Predicting Drug Metabolism: Experiment and/or Computation?

Johannes Kirchmair,[1,2*] Andreas H. Göller,[3] Dieter Lang,[4] Jens Kunze,[1] Bernard Testa,[5] Ian D. Wilson,[6] Robert C. Glen,[6,7] Gisbert Schneider[1*]

[1]Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, ETH Zürich, 8093 Zürich, Switzerland; [2]Center for Bioinformatics, University of Hamburg, 20146 Hamburg, Germany; [3]Global Drug Discovery, Bayer Pharma AG, 42096 Wuppertal; [4]Drug Metabolism, Bayer Pharma AG, 42096 Wuppertal, Germany; [5]Lausanne University Hospital, CH-1011 Lausanne, Switzerland; [6]Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ London, United Kingdom; The Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, CB2 1EW, United Kingdom.

[*]Correspondence to J. K. (kirchmair@zbh.uni-hamburg.de) and G. S. (gisbert.schneider@pharma.ethz.ch)

DEDICATION

We dedicate this work to the memory of Jörg Keldenich: experimentalist and pioneer for PK/PD and ADME/Tox modelling at Bayer.

ABSTRACT

Drug metabolism can produce metabolites with physicochemical and pharmacological properties that differ significantly from those of the parent drug. These mechanisms are an important factor in the high attrition rates currently found in drug discovery and development, including late-stage clinical trials. The relevance of drug metabolism for both safety and efficacy implies a vital requirement for efficient and reliable ways to predict drug metabolism *in vitro*, *in silico*, and in intact organisms. In this Perspective, we provide an overview of the state-of-the-art of experimental and computational approaches for investigating drug metabolism. We highlight the scope and limitations of these methods and indicate strategies to harvest the synergies resulting from combining measurement and prediction.

INTRODUCTION

Metabolism is a signature of living systems, and enables organisms to create a viable environment within which to perform the complex biochemical transformations that maintain homeostasis. The metabolic system has evolved as the main line of defence against foreign, hazardous substances, by transforming them into readily excretable metabolites. These xenobiotics also include synthetic drugs, in addition to naturally occurring substances (e.g. bacterial, fungal and herbal toxins), often produced as part of a specific defence mechanism. Recent studies have highlighted the often synergistic signalling between the host organism and the microbiome facilitated by metabolism[1]. For about 75% of all drugs metabolism is one of the major clearance pathways[2,3]. The process of biotransformation can produce metabolites with substantially altered physicochemical, physiological, pharmacological, and toxicological profiles[4-6]. Metabolism is the main factor mediating the activation, deactivation, toxification and detoxification of small molecules, and is a key determinant in the performance and safety of small molecules, including their uses in medicines, cosmetics, nutritional supplements and agrochemicals (Box 1).

<<Box 1 approx. here>>

The complexity and malleability of the metabolic system cannot be overstated, A plethora of diverse enzyme families are involved in xenobiotic metabolism, including cytochrome P450s (CYPs), dehydrogenases, flavin-containing monooxygenases, hydrolases, peroxidases, UDP-glucuronosyltransferases, sulfotransferases, and glutathione S-transferases[7]. Their expression patterns and substrate specificities can vary greatly among different species, which implies there is a risk of missing toxic metabolites formed in humans when extrapolating from *in vitro* and animal testing results. Expression patterns also differ between tissues and organs, and there are indications that metabolic enzymes engage in synergistic collaborations with transporters (e.g. CYP 3A4 and P-glycoprotein[8,9]). On top of this, many inter- and intra-individual factors need to be considered, such as gender differences and genetic polymorphisms, age, biological rhythms, pregnancy, intestinal flora, diseases, stress, lifestyle, diet and medication[4]. Hence even with the advanced technologies and knowledge available today, accurate prediction remains highly challenging.

Understanding metabolic processes on a molecular level of detail is of fundamental importance and key to successful drug discovery and development. Knowing the metabolic properties of a molecule can help to optimize the stability and in consequence the *in vivo* half-life and risk-benefit ratio of a drug. A plethora of experimental methods are available for investigating the metabolic fate of drugs at an unprecedented level of detail. However, these experimental approaches remain demanding with respect to scientific equipment, human expertise, cost and time, which have acted as major drivers for the development of computational tools for drug metabolism prediction. *In silico* methods allow the prediction of the metabolic fate of virtual compounds and to plan for the most promising strategy to optimize metabolic stability of project compounds *a priori*.

Experimental and theoretical approaches are (still) all too often regarded as separate domains. It is most important to realize that there is enormous potential for synergy in the combination of both areas, which will allow the analysis and prediction of drug metabolism to make a major leap forward.

BIOSYSTEMS AND ANALYTICAL METHODS FOR INVESTIGATING DRUG METABOLISM

CYPs are of major significance to drug metabolism, and the most relevant forms are expressed predominantly in the liver, but also in a whole variety of other organs at lower expression levels (Figure 1). Hence liver, or liver-derived *in vitro* systems are often the most convenient, interesting and important experimental model systems for metabolism studies, e.g. when considering the first-pass metabolism of an orally bioavailable drug. Because of the considerable variability in metabolism among different species it is essential and valuable early in drug discovery to use *in vitro* systems with human-derived material, since man is generally the target species. Several *in vitro* systems, as well as new *in vivo* approaches are available and can be specifically used depending on the issue or problem to be solved (Table 1). The simplest systems are recombinant enzymes (expressed together with coenzymes to achieve optimal catalytic activity). They can be used e.g. as a single CYP isoform system to assess which isoforms are involved in the metabolism of a compound or for drug interaction studies. From native material, e.g. liver, different enzyme families can be separated by centrifugation. Soluble cytosolic enzymes, e.g. sulfotransferases (SULTs) remain in the supernatant, whereas membrane-bound enzymes like CYPs, UDP-glucuronosyltransferases (UGTs) etc. are enriched in the pellet and after re-suspension, the material described as "microsomes" (membrane vesicles of the endoplasmic reticulum) are obtained. Microsomal preparations can be easily stored while retaining their functionality, and are convenient to use because they are also available in sufficient quantity for high throughput assays.

<< Figure 1 approx. here>>

<< Table 1 approx. here>>

Another subcellular matrix is the S9-fraction, which consists of both the cytosolic and microsomal fractions. In all subcellular fractions, due to dilution or washout of cofactors during the preparation process, cofactor-dependent enzymes in microsomes, e.g. CYPs or UGTs need to be specifically supplemented with the cofactor (NADPH in the case of CYPs) to regain enzymatic activity. The

metabolically most competent system is one having intact liver cells (hepatocytes), with all appropriate cofactors, coenzymes etc. present in physiological conditions and environment. Hepatocytes, while being less "robust" than microsomes, are useful when a more comprehensive picture of the hepatic metabolism of a selected compound is required. Both fresh and commercially produced cryopreserved primary hepatocytes, or immortalized hepatocyte-derived cell lines can be easily produced from preclinical species, and there is now more straightforward access to human derived material. New preclinical models such as "humanized mice" (either mice with specific human drug-metabolising genes inserted[10] or chimeric animals where human hepatocytes have replaced >90% of the murine hepatocytes in the liver[11]) are becoming available which may improve studies of human metabolism, drug-drug interactions etc. in an *in vivo* animal model very early in drug discovery or development, significantly before entering clinical studies[12,13].

Currently high-performance liquid chromatographic (HPLC) separation systems coupled to mass spectrometry (MS) is the workhorse for detecting and characterising drug metabolites. The metabolite data generated by LC-MS analyses are in general qualitative in the absence of synthetic reference material, since parent drug and the metabolites formed can have dramatically different MS-responses. Integrating UV-detectors in these systems can provide at least semi-quantitative data, providing that metabolism does not affect the chromophore, and this can be used to assess major or minor metabolic pathways. Clearly, quantitative data without standards (ideally stable isotope labelled if LC-MS is to be used for quantification), particularly from complex *in vivo* matrices can only be generated using radiolabeled compounds and these (like the synthetic standards of the metabolites themselves) are rarely available during the discovery phase. Depending upon the spectrometer used, with high resolution mass spectrometry (HRMS) the preferred option, a suite of MS-based experiments can be deployed to help characterize metabolites, e.g. MS/MS, MS^n, MS^e, mass defect filters (MDFs), neutral losses[14,15].

