Driving Discovery

State-of-the-art computational approaches can now be used to predict metabolism by cytochromes P450, but many issues still need to be addressed

Matthew Segall
at Optibrium Ltd

Cytochromes P450 (P450s) are a superfamily of enzymes present in all organisms. They contain a haem moiety that forms the catalytic centre and primarily function as monooxygenases, adding an oxygen to their substrates (see Figure 1) (1). P450s are highly relevant in drug discovery because they are responsible for Phase 1 metabolism of approximately 70-80% of drugs (2). Several P450 isoforms contribute to human drug metabolism; the most important of these are summarised in Figure 2. Many issues affecting potential drugs can arise from P450 metabolism:

- Rapid metabolism can lead to poor bioavailability or rapid clearance, resulting in low exposure and efficacy
- Co-administration of drugs can cause drug-drug interactions (DDIs), whereby inhibition or induction of a P450 isoform by one therapeutic can dramatically affect the metabolism of another, causing significant changes in exposure
- Genetic polymorphisms in patient populations affect the rates of metabolism by major P450 isoforms, including CYP2D6 and CYP2C19, generating large variabilities in exposure between patients
- Metabolism by P450s can lead to bioactivation of compounds and the formation of reactive or toxic metabolites

In an effort to reduce these issues, *in vitro* assays of P450 metabolism-related properties are widely used in the drug discovery process (3). High-throughput assays, such as measurement of turnover in human liver microsomes or inhibition of P450 isoforms, are applied as early screens. However, more detailed studies, such as metabolite identification, remain expensive and time-consuming, hence they are typically applied later in the process to a limited number of compounds.

There has also been a concerted effort to develop computational – or *in silico* – methods to predict the interactions of compounds with P450s. These can be inexpensively applied to large numbers of virtual compounds and provide detailed information to guide the design of new compounds before synthesis, overcoming P450-related issues.

Modelling Methods

To build and validate *in silico* models, high-quality data for a sufficient number of diverse compounds are essential. These data are most readily available in big pharmaceutical companies where *in vitro* assays have been run routinely for many years. However, even here, inconsistencies between assay protocols can limit the quantity of suitable data.

When relying on public domain data, it is essential to ensure that the quality and comparability of information from different sources are carefully assessed. Ideally, the results from *in vitro* assays should be referenced to clinical results to ensure their relevance; inappropriate assay conditions, such as high substrate concentrations that are not clinically relevant, can lead to misleading results.

Many methods have been applied to modelling P450-compound interactions:

Quantum Mechanical (QM)
These have been used to understand the mechanisms of P450-mediated metabolism and estimate reaction energetics (4). The main limitation of QM methods is that they are slow, although this can be mitigated by using faster semi-empirical methods.

Quantitative Structure-Activity Relationship (QSAR)
QSAR models use statistical and machine-learning methods to relate characteristics of compounds to their metabolic properties. However, they are held back by requirements for large data sets of compounds with comparable experimental measurements to build accurate and generally applicable models, particularly for complex metabolic properties.

Structure-Based Methods
These use 3D models of the structures of P450 isoforms to predict the binding of substrates and inhibitors. A challenge for structure-based methods is the relative scarcity of crystal structures of human P450 enzymes. Furthermore, some isoforms exhibit enormous flexibility, limiting the value of a single 'snapshot' from a crystal structure.

The remainder of this article will cover the most important P450 metabolism-related endpoints and, for each, how state-of-the-art *in silico* models and resulting predictions are used to guide the optimisation of compounds in drug discovery.

Inhibition

The inhibition of P450 isoforms by small molecules is most commonly predicted using QSAR methods based on measured inhibition data: IC_{50}, K_i or percentage inhibition. The variability of experimental inhibition measurements between different assay protocols – for example, different probe substrates – means that combining data from multiple sources can be very difficult.

