead of 68 for Meteor) of the observed in vivo CYP450 biotransformations were taken as basis for the evaluation.

The software packages

The three different metabolite prediction software tools were (i) Meteor (editor mode), Lhasa Ltd, v11; (ii) MetaSite, Molecular Discovery Ltd., v3.0.1 and (iii) StarDrop, Optibrium Ltd., v3.4. Chemical structures were drawn with CS ChemDraw Pro, CambridgeSoft, v4.5 and ISIS Draw, MDL information Systems, v2.4.

Testing of individual software packages
Both Meteor and MetaSite allow the user to customize the prediction settings. Only StarDrop had no selectable settings. The outcome of each of the various settings investigated, was evaluated and these different outcomes were compared to the \textit{in vivo} data. The goal of this software per software exercise was to (i) find which settings provided the best prediction for a given software package (INTRA); (ii) compare the three software packages to each other using their optimized settings (INTER) and (iii) investigate whether there was any added prediction value in combining the outcomes of individual software packages (illustrated in the Supplemental data).

\textbf{Meteor}

Each chemical structure was imported and predictions were performed, using each possible setting (absolute reasoning level (ARL) and relative reasoning level (RRL)). Absolute reasoning was tested only at the ‘probable’ and ‘plausible’ levels. Relative reasoning was tested on levels 0, 1, 2 for each absolute reasoning level. The outcomes for these six different setting combinations were compared to 68 \textit{in vivo} biotransformations from the reference data. From now on, when quoting a setting in Meteor we will use the format ‘ARL RRL’ (e.g. ‘plausible 1’).

Meteor predicts biotransformations rather than metabolic hotspots. This is because per locus of metabolism, several metabolites can be formed. For example, following $N$-dealkylation, three metabolites are formed \textit{in silico}: an amine, an alcohol, and further downstream oxidation of the alcohol to the acid. This complicates the comparison of the three packages, because where MetaSite or StarDrop would predict one site of metabolism correlating with $N$-dealkylation, Meteor would predict three metabolites.

\textbf{MetaSite}

Each chemical structure was imported into the software and a prediction of hotspots was performed. MetaSite allows the user to toggle between various reactivity settings \textit{post hoc}. We’ve tested the settings ‘without reactivity correction’, ‘with reactivity correction for the
substrate’, and ‘with reactivity correction for substrate and CYP’. Only ‘most probable’ and ‘intermediate probability’ sites (correlating with relative percentage intervals [0.50; 1.00] and [0.25; 0.50], respectively) were considered. The metabolic hotspots obtained with these six different settings were compared to 50 *in vivo* CYP450 biotransformations from the reference data. From now on, when quoting a setting in MetaSite we will use the format ‘considered SOM; reactivity correction setting’ e.g. (‘most probable & intermediately probable; on for substrate’).

**StarDrop**

Each chemical structure was imported into the software and a prediction of metabolism performed. StarDrop had no user-adaptable settings. Only ‘labile’ and ‘moderately labile’ positions were considered. The outcome was compared to 50 *in vivo* CYP450 biotransformations from the reference data.

**Testing the combination of software packages**

The individual outcomes of the different software packages were also combined to examine whether this could increase the prediction reliability. Several combinational approaches were tested and reported using the following approaches (Table 3).

**Intersection**

All three software packages predict CYP450 reactions, only Meteor also predicts phase I non-CYP450 and phase II reactions. Consequently, when evaluating which metabolites were predicted *in common* by the various packages, only CYP450 reactions were taken into account. This ‘intersection’ approach was used with the following combinations: Meteor – MetaSite; Meteor – StarDrop; MetaSite – StarDrop.

**Union**

In the ‘union’ approach, *all* the predictions that are made by each package are taken into account, including biotransformations that are predicted *in common*, as well as
biotransformations that are independently predicted by the software packages. We also investigated whether the different software packages were complementary to each other. By ‘complementary’, we mean the prediction of independently predicted biotransformations of a certain software package that were not seen in common to the 2 software packages. The following combinations were made: Meteor – MetaSite; Meteor – StarDrop; MetaSite – StarDrop.

Intersection+n

‘Intersection+2’ was the approach used to combine results from Meteor and MetaSite: CYP450 biotransformations which were common to these two packages, together with additional phase I non-CYP450 and phase II predictions from Meteor. ‘Intersection+3’ was the approach used to combine results from Meteor, MetaSite as well as StarDrop: CYP450 biotransformations common to (i) the three packages, and (ii) common to Meteor and MetaSite, Meteor and StarDrop, and MetaSite and StarDrop, together with additional phase I non-CYP450 and phase II predictions from Meteor.

Evaluation

Each prediction of a metabolite or biotransformation route was compared to the experimentally observed in vivo biotransformations by scoring a predicted metabolite or biotransformation route as ‘TRUE’, ‘FALSE’ or ‘MISSED’.