The experienced MS operator can use a variety of strategies including knowledge, rules and databases[16] for metabolite detection and identification (MetID) to deduce lists of masses to be monitored. Indeed, all of the major mass spectrometer manufacturers offer bespoke metabolite identification software for this purpose. This can be supplemented by *in silico* tools such as IsoScore[17] or Mass-MetaSite[18,19] that

predict the metabolic fate to guide metabolite detection and identification[20]. The expert system Meteor Nexus[21] includes functionality for linking metabolites predicted for phase I and II biotransformations to mass spectrometry data. It highlights matches in the graphical depiction of the metabolic tree. Also tools based on the SyGMa[22] metabolite prediction method have been successful in supporting MetID efforts for drugs[23], and recently an open source library for LC-MS data processing, MassCascade[24], has been released. Another method is to search for metabolites using multivariate statistical analysis of the data (this approach is derived from the developing field of metabonomics/metabolomics where normally the data is used to deduce the relationships between the metabolism and phenotypic readouts such as clinical endpoints[25]). Comparison of the MS data from a sufficient number of samples obtained from dosed and control *in vitro* incubations, or animal experiments, can be mined using techniques such as principal components analysis for changes in metabolic profiles. Whilst some of these changes may well result from alterations in the endogenous metabolites, other changes will be due to the production of drug metabolites[26,27].

Despite the availability of these MS-technologies, and the various *in silico* aids it is often difficult, or impossible, to assign a full constitutional structure. The use of techniques such as (LC-)NMR spectroscopy may enable a more complete structural characterization of unknown metabolites provided sufficient material is available[28]. However, full metabolite characterization is generally only undertaken on a "for cause" basis, or after a compound has entered full development[14]. In the initial stages of discovery programs a simple measure of compound stability (i.e., rate of substrate disappearance) in hepatic microsomes or primary hepatocytes may suffice. Based on the data generated here, those molecules with greater metabolic stability can then be chosen for further investigation and chemical optimization. This type of analysis can easily be automated, and medium- to high-throughput metabolic stability assays are routinely applied[29-31]. Such assays can also be applied to investigations of a compound's susceptibility to metabolism by specific enzyme systems.

Once appropriate structures have been selected for lead optimization, additional metabolic information helps to identify metabolic "hotspots" so that *in vivo* breakdown can be restricted and exposure improved. Rapid generic analysis based on a reversed-phase gradient chromatographic separation may be employed for

metabolite profiling. Such methods can indicate the general nature of the biotransformations occurring (hydroxylation, dealkylation, reduction, conjugation reactions etc.) and provide insights into the SoMs as well as the enzyme classes involved. Microsomal systems are also useful for examining candidates for the generation of reactive intermediates based on so-called "trapping" experiments[32,33]. Such screening aims to detect both "soft" and "hard" nucleophiles formed as reactive metabolites *via* panels of trapping reagents.

As programs progress through the discovery phases, studies of more complex systems such as hepatocytes will be deployed, giving a more complete picture of hepatic metabolism, including further information about the generation of reactive intermediates. Finally, keeping in mind potential issues with allometric scaling, studies in animals provide the opportunity to investigate the *in vivo* metabolism, distribution and excretion of a selected number of candidate compounds, often using a much more bespoke chromatographic separation to maximize the data recovery from samples such as urine, bile and tissue extracts. However, the difficulties of obtaining truly comprehensive metabolite profiles in the absence of radiolabelled materials (or some other tracer) in biofluids should not be underestimated. The presence of endogenous metabolites increases the complexity of detection while effects such as ion suppression may attenuate signals for candidate drug metabolites.

COMPUTATIONAL APPROACHES FOR INVESTIGATING DRUG METABOLISM

A wide array of computational methods and integrated approaches have been developed for the prediction of drug metabolism, and are distributed in the form of web services and as free and commercial software (Table 2)[34,35]. They may be classified as specific ("local") or comprehensive ("global") tools[36,37]. Specific models apply to certain biomolecules (mainly metabolic enzymes) and/or to specific metabolic reactions, while global models are in principle applicable to diverse biological systems (i.e. to any metabolic enzyme and biotransformation) and to most small organic compounds. The goal of many metabolism software packages generally lies in combining various tools and methods not for a single enzyme, but for the largest possible number of targets related to drug metabolism. The inclusion of other

functional proteins, which may synergize with metabolising enzymes, such as transporters (e.g. P-glycoprotein[9]) can also be envisaged.

Among the most common applications of computational methods is the determination of fundamental structural, functional and mechanistic properties of biomolecules related to drug metabolism, which enables the identification of metabolically labile positions (SoMs) in small organic molecules and the prediction of metabolites. Once the chemical structures of the actual metabolites are determined, a good starting position for predicting their reactivity, toxicity, bioactivity and other pharmacokinetic and pharmacodynamic properties has been obtained. An overview of the scope and limitations of computational methods is provided in Table 3.

<<Table 2 approx. here>>

<<Table 3 approx. here>>

Components to successful prediction of drug metabolism

Virtually any technique of computational chemistry has been adopted for the task of drug metabolism prediction, but it has become clear that the key to success is the integration of various methods and resources.

Component 1: Experimental data

Computational models are often (but not exclusively) based on experimental data, and the amount and quality of the available data will determine their coverage and performance. Experimental data are generally generated under controlled conditions, meaning that a number of variables are set and kept constant, leaving large portions of the space of possibilities unexplored. As an example, the concentration and distribution of metabolic enzymes *in vivo* may not be equivalent to those from which the experimental model was derived. Computational models can therefore be false, irrelevant, or, at best, incomplete.

Recently, efforts in making experimental data on small organic molecules available to the public have clearly increased[38], but data on drug metabolism is still

sparse. There are only a few databases of relevance to this topic in the public domain[35] (Table 4), and even proprietary data collections are surprisingly limited in their coverage. Data collections on metabolism have been built up over decades and are often not stored in a machine-readable format.

Assay technologies, protocols, requirements for new medicines, definition of targets and the chemical space of interest etc. have changed during the last years, resulting in heterogeneous datasets. Apart from the limited quantity and coverage, data quality is a further concern, including issues related to experimental variability and errors introduced during data collection, curation and manipulation[37,39,40]. When searching the literature (or in-house archives) for metabolic data, one is all too often confronted with unconvincingly documented or unsolved constitutional assignments, especially for minor metabolites. For example, sites of hydroxylation or conjugation (i.e. product regioselectivity) are often left unassigned. Such partly solved metabolic structures certainly have some utility, but they must be clearly designated as such to avoid an over- or misuse. Configurational (i.e. stereochemical) aspects of metabolite formation and characterization have received considerably less attention than constitutional aspects[41,42], despite their obvious pharmacological significance. For all these reasons, distilling high quality data that can be used for training computational models is a challenging and labour-intensive task that needs to be done by experts.

Experimental data such as bioactivities can be modelled using (Q)SAR techniques. These methods have evolved from univariate statistical approaches to multivariate machine learning techniques based on heuristics. Classical (Q)SAR methods apply linear regression techniques to fit experimental data. Regression (numerical) models have historically been used to model continuous responses (such as pK_D) from high-quality datasets, while classification (categorical) models are often favoured for modelling noisy data (e.g. when collected from different laboratories or from assays with a high intrinsic variability). Machine-learning methods generally outperform classical approaches on large and highly diverse datasets featuring complex non-linear relationships. With the rapid increase of available data these methods are becoming increasingly popular, despite obvious drawbacks such as limited interpretability[37]. However, it may be unwise to dismiss such models for the sole reason of lack of (linear) interpretability. Here, chemists can learn from other

disciplines, e.g. physics and engineering, that very successfully employ nonlinear modelling approaches for solving complex problems.