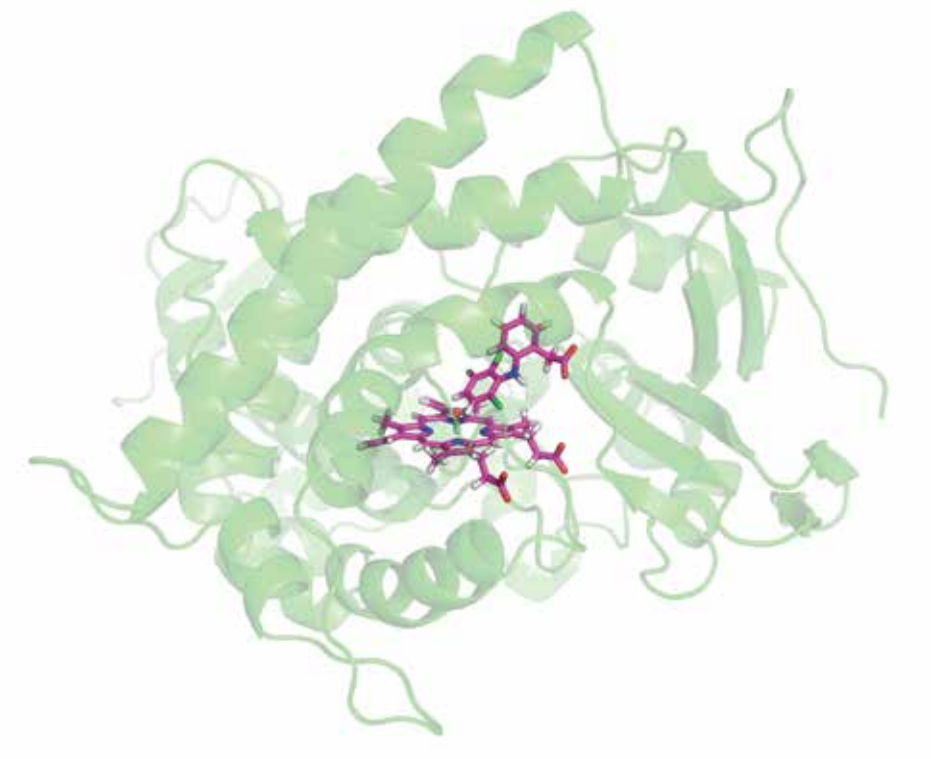

Image: ©Rasmus Leth

Figure 1: Illustrative structure of P450 isoform BM3 in which the reactive oxy-haem moiety (responsible for oxidation of the substrate) and the substrate diclofenac are highlighted

Some large datasets of consistent P450 inhibition information have been published, but the use of high-throughput assays means that these data are usually treated categorically (eg high/low) (5). The resulting predictions can be used to identify potential isoforms of concern and prioritise experimental studies (6). Quantitative models of P450 inhibition, based on high-quality, consistent data can be applied to predict K_i or IC_{50} values. These provide an assessment of the degree of inhibition and help to design new compounds with a reduced risk of high inhibition and DDIs.

Regioselectivity of Metabolism

Predicting the site of metabolism (SOM) of a compound by a specific P450 isoform has been the subject of research for decades. A review by Kirchmair *et al* gives an overview of the many methods that have been applied (7). Accurately predicting SOM and the proportion of metabolism observed at each site (the regioselectivity) requires modelling two contributing factors:

- The inherent reactivity of each site to oxidative attack, usually via abstraction of a hydrogen or direct oxidation by a highly reactive oxy-haem. Here, quantum mechanical simulations provide the most accurate methods to estimate the energetics of the reactions at each potential SOM (4,8)
- The different binding sites of each P450 isoform impose constraints on the orientation of substrates relative to the oxy-haem and hence the accessibility of a potential SOM. Therefore, individual models of these effects must be built for each isoform. QSAR approaches, based on experimental regioselectivity data and structure-based methods that dock potential substrates with 3D models of the binding sites, have been used for this (4,9,10)

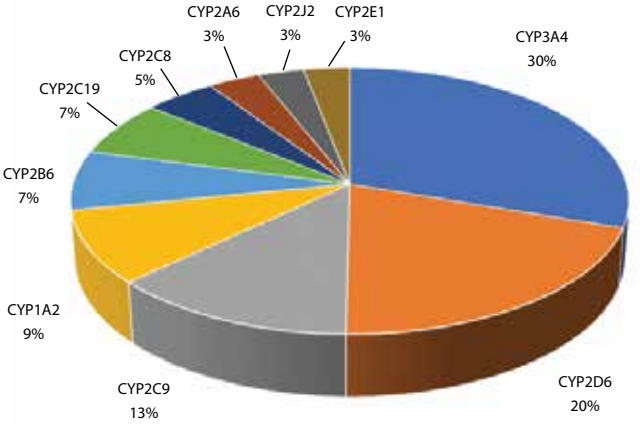

Figure 2: The primary P450 isoforms responsible for human drug metabolism, and the proportion of marketed drugs they have been observed to metabolise (2)

Figure 3 shows an example of regioselectivity predictions for the metabolism of Venlafaxine by P450 isoforms CYP3A4 and CYP2D6. Here you can see the different SOMs predicted for these isoforms in agreement with experimental observations. From predictions of the SOMs, it is usually straightforward to identify the resulting metabolites, as also illustrated in Figure 3. However, the formation of reactive metabolites usually occurs by less common mechanisms and remains difficult to reliably predict (11).