In the case of binary classification of data (as in this case) performance testing parameters such as sensitivity, specificity, accuracy and precision can be used. In the case of metabolism predictions compared to a reference set, accuracy and specificity cannot be used because the concept ‘true negative’ (used to calculate accuracy and specificity) is not appropriate. A ‘true negative’ would mean a metabolite NOT predicted in silico and NOT seen in vivo. This is an irrelevant situation in metabolism prediction, since one could then take any possible biotransformation, whether or not relevant to the query structure, into the comparison. For example, the human AME report indicates that RWJ-241947 is not
acetylated \textit{in vivo} (Supplemental data). RWJ-241947 acetylation is also not predicted \textit{in silico}. Thus, acetylation would be a ‘true negative’. Nevertheless, ‘true negatives’ don’t make any sense in the evaluation of the performance of a metabolism prediction package.

Our prediction quality assessment was then as follows

- \textit{True positive}: a metabolite predicted and also seen \textit{in vivo} was assigned the value \textbf{TRUE}

- \textit{False positive}: a metabolite predicted but NOT also seen \textit{in vivo} was assigned the value \textbf{FALSE}

- \textit{False negative}: a metabolite NOT predicted, but seen \textit{in vivo} was assigned the value \textbf{MISSED}

Precision describes the proportion of correctly predicted metabolites in the population of all \textit{in silico} metabolites of a test compound, and is defined as follows

\[
\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \times 100\%
\]

Sensitivity describes the ability of a system to predict a TRUE metabolite in the population of all \textit{in vivo} metabolites of a certain test compound, and is defined as follows

\[
\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \times 100\%
\]
Results

Results for the individual software packages

Meteor

Meteor was tested at 6 different settings: the ARL at ‘probable’ with RRL 0, 1 and 2 and the ARL at ‘plausible’ with RRL 0, 1 and 2. For each of these settings, the absolute amounts of TRUE, FALSE and MISSED biotransformations are shown in Figure 3.

The influence of the parameter settings is best visualized through a precision versus sensitivity plot (Figure 4). Using the ‘probable’ level, the sensitivity is ~20% (20% of the experimentally observed metabolites are correctly predicted) with a precision of ~40% (40% of the predicted metabolites are actually observed in vivo). In the ‘plausible’ level the precision lowers a few percent points but there is a very substantial increase in sensitivity. Indeed, when moving from ‘probable’ to ‘plausible’, many of the MISSED metabolites (not predicted, although experimentally observed) from the ‘probable’ level do become TRUE metabolites (predicted and observed) in the ‘plausible’ level.

Relative reasoning levels constrain the prediction output. The lower the relative reasoning level, the more restriction there is. The result of this constraint is exemplified in Figure 4, where a decrease in RRL (RRL 1 as compared to RRL 2) is associated with a strong decrease in sensitivity (especially at ‘plausible’ absolute reasoning) and only a small increase in precision. That is to say, constraining the prediction output (lower RRL) results in slightly more precise, yet less sensitive results. With the relative reasoning set to off (RRL=0) we can isolate the effect of the absolute reasoning filter. The arrows clearly indicate that when moving from ‘probable’ to ‘plausible’ ARL we gain a lot in sensitivity while precision remains about the same (grey for RRL0, black for RRL1, and white for RRL 0).

MetaSite
Reactivity correction in MetaSite can be set to (i) ‘off’, (ii) ‘on for substrate’ and (iii) ‘on for substrate and CYP’. We evaluated these three settings for both ‘most probable’ and ‘most probable & intermediately probable’ sites of metabolism (SOM), resulting in six different outcomes.

With the reactivity correction removed, precision is at its lowest (Figure 5) at ~20%. When considering ‘most probable & intermediately probable’ SOMs, the differences in sensitivity and precision between settings ‘on for substrate’ and ‘on for substrate and CYP’ is minimal. In Figure 5 the grey and especially the black arrow indicates that, when moving from ‘most probable’ to ‘most probable & intermediately probable’ the gain in sensitivity is accompanied by a loss in precision. Of course, also reactivity correction largely affects the sensitivity and precision (compare results 1 → 3 → 5).

The best predictions in MetaSite (combination of the highest sensitivity with the highest precision) were obtained when considering only the most probable metabolites, and using reactivity correction enabled for substrate.

StarDrop

We considered those hotspots considered to be ‘labile’ and ‘moderately labile’ positions, respectively, for each of the different SOMs for each query compound. Since StarDrop has no user-selectable settings, the prediction quality plot (Figure 5) has only one outcome with precision of ~45% and sensitivity of ~51%, which is comparable with the best outcome obtained with MetaSite.

Results for the combinations of software packages

The quality plots for software combinations, Figure 6 (CYP450 predictions only) and Figure 7 (phase 1 & phase 2 predictions) show (i) the optimum outcomes per individual software package (as described in the above sections), and (ii) the outcomes of a combination of software packages (as described: ‘intersection’, ‘union’, and ‘intersection+n’).
The results are again based on the combined outcomes for all 22 compounds in the reference set, resulting in a single scatter plot entry per setting or software combination.

In Figure 6 the black arrow indicates that when moving from the best setting in MetaSite to the combination of MetaSite in its best setting with Meteor ‘plausible 0’, a great increase in prediction precision is found. The white arrow suggests the same but for the combination of StarDrop with Meteor ‘plausible 0’. Using the ‘intersection’ approach for CYP450 predictions (Figure 6), precision can be increased by 20% with no reduction in sensitivity. The ‘union’ result shows an increase in sensitivity but again a great loss in precision.