(Q)SAR models have primarily been reported for the interaction of molecules with specific CYP isoforms, but also UGTs and sulfotransferases, and receptors involved in enzyme induction, in particular the CYP-inducing AhR (aryl hydrocarbon receptor), CAR (constitutive androstane receptor), and PXR (pregnane X receptor)[37,43,44]. On a broader scale, statistical and machine learning methods are used to predict comprehensive bioactivity spectra of small molecules. However, current mechanistic models generally do not take into account information beyond target annotation, and hence are limited in their ability to predict phenotypes[45]. Linking pathway information to targets can hence improve model accuracy. MetaCore and MetaDrug46,47, two pharmacology platforms that use comprehensive biological networks for estimating the pharmacological effects and risk of small molecules, harness this approach. However, the coverage and accuracy of annotated pathways is still a work in progress and there is often significantly less overlap between different databases of annotated pathways than one would expect. This is often compounded by lack of information on tissue expression levels and rate constants for metabolising enzymes. There clearly is an opportunity to invest additional effort in this area to further increase the reliability of computation models.

The majority of toxicity models available to date are trained solely on toxicological endpoints. These quantitative structure-toxicity relationship (QSTR) in general make use of only a few basic chemical descriptors in combination with classic linear modelling techniques. They serve primarily as hazard identification tools to support the general risk assessment. Only rarely they are derived for exposure-response relationships that allows the prediction of absolute toxicity in isolation[39].

Biotransformation data can be used to derive models for predicting both the sites and products of metabolism in an automated fashion. MetaPrint2D[48], for example, generates simple statistical models for SoM prediction from biotransformation databases. An extension of this software, MetaPrint2D-React[48], identifies and encodes the type of metabolic reaction observed for specific atom environments. The program is able to generate the chemical structures of likely metabolites by applying reaction rules to predicted SoMs. Realistically, such data mining approaches require thousands of biotransformation records for proper model

development, which with the present sparsity of data, is one of their major limitations. These methods therefore have a limited applicability.

<<Table 4 approx. here>>

Component 2: Expert knowledge

Given the limited quality and quantity of available data, and the wealth of empirical knowledge that medicinal chemists have accumulated over decades of research in drug metabolism, scientists have been keen to devise sets of rules (dictionaries) from expert human knowledge, and to develop reasoning engines to apply this knowledge to metabolite structure prediction[49]. Knowledge-based approaches such as Meteor[21] predict the sites and products of metabolism by scrutinizing a molecule of interest for the presence of target fragments. Their key advantage is the provision of the rational basis underlying a prediction (e.g. literature references and brief descriptions). However, these tools often generate a combinatorial explosion of predictions, as all possible combinations of transformations encoded by the dictionary can be performed, for several (usually three) generations of metabolites. Without pruning of the product trees produced, large numbers of metabolites may be presented to the user, and their analysis requires experience and diligence again. Hence the ranking of metabolites and the definition of adequate cut-off criteria are key challenges of knowledge-based approaches, which requires additional components to be integrated into such expert systems.

Rule-based systems are also firmly established in predictive toxicology, where they are used to interrogate molecules for structural elements linked to toxicity[50]. They have received considerable interest from regulatory authorities concerned with medicines registration (e.g. [51]) and are typically included as part of the guidance in investigating the toxic liability of new substances including drugs.

Component 3: Physicochemical properties

Sufficiently water-soluble compounds are likely to be excreted without undergoing metabolism. Expert systems and many other predictors make extensive use of this fact and other computed physicochemical properties such as log*P*

(octanol/water) or log*D*, the pH-dependent partition coefficient, as a means of metabolite ranking and filtering. A note of warning is apposite here, since more and more physicochemical parameters used in structure-metabolism relationships are calculated ones using common software whose limits and levels of precision are ignored by users. Also the effects of transporters on excretion of small molecules can be profound. For example glucose is almost 100% re-absorbed in the kidneys from the ultra filtrate despite being very hydrophilic.

Component 4: Target structure

Ligand-based methods need to deal with significant uncertainty about the target structure, specifically the ligand-receptor interaction site. Structure-based methods additionally take into account structural properties of the target, a key component for understanding protein-ligand interactions at an atomic level of detail. Only twelve years ago the first crystal structure of a human microsomal CYP (2C9) was resolved[52]. Today about one hundred crystal structures of CYPs are available covering most isoforms relevant to human drug metabolism. This is a respectable collection of potentially valuable data, but it is important to note that these structural models cover only a fraction of the enzymes' conformational space relevant to the binding of small molecules[53].

Automated ligand docking can be useful for examining whether a specific site on a molecule is likely to bind to a specific locus in a target protein. SoMs can be predicted by relating the proximity of ligand atoms in a computed docking pose to the catalytic centre of the target enzyme. The approach is generally able to correctly predict the approximate ligand orientation within the binding pocket (for binding sites of low flexibility) and provide a structural hypothesis for the observed biological response[53-55]. The technique is particularly valuable for rationalising e.g. diverging biological properties of enantiomers. In fact, we advocate the prediction of metabolic stereoselectivity be added as a benchmark of the maturity of substrate-enzyme docking simulations. Current automated docking methods are not particularly promising for 3D-QSAR modelling because of the poor performance of the available scoring functions[56]. Specifically, while enthalpic contributions to ligand-receptor binding for high affinity molecules are more or less captured by the various scoring functions that guide the ligand docking process, the entropic contributions (and other

more subtle effects like changes in the heat capacity of the complex) elude these techniques. There are remarkable differences in the performance of the various docking algorithms but the limitations of this method are primarily defined by (structural) data availability (in particular the coverage of relevant protein conformations), expert knowledge of the target, the algorithm itself and model parameterization.

Steric and chemical properties of target macromolecules can be described by molecular interaction fields (MIFs). MIFs are generated by embedding the target into a grid and computing interaction energies with a chemical probe for each grid node[57,58]. A probe is a small chemical fragment used to characterize a specific type of chemical features, most commonly hydrophobic features and hydrogen bonding regions. MIFs can be visualized as isopotential 3D maps indicating the directionality and topology of the interactions formed between a ligand and the receptor. MetaSite[59] is such an integrated software package for metabolism prediction that is based on MIFs. Among several other components, it features a fast docking surrogate method to relate the structure of small molecules to CYPs for SoM and metabolite prediction.

Component 5: Target flexibility

Metabolic enzymes and effectors involved in their regulation are known for remarkable ligand promiscuity. The plasticity and size of their binding sites (some of them have two or more) is a direct result of their function, which in the case of xenobiotic-metabolising enzymes is to provide a flexible and adaptable system for processing a wide range of substrates. Today, molecular dynamics (MD) simulations (often in combination with quantum chemical methods) are the most powerful theoretical approaches for analysing and predicting the interactions of protein-ligand pairs, and much of our knowledge about the structure, function, specificity and mechanisms of metabolic enzymes has been derived from these simulations[60,61].

MD simulations have revealed conformational changes induced by the binding of small molecules[62,63] to various CYP isoforms and established a relationship between substrate specificity and enzyme malleability[64]. They have been employed to study the solvation of the active sites of various CYP isoforms[65], and the active site access and egress pathways, which might play a key role in substrate selectivity and specificity[66,67]. Various methods have been developed to enhance conformational

sampling, including the application of an artificial force applied to the ligand in order to pull toward a specific (steered MD simulation[68]) or random (random accelerated MD simulation[69]) direction. The fact that the dynamic motions of CYPs are affected by their binding to membranes introduces another layer of complexity to their analysis[60].

Free energy (pathway) methods such as free energy perturbation[70], thermodynamic integration[71] and the Bennett acceptance ratio method[72] allow the calculation of the free energy of binding with a mean absolute error typically ~1.5 kcal/mol[73] (NB: the difference between a high and a low affinity analogue is only a few kcal/mol). These methods also come at high computational cost[24]. Faster free energy methods (such as LIE[74]) have been developed, but they are generally less accurate[73]. Free energy methods have been employed e.g. to study the binding of the acidic and lactone forms of atorvastatin to CYP3A4[75], the stereoselective metabolism of CYP substrates, and the impact of mutations on substrate affinities[76,77].