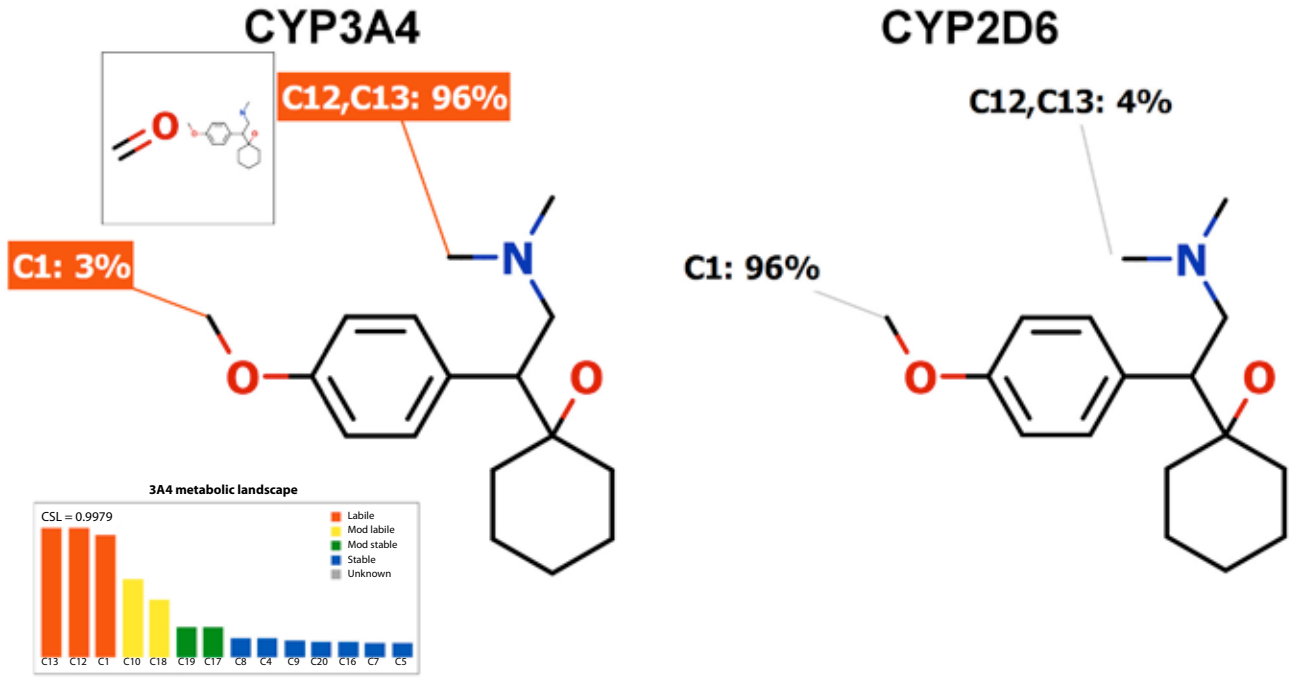

Figure 3: Example predictions of the regioselectivity of P450 metabolism of the drug Venlafaxine by CYP3A4 and CYP2D6, showing the different proportions of metabolism for each isoform, in agreement with experimental observations. Inset, the structure of the metabolite resulting from metabolism at the C12,C13 position is shown. In the bottom-left, an example of a metabolic landscape indicates the vulnerability of each site to metabolism by CYP3A4 on an absolute scale