In Figure 7 the white arrow represents the evolution from the result for ‘plausible 0’ in Meteor towards the result for a ‘union’ combination with MetaSite that is not useful; this ‘union’ approach does not improve prediction quality compared with the individual outcomes for Meteor. The black arrow in Figure 7 indicates the evolution towards the ‘intersection+n’ approach, which resulted in an increase in precision of ~20%, with little or no decrease in sensitivity, compared to Meteor results ‘plausible 0’ and ‘plausible 2’, respectively. Both ‘intersection+2’ (based on the common CYP450 predictions of Meteor and MetaSite and the phase I non-CYP450 and phase II predictions from Meteor) and ‘intersection+3’ (based on the common CYP450 predictions of Meteor, MetaSite and StarDrop and the phase I non-CYP450 and phase II predictions from Meteor) provide nearly similar sensitivity (~65%) and precision (~55%) quality values.
Discussion

**Optimal individual software settings**

Meteor in setting ‘plausible 1’ offered more precise but less sensitive predictions, compared to the other 5 Meteor results (Figure 4). In contrast, settings ‘plausible 2’ (default setting) and ‘plausible 0’ (Meteor’s most valuable setting) can be used to obtain more sensitive results. In our opinion the enhanced sensitivity is Meteor’s most valuable feature. However, it remains to be seen whether this enhanced sensitivity is really an advantage, especially in a discovery setting. Rule-based software such as Meteor is known to show a tendency for overprediction (high sensitivity/low precision) of metabolism (Cruciani et al., 2005) e.g. a piperidine moiety in a compound can undergo *N*-oxidation, but in another compound, the same piperidine moiety is oxidized on a carbon atom (*lactam formation*) of the heterocycle. Rule-based software that recognizes the piperidine moiety will generate these 2 metabolites (*N*-oxide and lactam). A high sensitivity/low precision is useless in lead optimization, because it is not able to guide the medicinal chemists. The high sensitivity denotes that a lot of information is there, but cannot be used with sufficient certainty due to the low precision. However, this pattern shouldn’t pose any problem if the scope of the intended research is to actively look for metabolites. In the context of metabolite hunting or metabolite explanation, Meteor already proved useful in software-assisted detection of metabolites (Naegele et al., 2010; Pelander et al., 2009; Valerio and Long, 2010).

The most valuable outcome for MetaSite and StarDrop is highlighted in Figure 5. The overall prediction pattern of these software tools yields more precise, albeit less sensitive predictions, compared to Meteor. Therefore, in the elucidation of a compound’s metabolites in early lead optimization and soft spot identification, StarDrop (Earnshaw, 2010; Segall et al., 2011; Segall et al., 2009) and MetaSite (Ahlstrom et al., 2007; Boyer et al., 2009; Myatt et al., 2010; Trunzer et al., 2008; Wu et al., 2010) are much more appropriate tools.

**Combining metabolite prediction software**
The ‘intersection’ combination of Meteor in the configuration ‘plausible 0’ with MetaSite in the configuration ‘most probable; on for substrate’ enhances prediction reliability, as it provides a precision level practically unsurpassed by any other setting or combination (Figure 6). Again, this is at the expense of sensitivity. By applying the ‘intersection’ approach, an increase in prediction precision is achieved by a reduction of false positive predictions made by the software packages, and underscores the added value of combining software packages which are based on quite different algorithms.

The ‘union’ combination of MetaSite and StarDrop (Figure 6) was not useful as such but could provide us with some valuable insights: a sensitivity of 65% is achieved by almost equal contributions of the 2 software tools, while in their ‘intersection’ combination (Figure 6) sensitivity represented only 37% (meaning few biotransformations in common). This would imply that MetaSite and StarDrop are –to a certain degree- complementary to each other in predicting in vivo metabolism. The ‘union’ combinations from Figure 7 did not yield more precise nor more sensitive results, compared to Meteor’s best individual settings (‘plausible 0’; ‘plausible 2’).

By making the ‘intersection+n’ combinations (sum of the common CYP450 predictions by 2 or 3 programs and the phase I non-CYP450 and phase II predictions from Meteor) we wanted to investigate how the prediction would be affected if Meteor supplied information only on phase I non-CYP450 and phase II predictions, while the precision in CYP450 predictions is maintained by using the ‘intersection’ approach. This procedure fits in neatly with the concept of taking the positive aspects of each individual software package and trying to circumvent, as much as possible, their limitations. Figure 7 demonstrates that precision is increased by nearly 20% for Meteor, 15% for StarDrop, and 5% for MetaSite, compared to the individual packages when used in isolation.

In a recently published review article (Tarcsay and Keseru, 2011), the performance of metabolite prediction softwares was also assessed. When we compare our investigation
results to the studies reviewed in this publication, we found that our prediction quality assessment points towards a substantially lower prediction quality compared to what these authors report. This could be due to the fact that a great difference in scoring methodology exists between both our studies. The studies mentioned by Tarcsay and Keseru report on scoring the accuracy of predictions (SOMs) of the different prediction tools based on the first, the first 2 or the first 3 ranked predicted metabolites per software tool and per CYP450 isoform. In our article we wanted to present a more generic approach, scoring the performance of the software outcome by letting it use its own constraints instead of the user choosing some cut off value. We also wanted to evaluate the value of combining both oxidative and conjugative biotransformation predictions. We acknowledge the primary importance of CYP450-mediated metabolism but reliable predictions of phase II metabolism are also of interest (Cubitt et al., 2011).