MD simulation methods come with considerable demands in computational power and human expertise. Therefore they have mostly been applied *a posteriori* for studying the time-dependent structural and electronic properties of the most important metabolic enzymes. GPU technology with faster MD algorithms is boosting capabilities to sample the phase space more comprehensively[78], but the setup of the simulation (protein preparation, ligand parameterization) in particular, and the respective analysis remain laborious and accessible to experts only. Data can be complex, "Big", and overwhelming, therefore there is a need to develop more efficient ways to decode the crucial bits of information. Various clustering techniques can be used to elucidate representative protein conformations that could be used for automated docking[79]. The many options to consider: water molecules, solvation, protonation and other factors, result in a very large number of possible receptor models. From a computational point of view this is generally not a limitation as molecules can be easily docked against thousands of protein structures within a short period of time. The problem lies again with the insufficient performance of scoring functions, leaving the question of which docking pose to trust.

Component 6: Reactivity

Reactivity is the major determinant of drug metabolism (at least for CYP-mediated biotransformations[80]), and quantum mechanical (QM) methods allow its study at an electronic level of detail. Regarding metabolism as a hypothetical two-step process, the molecule must fit into the active site, and then must be sufficiently activated by the enzyme to undergo metabolism. QM systems generally take into account only the most proximate protein environment that is directly involved in a chemical reaction but ignore effects originating from the more distant protein environment. But even with these abstract representations of enzyme systems, fairly accurate predictions for specific metabolic reactions are possible[81].

With MD simulations and QM methods having complementary properties, it comes as no surprise that the combination of both approaches, referred to as quantum mechanical/molecular mechanical (QM/MM) methods[82], has become a key technology for investigating enzyme reactions[61,83]. The idea of the QM/MM hybrid approach is to tackle large systems by describing the region where a chemical reaction takes place by a QM method while accounting for effects of the environment by MM methods. QM/MM methods have been extensively used for investigating the catalytic cycle of metabolic enzymes, primarily that of CYPs[61,83], but also of epoxide hydrolase[84] and glutathione S-transferases[85,86], for example. Reaction intermediates can be unstable with very short lifetimes, which makes them extremely challenging to observe with experimental approaches. Knowing their chemical structure is of immediate relevance to drug discovery, as it allows the rational design of molecules with specific binding properties (in particular substrates or inhibitors). This is an area of research where QM/MM methods can be effective.

A new direction to the QM calculation of full proteins is being explored utilising fragmentation methods like divide-and-conquer or fragment molecular orbital theory (FMO)[87], where the protein is split into terminally capped amino acid fragments to be calculated locally and an overall energy extracted from their pair interaction energies. Specifically designed GPU (graphics processing unit) based algorithms like TeraChem[88] allow for even QM molecular dynamics simulations of full proteins in an MM solvent environment.

Calculating molecular flexibility and/or reactivity is a non-trivial task, and it is important to note that such models depict one specific protein-ligand interactions or

enzyme mechanism only. They are also computationally fairly expensive, which is a further complication in the implementation of such methods in drug discovery workflows. Fast QSAR methods have therefore been preferred in an attempt to incorporate reactivity implicitly through chemical descriptors[89,90]. Also these are generally confined to a specific type of metabolic reaction. Machine learning methods, on the other hand, are able to encode a whole range of different metabolic reactions, which is one of the main reasons for the popularity of these methods, including support vector machines (SVMs), artificial neural networks (ANNs) and random forests. The types of chemical descriptors used for encoding SoMs, their numbers and level of sophistication are in fact remarkable. The spectrum ranges from FAME[91], a random forest model for phase I and II metabolism that uses only seven molecular descriptors, to RS-Predictor[92], an SVM model for CYP-mediated metabolism which employs more than 500 chemical descriptors.

Component 7: Metabolic networks - systems biology

Comprehensive models (simulators) of drug metabolism require the ability to correctly predict a whole cascade of events and properties of the system to allow the estimation of biological effects. This would involve accurate knowledge and prediction of the (i) concentrations and distribution of the drug, (ii) metabolic liabilities (SoMs), (iii) chemical structure of metabolites, (iv) interactions with pharmacologically and toxicologically relevant biomolecules, (v) reaction rates and (vi) tissue concentration and localisation of enzymes and cofactors. While prediction of likely metabolites is feasible, it is challenging to ascertain their pharmacological and toxicological relevance in the context of the biological system. Target prediction tools allow the identification of likely ligand-protein interactions and possibly extrapolation to the contribution of these interactions to prediction of phenotypic effects using QSAR techniques. However, pharmacological effects are strongly linked with reaction rates and metabolite concentrations in specific tissues, which themselves are to a significant extent influenced by many factors including the activity of transporters. Possibly, these will only be reliably estimated if binding, ligand recognition and unbinding processes (pharmacodynamics) are understood at an atomic level of detail. Metabolic rates are specific to an enzyme and a substrate. Regardless of the structural or chemical significance of changes to substrates, these

can have a huge impact on the binding and unbinding processes (which SoM and metabolite prediction algorithms typically do not include), and consequently, on metabolic rates[93]. Chemical modifications of one part of a molecule to increase its metabolic stability may in fact lead to *metabolic switching*, resulting in the expedited biotransformation at another position in the molecule. As a consequence, prediction of metabolic rates, if possible at all, is only feasible within an extremely narrow and well-defined chemical space[93]. Even if metabolic rates could be obtained, it is still a far cry from being able to estimate effects of metabolites on a biological system, in particular as metabolic processes and biological responses can be highly specific to the individual[13].

Integrated computational approaches

Integrated computational approaches combine a variety of data sources, models, and algorithms in order to boost applicability, information content, and significance and prediction success rates, with the ultimate goal of rendering a (more) complete picture of physiological processes. One common strategy is to combine ligand- and structure-based methods. It has been shown that both have their individual methodological advantages and disadvantages, and there is no clear preference for a single method[43,94] but a lot to gain by their combination. The targeted use of consensus approaches and composite modelling workflows represent further established strategies for improving the accuracy and significance of predictions, e.g. for the task of bioactivity profiling[95,96]. Pathway information and network analysis algorithms have been successfully integrated for predicting the pharmacokinetic and pharmacodynamic properties of drug molecules[46,47]. Integration of data types/resources and algorithms has also become a major driver of tools for regioselectivity and metabolite prediction. For example, knowledge-based systems use physiochemical property estimators and QC methods to flag potentially toxic metabolites and reduce false-positives rates. The latter has also been addressed by implementing docking as a post-filtering tool for generated metabolites[97]. MetaSite uses MIFs derived from the protein structure structural data, combined with a fingerprint-based algorithm (fast docking surrogate method) and a QC approach (reactivity estimator). The importance for including reactivity, steric accessibility and pharmacophoric constraints has been studied in detail for SMARTCyp[80,98,99], but in

fact most statistical and machine-learning approaches make use of descriptors that encode these components.

BEST OF BOTH WORLDS: INTEGRATION OF EXPERIMENT AND COMPUTATION

A broad portfolio of experimental and theoretical methods for studying metabolism at different levels of sophistication is available today. It is important to realise that when used independently, any of these approaches illuminate only specific aspects of drug metabolism and ignores most of the others. In contrast, the combination of various experimental and theoretical approaches will generate a fairly complete picture of a compound's metabolic fate. One would assume that synergies created by the integration of both domains has long been harnessed by the pharmaceutical industry, when in fact it seems that there is still some way to go to reach this point. The reasons for this situation are manifold: limited resources, costs of process restructuring, organizational structures, timelines for learning cycles, diverging (and in part incompatible) schools of thought, and the limitations of current methods. A key problem is the open availability of relevant and diverse data, which is often secured in pharmaceutical companies and CROs. Sharing these data is often impossible due to intellectual property constraints. However, solving this problem will provide data related to a broader chemical space and thereby increase the applicability domain of computational prediction models.