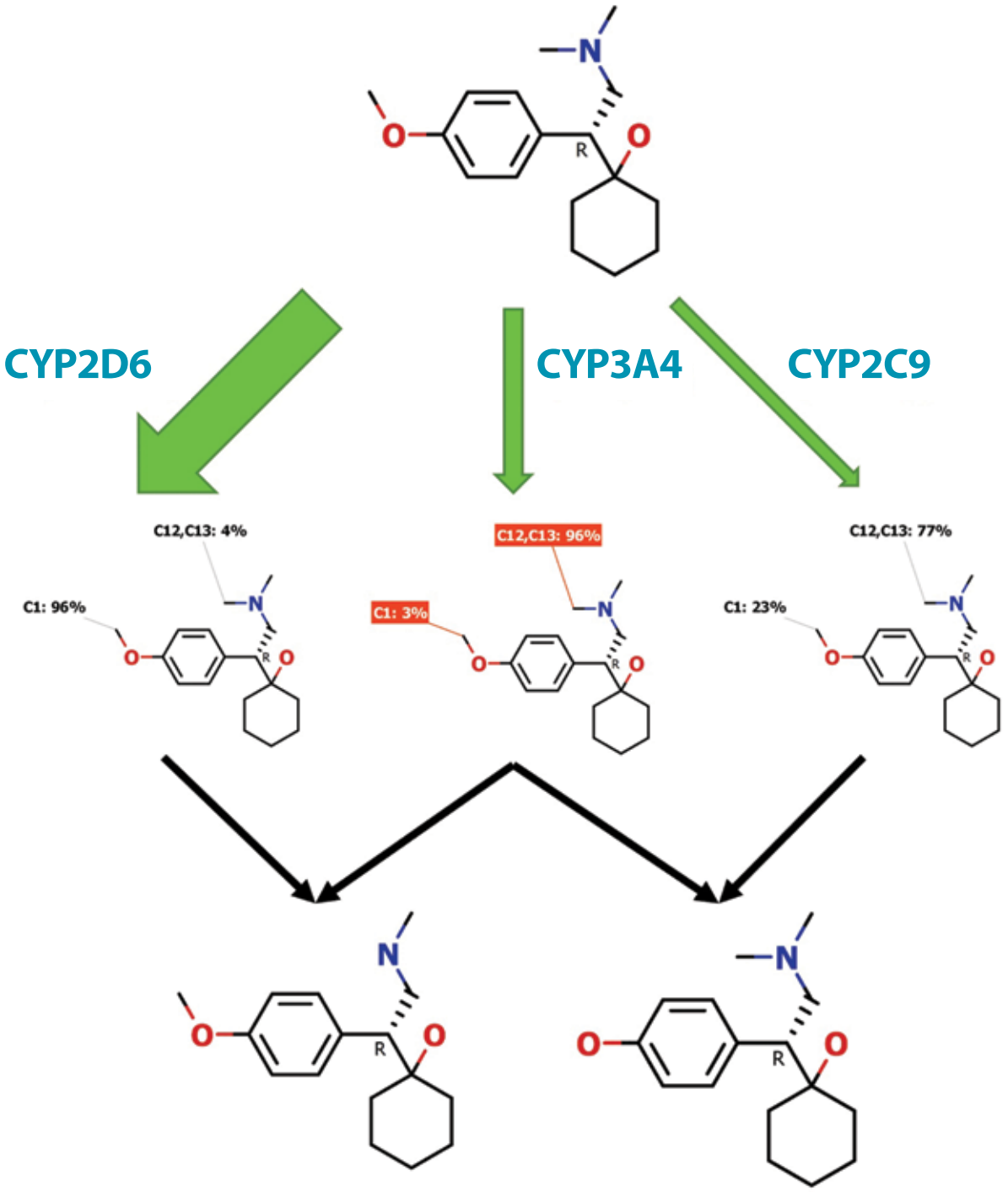

Figure 4: Prediction of significant P450 metabolites for Venlafaxine. The green arrows indicate the predicted major isoforms responsible for metabolism of Venlafaxine; the width of the arrows is proportional to the predicted contribution of each isoform. Below these are predictions of the regioselectivity of metabolism by each of the major isoforms, and the corresponding metabolites are shown at the bottom

Identifying the most important SOMs can help to redesign a compound to improve metabolic stability. However, the relative proportion of metabolism only tells part of the story; it is also essential to understand the vulnerability of each site in absolute terms. Methods that estimate the energetics of the reaction mechanism for each site can therefore provide additional information to guide optimisation of metabolic stability (4). An example of this is shown as a 'metabolic landscape' in Figure 3.

Isoforms Responsible for Metabolism

Multiple P450 isoforms are involved in drug metabolism and may contribute to the metabolism of even a single compound. It is therefore helpful to predict the specificity of each isoform, ie which are responsible for the metabolism of a given compound. This indicates which regioselectivity prediction(s) are most relevant. Additionally, if the metabolism of a compound is dominated by a single isoform, this increases the risk of DDIs;

inhibition of this isoform by a co-administered drug could dramatically reduce its metabolic clearance. Ideally, a compound would be cleared via multiple pathways to reduce this risk, provided the overall clearance is not too high.

If a compound inhibits an isoform, it does not necessarily mean that it will be a substrate. Therefore, inhibition data are not sufficient to predict which P450 isoforms are likely to metabolise a compound. Instead, data must be carefully collected relating to which isoforms contribute to the metabolism of compounds. Models of 'which P450' are less common than those of P450 inhibition; however, similar QSAR approaches are often applied to this question (12-14).