**Limitations and Pitfalls**

An important remark is that although we consider the *in vivo* metabolite profile in human as the golden standard, this approach has some limitations. The metabolite profile may change as the dose level (saturation of processes) or extent of absorption (formulation dependent) changes, or when transporters affect the exposure of the compound to the metabolizing enzymes. However, similar concerns also apply when comparing the software outputs against *in vitro* metabolite profiles in liver microsomes or hepatocytes.

In Meteor we found that there is no correlation between the absolute reasoning level of predicted metabolites and their abundance *in vivo*, irrespective of the terminology probable, plausible, etc. Because Meteor does not consider the shape of the molecule in making predictions, Meteor over-emphasizes simple oxidations. Indeed, allowing Meteor to evaluate metabolic transformations down to the ‘equivocal’ level, results in completely indiscriminate hydroxylation. Finally, *N*-glucuronidations at amine moieties are often found *in vivo* for the test set but they are only predicted at the ‘equivocal’ absolute reasoning level which is much too low for that particular biotransformation in humans. *N*-acetylations are heavily
overpredicted while N-oxidations at morpholine and piperidine rings are underpredicted, again for this test set.

The reactivity correction provides the best results when set at ‘on for substrate’, thus without considering the reaction mechanism of the CYP450 enzyme. One would expect more precision of the prediction profile in MetaSite considering the CYP450 mechanism additionally, but surprisingly this is not the case. Sometimes MetaSite identifies an in silico SOM right next to what the in vivo results indicate as a SOM. The fine-tuning of relative distances inside the active site of virtual CYP450 enzymes could possibly help to resolve this problem. In Metasite too, some important aliphatic/aromatic hydroxylations, that were seen for the test set, are not predicted.

For StarDrop one cannot attach much value to the absolute factor in the metabolism prediction because it does not correlate with the in vivo abundance of metabolism, irrespective of the terminology labile positions, stable positions, etc. N-oxidations and some aliphatic hydroxylations, in our case, were not predicted by StarDrop.

**Conclusion**

In this study, we have evaluated and compared the metabolite prediction software packages Meteor, MetaSite, and StarDrop, in terms of precision and sensitivity. We conclude that the state-of-the-art individual metabolite prediction software has many advantageous features but still needs refinement to obtain an acceptably useful prediction profile. We found that intelligent combinations, i.e. combining packages based on different mechanistic principles could prove useful and should be pursued in order to increase the prediction precision. Depending on the scope of the intended research either sensitivity or precision should be prioritized. The answer to “what will happen” in drug metabolism should be sought in precise prediction tools; the answer to “what might happen” in drug metabolism should be sought in sensitive prediction tools.
Authorship contributions

*Participated in research design*: T’jollyn, and Mannens

*Conducted experiments*: T’jollyn, and Coe

*Contributed new reagents or analytical tools*: /

*Performed data analysis*: T’jollyn, and Mannens

*Wrote or contributed to the writing of the manuscript*: T’jollyn, Boussery, Mortishire-Smith, Coe, De Boeck, Van Bocxlaer, and Mannens
References


Legends for figures

Figure 1: sensitivity and precision classification and likely use of the output

Figure 2: chemical structures of the 22 compounds of the test set.

Figure 3: absolute number of biotransformations in function of six different settings in Meteor. TRUE and FALSE predictions are made by the software; MISSED predictions are not seen *in silico*

Figure 4: sensitivity/precision plot representing the outcomes for six different settings in Meteor. Most promising settings are encircled and will be used later on for making combinations of software packages (note: dots 1 and 3 coincide)

Figure 5: sensitivity/precision plot representing the outcomes for six settings in MetaSite and one setting in StarDrop, all concerning CYP450 predictions. Most promising settings are encircled per software package and will be used later on for making combinations of software packages

Figure 6: sensitivity/precision plot representing the outcomes for the best individual results for MetaSite and StarDrop and for the ‘intersection’ and ‘union’ combination of software packages, concerning CYP450 predictions. Three combinations out of five that show an increased prediction precision and still an appreciable sensitivity, are encircled

Figure 7: sensitivity/precision plot representing the outcomes for the best individual results for Meteor and for the ‘union’ and ‘intersection+n’ combination of software packages, concerning phase I & II predictions. Two combinations out of six show an increased precision outcome and are encircled
Table 1: An overview of the 3 metabolite prediction software packages that will be discussed in this article