Departments for computational and experimental research of drug metabolism are traditionally operated as well-separated "silos". Scientists of both areas have different backgrounds, use distinct terminologies, and are focused on applying their acquired expertise. Depending on the organizational setup of research organisations or projects, the level of integration of drug metabolism research varies greatly. Computational tools may be used by experts in drug metabolism and/or by computational or medicinal chemists. The latter tend to focus on SAR and synthesis. Computational chemists may not have expert knowledge in pharmacokinetics, and sophisticated theoretical methods may not necessarily be within the comfort zone of specialists in drug metabolism. Close interplay of the different disciplines is therefore of utmost importance to tackle the problem of metabolism prediction. Efficient communication allows the development of appropriate models based on in-house data

to guide metabolic optimization. Early access to data on microsomal or hepatocyte stability is required. A recent example for the successful liaison of experimental and theoretical approaches is the optimisation of the metabolic stability of mineralocorticoid receptor antagonists reported in Box 2. But there is room for improvement, in particular with respect to modelling CYP induction, the rate and extent of metabolic reactions, and the prediction of phase II metabolism.

<< Box 2 approx. here>>

CONCLUSIONS AND FUTURE DIRECTIONS

A plethora of powerful *in vitro* assay and analytical technologies are at our disposal, allowing us in principle to obtain a fairly complete and accurate picture of metabolic processes. Whilst it is likely that robust and cost-effective metabolite generating systems such as hepatic microsomes will continue to provide a solid bedrock for metabolite generation the development of so-called 3D bioreactor ("organ on a chip") technology is a rapidly developing field. These systems provide a much more "organ-like environment" for cells, which can be maintained for long periods whilst being exposed to drugs in order to assess metabolism and toxicity[100-102]. If, as seems likely, such bioreactors (which are not limited to hepatocytes) can be made into a robust and easily deployed technology this will represent a major advance for *in vitro* techniques.

Analytical techniques for drug metabolite detection and identification are likely to remain focussed on the use of MS-based technologies of ever increasing levels of sophistication and sensitivity for the foreseeable future. The development of robust miniaturized separation systems will further drive down the sample requirements[103]. Another promising development is the application of ion mobility spectrometry (IMS), which can enhance metabolite separation and, by additional characterization capability, improve SoMs identification. The approach is based on separating isomeric species depending on differences in their IMS drift times, which are linearly proportional to the collision cross-section reflecting physical size and shape[104,105]. Fundamental progress in SoM and metabolite prediction would result

from improvements in the reporting of additional metadata on the origin (tissues, genomes, experimental conditions etc.) of samples, complete identification of major and minor products, improved experimental and statistical software tools for the normalisation of mass spectral data, integration with transport phenomena and experimental data on the concentrations and rates of reaction.

Rapid advances can also be expected in the development of new animal models, either based on the engraftment of human cells into animals such as those seen in the "chimeric humanized" mice or animals genetically modified to contain human drug-metabolising enzymes and transporters. These models can be expected to provide much more accurate predictions of the metabolic and pharmacokinetic properties of candidate drugs in humans[10,11], but certainly raise ethical issues.

Experimental approaches to investigate metabolism come with substantial demands in technical resources and human expertise. This is where computational chemistry can introduce a big leap forward in metabolism research. Computational prediction of drug metabolism is still a young field of research but *in silico* methods do not only have the potential to guide or (partially) replace experimental efforts, they may also be the method of choice to investigate catalytic processes and highly reactive species. The most obvious synergy to gain from the combination of both domains is from coupling of software and spectroscopic instruments for more efficient metabolite identification and characterization. Manufacturers of mass spectrometers are vigorously pursuing this.

Today drug metabolism is typically addressed in the industrial setting rather than academia, with the consequence that software and data are generally withheld and not available to the scientific community at large. However, the tide seems to be turning and the field of cheminformatics appears to be finally making much more program code and data available to the scientific community. Data on the sites and products of metabolism, metabolite concentrations and tissue locations of metabolising enzymes are being harvested and incorporated into various databases which address different aspects of metabolism, including drug and agrochemical metabolism in a variety of species of plants and animals. The information stored ranges from data that is semantically rich to numerical data on concentrations and locations of metabolites. Annotated metabolic pathways and data on transporters complement this picture.

Metabolism is an example of an emergent process, and the complexity of the system is influenced by many extraneous and individual factors. Current computational approaches take little account of these and other factors including transporters, proper consideration of metabolising enzymes other than CYPs, environmental factors, and in particular the concentrations, metabolic networks and rates of metabolism of metabolising enzymes. Substantial progress is being made in gathering data for pharmacokinetic/pharmacodynamic (PK/PD) modelling, and this will have a significant effect on prediction not just which metabolites are present, but also their concentrations and flux over time. In addition, the emerging field of metabonomics attempts to connect the metabolic state of an individual or population to phenotypic measures. This covers the whole gamut of metabolic research, instrument design and measurement protocols to machine learning and clinical prediction. There is enormous potential in this area to combine all the experimental and theoretical methods to improve patient stratification and clinical outcomes.

We will continue to see substantial gains in computational power, driven by GPU and cloud technology, increasingly efficient algorithms, and advances in parallel computing. This will open up many new avenues for the development and application of highly accurate methods for the simulation of biomolecules and indeed whole biosystems.

For experimental and theoretical approaches alike, human expertise remains an important ingredient. The broad range of disciplines involved at the cutting edge of predictive drug metabolism is impressive. It draws on chemistry (physical, organic-synthetic, analytical, medicinal etc.), biology (biochemistry, enzymology, genetic, epigenetics etc.), pharmacology (molecular, clinical, pharmacokinetic, toxicology, therapeutics etc.), and computational components (software development, quantum chemistry, simulations, statistics, machine learning etc.). For all of these experimental and theoretical domains of research, there is no technology available to date that is superior in all relevant scenarios. Bringing experimentalists and theoretical scientists closer together will result in progress. Ideally, a feedback loop of experiment, design and testing would greatly improve predictive models, as much of the present work has concentrated on retrospective analysis. In the future, closer collaboration is needed to further develop these models. The central role that metabolism plays in biology and the requirement that predictive models be incorporated into a wide variety of research

and development programs (e.g. drug discovery, environmental science, clinical phenotyping, pesticide development) means that significant resources will be applied and progress in understanding and in generating predictive models will continue.

REFERENCES

1 Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336**, 1262-1267 (2012).

2 Wienkers, L. C. & Heath, T. G. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat. Rev. Drug Discov.* **4**, 825-833 (2005).

3 Di, L. The role of drug metabolizing enzymes in clearance. *Expert Opin. Drug Metab. Toxicol.* **10**, 379-393 (2014).

4 Testa, B. in *Drug Metabolism Prediction* (ed J. Kirchmair) 3-25 (Wiley-VCH, Weinheim, 2014).

5 Testa, B. Drug metabolism for the perplexed medicinal chemist. *Chem. Biodivers.* **6**, 2055-2070 (2009).

6 Kirchmair, J. *et al.* How do metabolites differ from their parent molecules and how are they excreted? *J. Chem. Inf. Model.* **53**, 354-367 (2013).

7 Testa, B., Pedretti, A. & Vistoli, G. Reactions and enzymes in the metabolism of drugs and other xenobiotics. *Drug Discov. Today* **17**, 549-560 (2012).

8 van Waterschoot, R. A. B. & Schinkel, A. H. A critical analysis of the interplay between cytochrome P450 3A and P-glycoprotein: Recent insights from knockout and transgenic mice. *Pharmacol. Rev.* **63**, 390-410 (2011).

9 Thomas, S. & Dimelow, R. J. in *Drug Metabolism Prediction* (ed J. Kirchmair) 373-396 (Wiley-VCH, Weinheim, 2014).

10 Scheer, N. & Wolf, C. R. Genetically humanized mouse models of drug metabolizing enzymes and transporters and their applications. *Xenobiotica* **44**, 96-108 (2014).

11 Kitamura, S. & Sugihara, K. Current status of prediction of drug disposition and toxicity in humans using chimeric mice with humanized liver. *Xenobiotica* **44**, 123-134 (2014).