Metabolite Profile

Combining predictions of P450 isoform specificity with regioselectivity can indicate the expected profile of Phase 1 metabolites to which a patient will be exposed (see Figure 4, page 23). The relative abundances of the different isoforms also affect the predicted profile of metabolites, which varies significantly between different tissues. It is important to identify potentially active, reactive or toxic Phase 1 metabolites, and this profile will also be indicative of the pathways for Phase 2 metabolism – eg conjugation or other routes of elimination.

Induction

The *in silico* prediction of induction of P450 isoforms has been less well-studied, probably due to the limited availability of data for a wide range of chemical diversity. The main approach for modelling this property has been through building QSAR models of compound interactions with the pregnane X receptor and constitutive androstane receptor, which have been implicated in P450 induction (15,16). Accurate prediction of P450 induction would be useful to identify potential risks of DDIs.

Final Predictions

The field of P450 metabolism prediction is very active, and recent developments have improved the reliability and applicability of models in practical drug discovery projects, which are now routinely used to guide compound optimisation. Nevertheless, much work remains, and a particularly important area of research is prediction of the formation of reactive and toxic metabolites, which have been implicated in drug-induced liver injury and idiosyncratic toxicity. Better understanding of P450 induction would also complete the range of predictive models important for drug-P450 interactions.

References

1. Guengerich FP, Cytochrome P450s and other enzymes in drug metabolism and toxicity, *AAPS J* 8(1): E101-11, 2006
2. Zanger UM and Schwab M, Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation, *Pharmacol Ther* 138(1): pp103-41, 2013
3. Tarbit MH and Berman J, High-throughput approaches for evaluating absorption, distribution, metabolism and excretion properties of lead compounds, *Curr Opin Chem Biol* 2(3): pp411-6, 1998
4. Tyzack JD *et al*, Predicting regioselectivity and lability of cytochrome P450 metabolism using quantum mechanical simulations, *J Chem Inf Model* 56(11): pp2,180-93, 2016
5. Veith H *et al*, Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries, *Nat Biotechnol* 27(11): pp1,050-5, 2009
6. Rostkowski M *et al*, WhichCyp: Prediction of cytochromes P450 inhibition, *Bioinformatics* 29(16): pp2,051-2, 2013
7. Kirchmair J *et al*, Computational prediction of metabolism: Sites, products, SAR, P450 enzyme dynamics, and mechanisms, *J Chem Inf Model* 52(3): pp617-48, 2012
8. Rydberg P *et al*, SMARTCyp: A 2D method for prediction of cytochrome P450-mediated drug metabolism, *ACS Med Chem Lett* 1(3): p96-100, 2010
9. Zaretzki J *et al*, RS-Predictor models augmented with SMARTCyp reactivities: Robust metabolic regioselectivity predictions for nine CYP isozymes, *J Chem Inf Model* 52(6): pp1,637-59, 2012
10. Tyzack JD *et al*, Prediction of cytochrome P450 xenobiotic metabolism: Tethered docking and reactivity derived from ligand molecular orbital analysis, *J Chem Inf Model* 53(6): pp1,294-305, 2013
11. Leth R *et al*, Density functional theory study on the formation of reactive benzoquinone imines by hydrogen abstraction, *J Chem Inf Model* 55(3): pp660-6, 2015
12. Mishra NK *et al*, Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule, *BMC Pharmacol* 10: p8, 2010
13. Terfloth L *et al*, Ligand-based models for the isoform specificity of cytochrome P450 3A4, 2D6, and 2C9 substrates, *J Chem Inf Model* 47(4): p1,688-701, 2007
14. Hunt PA and Segall MD, Which P450 – A multi-class categorical models to predict the major metabolising CYP450 isoform for a compound, (submitted), 2017
15. Wilson TM and Kliewer SA, PXR, CAR and drug metabolism, *Nature Rev Drug Discov* 1(4): pp259-66, 2002
16. Ekins S *et al*, Challenges predicting ligand-receptor interactions of promiscuous proteins: The nuclear receptor PXR, *PLoS Comput Biol* 5(12): e1000594, 2009

About the author

Matthew Segall has an MSc in computation from the University of Oxford, UK, and a PhD in theoretical physics from the University of Cambridge, UK. At Camitro, ArQule and Inpharmatica, he led a team developing predictive absorption, distribution, metabolism and excretion (ADME) models and intuitive decision-support tools for drug discovery. In 2006, Matt became responsible for management of Inpharmatica's ADME business and, following acquisition of Inpharmatica, he took up the role of Senior Director of BioFocus DPI's ADMET division. In 2009, he led a management buyout of the StarDrop business to found Optibrium Ltd, which develops software for small molecule design, optimisation and data analysis.

Email: matt.segall@optibrium.com