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Displayed</th>
<th>Phase I</th>
<th>Phase II</th>
<th>CYP450 isoforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meteor v11</td>
<td>Rule-based</td>
<td>Metabolites</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>MetaSite v3.0.1</td>
<td>Docking+ reactivity correction</td>
<td>Metabolic hotspots</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>StarDrop v3.4</td>
<td>Quantum mechanics</td>
<td>Metabolic hotspots</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>
Table 2 represents the test set of 22 compounds that was submitted to the different software packages and for which *in vivo* mass balance studies were available as reference.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Indication/target</th>
<th>single oral dose (mg)</th>
<th>Metabolic clearance normalized (%) (TR-UD)/TR</th>
<th>Type of biotransformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 RWJ-241947</td>
<td>anti-hyperglycemic agent</td>
<td>80</td>
<td>85.1</td>
<td>O-dealkylation Thiazolidinedione open</td>
</tr>
<tr>
<td>2 levocabastine</td>
<td>H1-antagonist</td>
<td>1</td>
<td>21.8</td>
<td>Acyl glucuronidation</td>
</tr>
<tr>
<td>3 risperidone</td>
<td>anti-psychotic agent</td>
<td>1</td>
<td>81.7</td>
<td>N-dealkylation Aliphatic oxidation Benzisoxazole open</td>
</tr>
<tr>
<td>4 vorozole</td>
<td>aromatase inhibitor</td>
<td>2,5</td>
<td>99.8</td>
<td>N-demethylation N-dealkylation N-oxidation N-glucuronidation</td>
</tr>
<tr>
<td>5 cisapride</td>
<td>gastroprokinetic agent</td>
<td>10</td>
<td>89.6</td>
<td>N-dealkylation O-dealkylation Aromatic oxidation</td>
</tr>
<tr>
<td>6 liarozole</td>
<td>anti-tumor agent</td>
<td>150</td>
<td>85.6</td>
<td>Benzimidazole oxidation N-glucuronidation N-dealkylation</td>
</tr>
<tr>
<td>7 ridogrel</td>
<td>anti-coagulant</td>
<td>5</td>
<td>98.9</td>
<td>N-oxidation β-oxidation Acyl glucuronidation</td>
</tr>
<tr>
<td>8 prucalopride</td>
<td>chronic constipation</td>
<td>0,5</td>
<td>12.1</td>
<td>N-dealkylation Oxidative O-demethylation</td>
</tr>
<tr>
<td></td>
<td>Drug Name</td>
<td>Class</td>
<td>CYP450 Enzyme</td>
<td>Mechanism</td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>------------------------------</td>
<td>---------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>9</td>
<td>sabeluzole</td>
<td>Alzheimer</td>
<td></td>
<td>N-oxidation</td>
</tr>
<tr>
<td>10</td>
<td>tipifarnib</td>
<td>anti-tumor agent</td>
<td></td>
<td>N-dealkylation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Aromatic oxidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Glucuronidation</td>
</tr>
<tr>
<td>11</td>
<td>norcisapride</td>
<td>gastroprokinetic agent</td>
<td></td>
<td>Oxidative deamination</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Loss of imidazole</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-demethylation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-glucuronidation</td>
</tr>
<tr>
<td>12</td>
<td>carisbamate</td>
<td>anti-epileptic agent</td>
<td></td>
<td>Carbamate hydrolysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Glucuronidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Epimerization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Aromatic oxidation</td>
</tr>
<tr>
<td>13</td>
<td>JNJ-37822681</td>
<td>D2-antagonist</td>
<td></td>
<td>N-dealkylation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-oxidation</td>
</tr>
<tr>
<td>14</td>
<td>canagliflozine</td>
<td>SGLT2 inhibitor</td>
<td></td>
<td>Glucuronidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Benzylic oxidation</td>
</tr>
<tr>
<td>15</td>
<td>doripenem</td>
<td>broad spectrum carbapenem</td>
<td>500 (iv)</td>
<td>β-lactam ring open</td>
</tr>
<tr>
<td>16</td>
<td>darunavir</td>
<td>anti-HIV-1 agent</td>
<td></td>
<td>Carbamate hydrolysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Aliphatic oxidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Aromatic oxidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Glucuronidation</td>
</tr>
<tr>
<td>17</td>
<td>etravirine</td>
<td>anti-HIV-1 agent</td>
<td>800</td>
<td>Methyl oxidation</td>
</tr>
<tr>
<td>18</td>
<td>rilpivirine</td>
<td>anti-HIV-1 agent</td>
<td>150</td>
<td>Methyl oxidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>α,β-unsaturated bond</td>
</tr>
</tbody>
</table>

This article has not been copyedited and formatted. The final version may differ from this version.
<table>
<thead>
<tr>
<th>No.</th>
<th>Compound</th>
<th>Category</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Metabolism/Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>galantamine</td>
<td>Acetylcholinesterase inhibitor</td>
<td>4</td>
<td>66.7</td>
<td>O-demethylation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-demethylation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-oxidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Epimerization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Glucuronidation</td>
</tr>
<tr>
<td>20</td>
<td>R228060</td>
<td>anti-depressant</td>
<td>200</td>
<td>0.4</td>
<td>N-acetylation</td>
</tr>
<tr>
<td>21</td>
<td>paliperidone</td>
<td>anti-psychotic agent</td>
<td>1</td>
<td>34.8</td>
<td>N-dealkylation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alcohol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>dehydrogenation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Aliphatic oxidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Benzisoxazole open</td>
</tr>
<tr>
<td>22</td>
<td>JNJ-31001074</td>
<td>ADHD</td>
<td>10</td>
<td>87.8</td>
<td>N-dealkylation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ring oxidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N-oxidation</td>
</tr>
</tbody>
</table>

*iv: intravenous; TR= total radioactivity in excreta, UD= unchanged drug; † Carbamate hydrolysis is considered as ester hydrolysis by esterases; ^ GSH conjugation is considered phase 2
Table 3 represents an overview of combinations of software packages that were evaluated in this project. Column one describes the type of combination, columns two to four indicate which software setting was used to obtain the combination outcome.