12 Glube, N. D. & Duchateau, G. in *Drug Metabolism Prediction* (ed J. Kirchmair) 415-440 (Wiley-VCH, Weinheim, 2014).

13 Ferguson, S. & Bonzo, J. A. in *Drug Metabolism Prediction* (ed J. Kirchmair) 441-484 (Wiley-VCH, Weinheim, 2014).

14 Wilson, I. D. in *Drug Metabolism Prediction* (ed J. Kirchmair) 485-498 (Wiley-VCH, Weinheim, 2014).

15 Zhu, M. S., Zhang, H. Y. & Humphreys, W. G. Drug metabolite profiling and identification by high-resolution mass spectrometry. *J. Biol. Chem.* **286**, 25419-25425 (2011).

16 Anari, M. R., Sanchez, R. I., Bakhtiar, R., Franklin, R. B. & Baillie, T. A. Integration of knowledge-based metabolic predictions with liquid chromatography data-dependent tandem mass spectrometry for drug metabolism studies: Application to studies on the biotransformation of indinavir. *Anal. Chem.* **76**, 823-832 (2004).

17 Leclercq, L. *et al.* IsoScore: Automated localization of biotransformations by mass spectrometry using product ion scoring of virtual regioisomers. *Rapid Commun. Mass Spectrom.* **23**, 39-50 (2009).

18 Zamora, I., Fontaine, F., Serra, B. & Plasencia, G. High-throughput, computer assisted, specific MetID. A revolution for drug discovery. *Drug Discov. Today Technol.* **10**, e199-e205 (2013).

19 Bonn, B., Leandersson, C., Fontaine, F. & Zamora, I. Enhanced metabolite identification with MSE and a semi-automated software for structural elucidation. *Rapid Commun. Mass Spectrom.* **24**, 3127-3138 (2010).

20 Wishart, D. S. Advances in metabolite identification. *Bioanalysis* **3**, 1769-1782 (2011).

21 Marchant, C. A., Briggs, K. A. & Long, A. In silico tools for sharing data and knowledge on toxicity and metabolism: Derek for Windows, Meteor, and Vitic. *Toxicol. Mech. Method.* **18**, 177-187 (2008).

22 Ridder, L. & Wagener, M. SyGMa: Combining expert knowledge and empirical scoring in the prediction of metabolites. *ChemMedChem* **3**, 821-832 (2008).

23 Jacobs, P. L. *et al.* Identification of drug metabolites in human plasma or serum integrating metabolite prediction, LC-HRMS and untargeted data processing. *Bioanalysis* **5**, 2115-2128 (2013).

24 Beisken, S., Earll, M., Portwood, D., Seymour, M. & Steinbeck, C. MassCascade: Visual programming for LC-MS data processing in metabolomics. *Mol. Inf.* **33**, 307-310 (2014).

25 Nicholson, J. K. & Lindon, J. C. Systems biology: Metabonomics. *Nature* **455**, 1054-1056 (2008).

26 Giri, S., Krausz, K. W., Idle, J. R. & Gonzalez, F. J. The metabolomics of (+/-)-arecoline 1-oxide in the mouse and its formation by human flavin-containing monooxygenases. *Biochem. Pharmacol.* **73**, 561-573 (2007).

27 Athersuch, T. J., Nicholson, J. K. & Wilson, I. D. Isotopic enrichment enhancement in metabonomic analysis of UPLC-MS data sets. *J. Labelled Compd. Rad.* **50**, 303-307 (2007).

28 Wilson, I. D. in *NMR Spectroscopy in Pharmaceutical Analysis* (eds U. Holzgrabe, I. Wawer, & B. Diehl) Ch. Chapter 7 - Hypernation and concatenation: Multiple on-line spectroscopic analysis for drug and natural product characterisation, 449-469 (Elsevier, Oxford, 2008).

29 Chovan, L. E., Black-Schaefer, C., Dandliker, P. J. & Lau, Y. Y. Automatic mass spectrometry method development for drug discovery: Application in metabolic stability assays. *Rapid Commun. Mass Spectrom.* **18**, 3105-3112 (2004).

30 Jenkins, K. M. *et al.* Automated high throughput ADME assays for metabolic stability and cytochrome P450 inhibition profiling of combinatorial libraries. *J. Pharmaceut. Biomed.* **34**, 989-1004 (2004).

31 Youdim, K. A. & Saunders, K. C. A review of LC-MS techniques and high-throughput approaches used to investigate drug metabolism by cytochrome P450s. *J. Chromatograph. B* **878**, 1326-1336 (2010).

32 Ma, S. & Zhu, M. Recent advances in applications of liquid chromatography–tandem mass spectrometry to the analysis of reactive drug metabolites. *Chem.-Biol. Interact.* **179**, 25-37 (2009).

33 Yan, Z., Caldwell, G. W. & Maher, N. Unbiased high-throughput screening of reactive metabolites on the linear ion trap mass spectrometer using polarity switch and mass tag triggered data-dependent acquisition. *Anal. Chem.* **80**, 6410-6422 (2008).

34 Tan, L. & Kirchmair, J. in *Drug Metabolism Prediction* (ed J. Kirchmair) 29-51 (Wiley-VCH, Weinheim, 2014).

35 Wishart, D. S. in *Drug Metabolism Prediction* (ed J. Kirchmair) 53-74 (Wiley-VCH, Weinheim, 2014).

36 Ekins, S., Mestres, J. & Testa, B. In silico pharmacology for drug discovery: Methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* **152**, 9-20 (2007).

37 Phuangsawai, O., Hannongbua, S. & Gleeson, M. P. in *Drug Metabolism Prediction* (ed J. Kirchmair) 321-349 (Wiley-VCH, Weinheim, 2014).

38 Koutsoukas, A. *et al.* From in silico target prediction to multi-target drug design: Current databases, methods and applications. *J. Proteomics* **74**, 2554-2574 (2011).

39 Gleeson, M. P. *et al.* The challenges involved in modeling toxicity data in silico: A review. *Curr. Pharm. Des.* **18**, 1266-1291 (2012).

40 Tiikkainen, P., Bellis, L., Light, Y. & Franke, L. Estimating error rates in bioactivity databases. *J. Chem. Inf. Model.* **53**, 2499-2505 (2013).

41 Testa, B. Organic stereochemistry. Part 7: The concept of substrate stereoselectivity in biochemistry and xenobiotic metabolism. *Helv. Chim. Acta* **96**, 1203-1234 (2013).

42 Testa, B. Organic stereochemistry. Part 8: Prostereoisomerism and the concept of product stereoselectivity in biochemistry and xenobiotic metabolism. *Helv. Chim. Acta* **96**, 1409-1451 (2013).

43 Kirchmair, J. *et al.* Computational prediction of metabolism: Sites, products, SAR, P450 enzyme dynamics, and mechanisms. *J. Chem. Inf. Model.* **52**, 617-648 (2012).

44 Kaserer, T., Temml, V. & Schuster, D. in *Drug Metabolism Prediction* (ed J. Kirchmair) 351-372 (Wiley-VCH, Weinheim, 2014).

45 Bender, A. in *Drug Metabolism Prediction* (ed J. Kirchmair) 397-411 (Wiley-VCH, Weinheim, 2014).

46 Ekins, S. *et al.* A combined approach to drug metabolism and toxicity assessment. *Drug Metab. Dispos.* **34**, 495-503 (2006).

47 Ekins, S. *et al.* Algorithms for network analysis in systems-ADME/tox using the metacore and MetaDrug platforms. *Xenobiotica* **36**, 877-901 (2006).

48 Adams, S. E. *Molecular similarity and xenobiotic metabolism* Ph.D. thesis, University of Cambridge, (2010).

49 Judson, P. N. in *Drug Metabolism Prediction* (ed J. Kirchmair) 293-318 (Wiley-VCH, Weinheim, 2014).

50 Stepan, A. F. *et al.* Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: A perspective based on the critical examination of trends in the top 200 drugs marketed in the United States. *Chem. Res. Toxicol.* **24**, 1345-1410 (2011).

51 U. S. Food and Drug Administration. Guidance for industry: Genotoxic and carcinogenic impurities in drug substances and products: Recommended approaches. (2008).

52 Williams, P. A. *et al.* Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **424**, 464-468 (2003).

53 Leach, A. G. & Kidley, N. J. in *Drug Metabolism Prediction* (ed J. Kirchmair) 103-132 (Wiley-VCH, Weinheim, 2014).

54 Oostenbrink, C. in *Drug Metabolism Prediction* (ed J. Kirchmair) 243-263 (Wiley-VCH, Weinheim, 2014).

55 Sykes, M. J., McKinnon, R. A. & Miners, J. O. Prediction of metabolism by cytochrome P450 2C9: Alignment and docking studies of a validated database of substrates. *J. Med. Chem.* **51**, 780-791 (2008).

56 Warren, G. L. *et al.* A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **49**, 5912-5931 (2006).

57 Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849-857 (1985).

58 Broccatelli, F. & Brown, N. in *Drug Metabolism Prediction* (ed J. Kirchmair) 223-242 (Wiley-VCH, Weinheim, 2014).

59 Cruciani, G. *et al.* MetaSite: Understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.* **48**, 6970-6979 (2005).

60 Mustafa, G., Yu, X. & Wade, R. C. in *Drug Metabolism Prediction* (ed J. Kirchmair) 77-101 (Wiley-VCH, Weinheim, 2014).

61 Shaik, S., Chen, H., Usharani, D. & Thiel, W. in *Drug Metabolism Prediction* (ed J. Kirchmair) 133-177 (Wiley-VCH, Weinheim, 2014).

62 Berka, K. *et al.* Binding of quinidine radically increases the stability and decreases the flexibility of the cytochrome P450 2D6 active site. *J. Inorg. Biochem.* **110**, 46-50 (2012).

63 Porubsky, P. R., Battaile, K. P. & Scott, E. E. Human cytochrome P450 2E1 structures with fatty acid analogs reveal a previously unobserved binding mode. *J. Biol. Chem.* **285**, 22282-22290 (2010).

64 Skopalík, J., Anzenbacher, P. & Otyepka, M. Flexibility of human cytochromes P450: Molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. *J. Chem. Phys. B* **112**, 8165-8173 (2008).

65 Hendrychova, T., Berka, K., Navratilova, V., Anzenbacher, P. & Otyepka, M. Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. *Curr. Drug Metab.* **13**, 177-189 (2012).

66 Cojocaru, V., Winn, P. J. & Wade, R. C. Multiple, ligand-dependent routes from the active site of cytochrome P450 2C9. *Curr. Drug Metab.* **13**, 143-154 (2012).

67 Cojocaru, V., Winn, P. J. & Wade, R. C. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta* **1770**, 390-401 (2007).

68 Grubmuller, H., Heymann, B. & Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **271**, 997-999 (1996).

69 Ludemann, S. K., Lounnas, V. & Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.* **303**, 797-811 (2000).

70 Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **22**, 1420 (1954).

71 Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**, 300-313 (1935).

72 Bennett, C. H. Efficient estimation of free-energy differences from Monte-Carlo data. *J. Comput. Phys.* **22**, 245-268 (1976).

73 Williamson, M. J. in *Drug Metabolism Prediction* (ed J. Kirchmair) 179-197 (Wiley-VCH, Weinheim, 2014).

74 Aqvist, J., Medina, C. & Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7**, 385-391 (1994).

75 Jacobsen, W. *et al.* Lactonization is the critical first step in the disposition of the 3-hydroxy-3-methylglutaryl-CoA reductase inhibitor atorvastatin. *Drug Metab. Dispos.* **28**, 1369-1378 (2000).

76 de Beer, S. B. A., Venkataraman, H., Geerke, D. P., Oostenbrink, C. & Vermeulen, N. P. E. Free energy calculations give insight into the stereoselective hydroxylation of α-ionones by engineered cytochrome P450 BM3 mutants. *J. Chem. Inf. Model.* **52**, 2139-2148 (2012).

77 Nagy, G. & Oostenbrink, C. Rationalization of stereospecific binding of propranolol to cytochrome P450 2D6 by free energy calculations. *Eur. Biophys. J.* **41**, 1065-1076 (2012).

78 Götz, A. W. *et al.* Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theor. Comput.* **8**, 1542-1555 (2012).

79 Amaro, R. E. & Li, W. W. Emerging methods for ensemble-based virtual screening. *Curr. Top. Med. Chem.* **10**, 3-13 (2010).

80 Rydberg, P., Rostkowski, M., Gloriam, D. E. & Olsen, L. The contribution of atom accessibility to site of metabolism models for cytochromes P450. *Mol. Pharm.* **10**, 1216-1223 (2013).

81 Korzekwa, K. R., Jones, J. P. & Gillette, J. R. Theoretical studies on cytochrome P-450 mediated hydroxylation: A predictive model for hydrogen atom abstractions. *J. Am. Chem. Soc.* **112**, 7042-7046 (1990).

82 Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227-249 (1976).

83 Shaik, S. *et al.* P450 enzymes: Their structure, reactivity, and selectivity—modeled by QM/MM calculations. *Chem. Rev.* **110**, 949-1017 (2010).

84 Lonsdale, R., Hoyle, S., Grey, D. T., Ridder, L. & Mulholland, A. J. Determinants of reactivity and selectivity in soluble epoxide hydrolase from quantum mechanics/molecular mechanics modeling. *Biochemistry* **51**, 1774-1786 (2012).

85 Ridder, L., Rietjens, I. M. C. M., Vervoort, J. & Mulholland, A. J. Quantum mechanical/molecular mechanical free energy simulations of the glutathione S-transferase (M1-1) reaction with phenanthrene 9,10-oxide. *J. Am. Chem. Soc.* **124**, 9926-9936 (2002).

86 Bowman, A. L., Ridder, L., Rietjens, I. M. C. M., Vervoort, J. & Mulholland, A. J. Molecular determinants of xenobiotic metabolism: QM/MM simulation of the conversion of 1-chloro-2,4-dinitrobenzene catalyzed by M1-1 glutathione S-transferase. *Biochemistry* **46**, 6353-6363 (2007).

87 Gordon, M. S., Fedorov, D. G., Pruitt, S. R. & Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.* **112**, 632-672 (2012).

88 Ufimtsev, I. S. & Martinez, T. J. Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theor. Comput.* **5**, 2619-2628 (2009).

89 Hennemann, M. *et al.* CypScore: Quantitative prediction of reactivity toward cytochromes P450 based on semiempirical molecular orbital theory. *ChemMedChem* **4**, 657-669 (2009).

90 Singh, S. B., Shen, L. Q., Walker, M. J. & Sheridan, R. P. A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. *J. Med. Chem.* **46**, 1330-1336 (2003).

91 Kirchmair, J. *et al.* FAst MEtabolizer (FAME): A rapid and accurate Predictor of sites of metabolism in multiple species by endogenous enzymes. *J. Chem. Inf. Model.* **53**, 2896-2907 (2013).

92 Zaretzki, J. *et al.* RS-Predictor: A new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP 3A4. *J. Chem. Inf. Model.* **51**, 1667-1689 (2011).

93 Guengerich, F. P. in *Drug Metabolism Prediction* (ed J. Kirchmair) 199-219 (Wiley-VCH, Weinheim, 2014).

94 Rydberg, P. in *Drug Metabolism Prediction* (ed J. Kirchmair) 265-291 (Wiley-VCH, Weinheim, 2014).

95 Reker, D., Rodrigues, T., Schneider, P. & Schneider, G. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc. Natl. Acad. Sci. U S A.* **111**, 4067-4072 (2014).

96 Meslamani, J., Bhajun, R., Martz, F. & Rognan, D. Computational profiling of bioactive compounds using a target-dependent composite workflow. *J. Chem. Inf. Model.* **53**, 2322-2333 (2013).

97 Tarcsay, Á., Kiss, R. & Keserű, G. M. Site of metabolism prediction on cytochrome P450 2C9: A knowledge-based docking approach. *J. Comput.-Aided Mol. Des.* **24**, 399-408 (2010).

98 Rydberg, P. & Olsen, L. Ligand-based site of metabolism prediction for cytochrome P450 2D6. *ACS Med. Chem. Lett.* **3**, 69-73 (2012).

99 Rydberg, P. & Olsen, L. Predicting drug metabolism by cytochrome P450 2C9: Comparison with the 2D6 and 3A4 isoforms. *ChemMedChem* **7**, 1202-1209 (2012).