<table>
<thead>
<tr>
<th>Combination approach</th>
<th>Meteor</th>
<th>MetaSite</th>
<th>StarDrop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intersection</td>
<td>‘plausible 2’</td>
<td>‘most probable; on for substrate’</td>
<td>no</td>
</tr>
<tr>
<td>Intersection</td>
<td>‘plausible 0’</td>
<td>‘most probable; on for substrate’</td>
<td>no</td>
</tr>
<tr>
<td>Intersection</td>
<td>‘plausible 0’</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Intersection</td>
<td>no</td>
<td>‘most probable; on for substrate’</td>
<td>yes</td>
</tr>
<tr>
<td>Union</td>
<td>‘plausible 2’</td>
<td>‘most probable; on for substrate’</td>
<td>no</td>
</tr>
<tr>
<td>Union</td>
<td>‘plausible 0’</td>
<td>‘most probable; on for substrate’</td>
<td>no</td>
</tr>
<tr>
<td>Union</td>
<td>‘plausible 1’</td>
<td>‘most probable; on for substrate’</td>
<td>no</td>
</tr>
<tr>
<td>Union</td>
<td>‘plausible 1’</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Union</td>
<td>no</td>
<td>‘most probable; on for substrate’</td>
<td>yes</td>
</tr>
</tbody>
</table>
Sensitivity/precision: usefulness

Figure 1

- **High sensitivity/low precision**: most in vivo metabolites are predicted, but also a lot of false predictions are made
  - Results are sensitive, but not precise

- **Low sensitivity/low precision**: few in vivo metabolites are predicted and also a lot of false predictions are made
  - Results are useless

- **High sensitivity/high precision**: most in vivo metabolites are predicted and not many false predictions are made
  - Results are sensitive & precise

- **Low sensitivity/high precision**: few in vivo metabolites are predicted, but not many false predictions are made
  - Results are precise, but not sensitive
Figure 2a
Figure 2b

Tipifarnib

Norcisapride

Carisbamate

JNJ-37822681

Canagliflozin

Doripenem

Duranavir
Figure 2c
### Meteor v11

<table>
<thead>
<tr>
<th></th>
<th>PROBABLE RRL0</th>
<th>PROBABLE RRL1</th>
<th>PROBABLE RRL2</th>
<th>PLAUSIBLE RRL0</th>
<th>PLAUSIBLE RRL1</th>
<th>PLAUSIBLE RRL2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MISSED</td>
<td>54</td>
<td>56</td>
<td>54</td>
<td>18</td>
<td>34</td>
<td>23</td>
</tr>
<tr>
<td>FALSE</td>
<td>24</td>
<td>18</td>
<td>24</td>
<td>101</td>
<td>48</td>
<td>86</td>
</tr>
<tr>
<td>TRUE</td>
<td>14</td>
<td>12</td>
<td>14</td>
<td>50</td>
<td>34</td>
<td>45</td>
</tr>
</tbody>
</table>

**Figure 3**
The outcomes for the individual settings for Meteor

Figure 4
Overview of the outcomes for the individual settings concerning CYP predictions

1. MetaSite ('most probable_off')
2. MetaSite ('most prob & interm prob_off')
3. MetaSite ('most probable_on for substrate')
4. MetaSite ('most prob & interm prob_on for substrate')
5. MetaSite ('most prob_on for substrate and CYP')
6. MetaSite ('most prob & interm prob_on for substrate and CYP')
7. StarDrop

Figure 5
Overview of best individual and combinational results concerning CYP predictions

1. Best individual settings for MetaSite
2. Best individual setting for StarDrop
3. Intersection (Meteor 'plausible 2' + MetaSite 'most probable_on for substrate')
4. Intersection (Meteor 'plausible 0' + MetaSite 'most probable_on for substrate')
5. Intersection (Meteor 'plausible 0' + StarDrop)
6. Intersection (MetaSite 'most probable_on for substrate' + StarDrop)
7. Union (MetaSite 'most probable_on for substrate' + StarDrop)
Overview of best individual and combinatorial results concerning phase I and phase II predictions

- 1 △ Meteor individual setting ‘plausible 0’
- 2 △ Meteor individual setting ‘plausible 2’
- 3 ● union (Meteor ‘plausible 2’ + MetaSite ‘most probable_on for substrate’)
- 4 ● union (Meteor ‘plausible 0’ + MetaSite ‘most probable_on for substrate’)
- 5 ● union (Meteor ‘plausible 1’ + MetaSite ‘most probable_on for substrate’)
- 6 ○ union (Meteor ‘plausible 1’ + StarDrop)
- 7 ■ intersection+2 (Meteor ‘plausible 0’ + MetaSite ‘most probable_on for substrate’)
- 8 ■ intersection+3 (Meteor ‘plausible 0’ + MetaSite ‘most probable_on for substrate’ + StarDrop)

Figure 7
Evaluation of three state-of-the-art metabolite prediction software packages (Meteor, MetaSite, and StarDrop) through independent and synergistic use

T'jollyn H., Boussery K., Mortishire-Smith R.J., Coe K., De Boeck B., Van Bocxlaer J.F., Mannens G.

**Supplemental data on RWJ-241947: illustration of the evaluation method**

**Information from the human AME report**

Information from the *in vivo* mass balance study of radiolabeled RWJ-241947 was obtained from the AME study. The amount of every metabolite detected in urine and/or faeces was expressed as a percentage of the dose administered. The total metabolic scheme is depicted in Figure 1 and the presence of the individual metabolites in urine and faeces is shown in Figure 2.

![Metabolic Scheme](image)

**Figure 1:** complete metabolic scheme of RWJ-241947 from the human AME study
Figure 2: chart represents percentages of administered dose of parent RWJ-241947 (UD) and its different metabolites in excreta.

Primary metabolic pathways

Next we extracted the primary metabolic pathways, originating in 1 step from the parent compound. The percentages of different metabolites in excreta that originated from one primary metabolic pathway were summed in order to obtain the overall importance of that metabolic path (Table 1), (Figure 3).

Table 1: percentages represent the main metabolic pathways by which RWJ-241947 is metabolized and excreted in urine and/or faeces

<table>
<thead>
<tr>
<th>Type</th>
<th>Metabolites</th>
<th>Urine (0-336h) Mean (%)</th>
<th>Faeces Mean (%)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-dealkylation</td>
<td>1,2,3</td>
<td>53.2</td>
<td>1.2</td>
<td>54.4</td>
</tr>
<tr>
<td>Thiazolidinedione open</td>
<td>6,7</td>
<td>3.2</td>
<td>0.8</td>
<td>4</td>
</tr>
<tr>
<td>-</td>
<td>UD</td>
<td>0</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>
Figure 3: chart represents the relative contribution of the main metabolic ways and excretion of unchanged drug (UD) in the metabolism of the drug compound. Percentages are relative to the total dose administered.
DRUG METABOLISM AND DISPOSITION

The software prediction output

Meteor

The prediction

Table 2: represents the prediction of Meteor for compound RWJ-241947 on the level ‘plausible 2’

<table>
<thead>
<tr>
<th>Prediction level</th>
<th>Predicted biotransformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROBABLE</td>
<td>Benzylic Hydroxylation (between the naphthalene and thiazolidine-2,4-dione ring)</td>
</tr>
<tr>
<td>PROBABLE</td>
<td>Sulphoxide Formation from Thioethers</td>
</tr>
<tr>
<td>PROBABLE</td>
<td>Oxidative O-dealkylation</td>
</tr>
<tr>
<td>PLAUSIBLE</td>
<td>4-Hydroxylation of 1,2-Disubstituted Benzenes</td>
</tr>
<tr>
<td>PLAUSIBLE</td>
<td>Oxidative Ring Opening of Thiazolidine-2,4-diones</td>
</tr>
</tbody>
</table>

In the case of RWJ-241947, Meteor predicts a benzylic hydroxylation relative to the naphthalene ring, sulphoxide formation, and an oxidative O-dealkylation on a probable level. On the plausible level a hydroxylation of 1,2-disubstituted benzenes, and oxidative ring opening of the thiazolidinedione ring are predicted. The biotransformations ‘sulphoxide formation’ and ‘ring opening of thiazolidinediones’ are considered as one biotransformation for RWJ-241947 because a sulphoxide formation could be the precursor of the ring opening. As is discussed in the main article, indiscriminate oxidation also occurs on this compound structure, of which the ‘Benzylic Hydroxylation’ and ‘4-Hydroxylation of 1,2-Disubstituted Benzenes’ are examples.

The evaluation

Table 3: represents the type of biotransformation as predicted by Meteor, labeled true, false, or missed

<table>
<thead>
<tr>
<th>Value</th>
<th>TRUE</th>
<th>FALSE</th>
<th>MISSED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotransformation</td>
<td>Oxidative O-dealkylation</td>
<td>4-Hydroxylation of 1,2-disubstituted benzenes</td>
<td>None</td>
</tr>
<tr>
<td>Thiazolidinedione</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ring opening</td>
<td></td>
<td>Benzylic Hydroxylation</td>
<td></td>
</tr>
</tbody>
</table>

In this project we considered only the ‘probable’ and the ‘plausible’ ARL with RRL 0, 1, and 2. In this example we calculated sensitivity and precision for setting ‘plausible 2’, for compound RWJ-241947, based on the data from Table 3.

\[
\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} = \frac{2}{2 + 0} = 1
\]

\[
\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} = \frac{2}{2 + 2} = 0.5
\]
**DRUG METABOLISM AND DISPOSITION**

**MetaSite**

The prediction

![Figure 4: displays RWJ-241947 with the metabolic hotspots as predicted by MetaSite with 'reactivity correction on for substrate and CYP'. Colors on the query compound correlate with those from Figure 5](image)

In the case of RWJ-241947, an oxidation is predicted with a high probability at the benzylic position on the fluorobenzyl (highlighted in blue), and on the sulfur atom of the thiazolidinedione ring (red). An oxidation with lower probability, relative to the first scores, is predicted on the benzylic carbon adjacent to the naphthalene moiety, and on the benzene ring, *meta* to the fluoro atom (brown).

The evaluation

**Table 4:** represents the type of biotransformation as predicted by MetaSite, labeled true, false, or missed

<table>
<thead>
<tr>
<th>Value</th>
<th>TRUE</th>
<th>FALSE</th>
<th>MISSED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotransformation</td>
<td>Benzylic Oxidation</td>
<td>Benzene Oxidation</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulphoxide Formation</td>
<td>Oxidation adjacent to naphthalene</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this project we considered only the ‘most probable’ (red) and the ‘intermediately probable’ (brown) predicted positions of MetaSite. In this example we calculated sensitivity and precision for setting ‘most prob & intermediately prob; on for substrate and CYP’, for compound RWJ-241947, based on the data from Table 4.

\[
\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} = \frac{2}{2 + 0} = 1
\]

\[
\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} = \frac{2}{2 + 2} = 0.5
\]
**DRUG METABOLISM AND DISPOSITION**

**StarDrop**

The prediction

<table>
<thead>
<tr>
<th>Site</th>
<th>3A4 Lability</th>
<th>Hydrogens</th>
<th>3A4 Ratio %</th>
<th>2D6 Ratio %</th>
<th>2C9 Ratio %</th>
</tr>
</thead>
<tbody>
<tr>
<td>S23</td>
<td>Labile</td>
<td>0</td>
<td>0%</td>
<td>99%</td>
<td>97%</td>
</tr>
<tr>
<td>C9</td>
<td>Mod Stable</td>
<td>2</td>
<td>90%</td>
<td>10%</td>
<td>1%</td>
</tr>
<tr>
<td>C5</td>
<td>Mod Stable</td>
<td>3</td>
<td>1%</td>
<td>99%</td>
<td>1%</td>
</tr>
<tr>
<td>Cl1</td>
<td>Mod Stable</td>
<td>1</td>
<td>1%</td>
<td>99%</td>
<td>1%</td>
</tr>
<tr>
<td>Cl4</td>
<td>Mod Stable</td>
<td>1</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Cl1</td>
<td>Mod Stable</td>
<td>1</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Cl2</td>
<td>Mod Stable</td>
<td>1</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Cl5</td>
<td>Mod Stable</td>
<td>1</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Cl6</td>
<td>Mod Stable</td>
<td>2</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Cl6</td>
<td>Stable</td>
<td>1</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Cl9</td>
<td>Stable</td>
<td>1</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Cl2</td>
<td>Stable</td>
<td>1</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Figure 6: showing a screenshot displaying the lability of compound RWJ-241947 to 3A4, the number of hydrogens situated on that position, and the different involvement ratios for 3A4, 2D6, and 2C9

In the case of RWJ-241947, StarDrop predicts the sulfur atom in the thiazolidinedione ring to be a labile position. The benzylic position on the fluorobenzyl ring is predicted as a moderately labile site of metabolism.

The evaluation

Table 5: represents the type of biotransformation as predicted by StarDrop, labeled true, false, or missed

<table>
<thead>
<tr>
<th>Value</th>
<th>TRUE</th>
<th>FALSE</th>
<th>MISSED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotransformation</td>
<td>Sulphoxide Formation</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Benzylic Oxidation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this project we considered only the ‘labile’ and the ‘moderately labile’ predicted positions of StarDrop. In this example we calculated sensitivity and precision for compound RWJ-241947, based on the data from table 5.

Sensitivity = \[
\frac{\text{true positives}}{\text{true positives} + \text{false negatives}} = \frac{2}{2 + 0} = 1
\]

Precision = \[
\frac{\text{true positivites}}{\text{true positivites} + \text{false positives}} = \frac{2}{2 + 0} = 1
\]
**Figure Legend**

Figure 1: complete metabolic scheme of RWJ-241947 from the human AME study

Figure 2: chart represents percentages of administered dose of parent RWJ-241947 (UD) and its different metabolites in excreta

Figure 3: chart represents the relative contribution of the main metabolic ways and excretion of unchanged drug (UD) in the metabolism of the drug compound. Percentages are relative to the total dose administered

Figure 4: displays RWJ-241947 with the metabolic hotspots as predicted by MetaSite with ‘reactivity correction on for substrate and CYP’. Colors on the query compound correlate with those from Figure 5

Figure 5: presents a more detailed view on the different scores that are assigned to every atom on the query compound

Figure 6: showing a screenshot displaying the lability of compound RWJ-241947 to 3A4, the number of hydrogens situated on that position, and the different involvement ratios for 3A4, 2D6, and 2C9

Figure 7: this screenshot from StarDrop displays the 3A4 Metabolic Landscape, only accounting for the influence of 3A4 metabolism on the compound RWJ-241947. A score per atom is attributed and displayed on the chemical structures, as well as in a bar chart
Table legend

Table 1: percentages represent the main metabolic pathways by which RWJ-241947 is metabolized and excreted in urine and/or faeces.

Table 2: represents the prediction of Meteor for compound RWJ-241947 on the level ‘plausible 2’.

Table 3: represents the type of biotransformation as predicted by Meteor, labeled true, false, or missed.

Table 4: represents the type of biotransformation as predicted by MetaSite, labeled true, false, or missed.

Table 5: represents the type of biotransformation as predicted by StarDrop, labeled true, false, or missed.