100 Sivaraman, A. *et al.* A microscale in vitro physiological model of the liver: Predictive screens for drug metabolism and enzyme induction. *Curr. Drug Metab.* **6**, 569-591 (2005).

101 van Midwoud, P. M., Verpoorte, E. & Groothuis, G. M. M. Microfluidic devices for in vitro studies on liver drug metabolism and toxicity. *Integr. Biol.* **3**, 509 (2011).

102 Khetani, S. R. & Bhatia, S. N. Microscale culture of human liver cells for drug development. *Nat. Biotechnol.* **26**, 120-126 (2007).

103 Rainville, P. D., Smith, N. W., Wilson, I. D., Nicholson, J. K. & Plumb, R. S. Addressing the challenge of limited sample volumes in in vitro studies with capillary-scale microfluidic LC–MS/MS. *Bioanalysis* **3**, 873-882 (2011).

104 Dear, G. J. *et al.* Sites of metabolic substitution: Investigating metabolite structures utilising ion mobility and molecular modelling. *Rapid Commun. Mass Spectrom.* **24**, 3157-3162 (2010).

105 Shimizu, A. & Chiba, M. Ion mobility spectrometry-mass spectrometry analysis for the site of aromatic hydroxylation. *Drug Metab. Dispos.* **41**, 1295-1299 (2013).

106 Bodor, N. & Buchwald, P. Soft drug design: General principles and recent applications. *Med. Res. Rev.* **20**, 58-101 (2000).

107 Rautio, J. *et al.* Prodrugs: Design and clinical applications. *Nat. Rev. Drug Discov.* **7**, 255-270 (2008).

108 Smith, D. A. Do prodrugs deliver? *Curr. Opin. Drug Disc.* **10**, 550-559 (2007).

109 Testa, B. Prodrugs: Bridging pharmacodynamic/pharmacokinetic gaps. *Curr. Opin. Chem. Biol.* **13**, 338-344 (2009).

110 Huttunen, K. M., Raunio, H. & Rautio, J. Prodrugs - from serendipity to rational design. *Pharmacol. Rev.* **63**, 750-771 (2011).

111 Ettmayer, P., Amidon, G. L., Clement, B. & Testa, B. Lessons learned from marketed and investigational prodrugs. *J. Med. Chem.* **47**, 2393-2404 (2004).

112 Fura, A. *et al.* Discovering drugs through biological transformation: Role of pharmacologically active metabolites in drug discovery. *J. Med. Chem.* **47**, 4339-4351 (2004).

113 Obach, R. S. Pharmacologically active drug metabolites: Impact on drug discovery and pharmacotherapy. *Pharmacol. Rev.* **65**, 578-640 (2013).

114 Testa, B. & Krämer, S. D. The biochemistry of drug metabolism - an introduction. Part 1. Metabolism and bioactivity. *Chem. Biodivers.* **3**, 1053-1101 (2006).

115 Vandenberg, J. I. *et al.* hERG K+ channels: Structure, function, and clinical significance. *Physiol. Rev.* **92**, 1393-1478 (2012).

116 Guengerich, F. P. & MacDonald, J. S. Applying mechanisms of chemical toxicity to predict drug safety. *Chem. Res. Toxicol.* **20**, 344-369 (2007).

117 Tang, W. & Lu, A. Y. H. Metabolic bioactivation and drug-related adverse effects: Current status and future directions from a pharmaceutical research perspective. *Drug Metab. Rev.* **42**, 225-249 (2010).

118 Esser, C. in *Comprehensive Medicinal Chemistry* (eds B. Testa & H. van de Waterbeemd) 215-229 (Elsevier, Oxford, 2007).

119 Sullivan, E. A. & Shulman, K. I. Diet and monoamine oxidase inhibitors: A re-examination. *Can. J. Psychiatry* **29**, 707-711 (1984).

120 Quinn, D. K. & Stern, T. A. Linezolid and serotonin syndrome. *Prim. Care Companion J. Clin. Psychiatry* **11**, 353-356 (2009).

121 Gillman, P. K. Monoamine oxidase inhibitors, opioid analgesics and serotonin toxicity. *Br. J. Anaesth.* **95**, 434-441 (2005).

122 Kell, D. B., Dobson, P. D., Bilsland, E. & Oliver, S. G. The promiscuous binding of pharmaceutical drugs and their transporter-mediated uptake into cells: What we (need to) know and how we can do so. *Drug Discov. Today* **18**, 218-239 (2013).

123 Thede, K. *et al.* Preparation of 3-hydroxyalkylindoles as mineralocorticoid receptor inhibitors. WO2009156091A1 (2009).

124 Weisel, M., Proschak, E. & Schneider, G. PocketPicker: Analysis of ligand binding-sites with shape descriptors. *Chem. Cent. J.* **1**, 7 (2007).

125 MEXAlert (CompuDrug International, Inc., Bal Harbor, FL).

126 QikProp (Schrödinger, New York, NY).

127 Li, J., Schneebeli, S. T., Bylund, J., Farid, R. & Friesner, R. A. IDSite: An accurate approach to predict P450-mediated drug metabolism. *J. Chem. Theory Comput.* **7**, 3829-3845 (2011).

128 Rydberg, P., Gloriam, D. E., Zaretzki, J., Breneman, C. & Olsen, L. SMARTCyp: A 2D method for prediction of cytochrome P450-mediated drug metabolism. *ACS Med. Chem. Lett.* **1**, 96-100 (2010).

129 Rydberg, P., Gloriam, D. & Olsen, L. The SMARTCyp cytochrome P450 metabolism prediction server. *Bioinformatics (Oxford, England)* **26**, 2988-2989 (2010).

130 StarDrop (Optibrium, Ltd., Cambridge, UK).

131 ADMET Predictor (Simulations Plus, Inc., Lancaster, CA).

132 Percepta Platform (ACD/Labs, Toronto, ON).

133 Zaretzki, J. *et al.* RS-WebPredictor: A server for predicting CYP-mediated sites of metabolism on drug-like molecules. *Bioinformatics* **29**, 497-498 (2013).

134 Darvas, F. MetabolExpert: An expert system for predicting metabolism of substances. *QSAR Environ. Toxicol.*, 71-81 (1987).

135 Klopman, G., Dimayuga, M. & Talafous, J. META 1. A program for the evaluation of metabolic transformation of chemicals. *J. Chem. Inf. Model.* **34**, 1320-1325 (1994).

136 MetaDrug (Thomson Reuters, New York, NY).

137 Mekenyan, O. G., Dimitrov, S. D., Pavlov, T. S. & Veith, G. D. A systematic approach to simulating metabolism in computational toxicology. I. The TIMES heuristic modelling framework. *Curr. Pharm. Des.* **10**, 1273-1293 (2004).

138 Gao, J., Ellis, L. B. & Wackett, L. P. The University of Minnesota Pathway Prediction System: Multi-level prediction and visualization. *Nucleic Acids Res.* **39**, W406-411 (2011).

139 JChem Metabolizer (ChemAxon, Kft., Budapest, Hungary).

140 Terfloth, L., Bienfait, B. & Gasteiger, J. Ligand-based models for the isoform specificity of cytochrome P450 3A4, 2D6, and 2C9 substrates. *J Chem Inf Model* **47**, 1688-1701 (2007).

141 Mishra, N. K., Agarwal, S. & Raghava, G. P. Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule. *BMC Pharmacol.* **10**, 8 (2010).

142 Rostkowski, M., Spjuth, O. & Rydberg, P. WhichCyp: Prediction of cytochromes P450 inhibition. *Bioinformatics* **29**, 2051-2052 (2013).

143 Faulon, J. L., Visco, D. P. & Pophale, R. S. The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSPR studies. *J. Chem. Inf. Model.* **43**, 707-720 (2003).

144 Vedani, A., Dobler, M. & Smieško, M. VirtualToxLab — a platform for estimating the toxic potential of drugs, chemicals and natural products. *Toxicol. Appl. Pharmacol.* **261**, 142-153 (2012).

145 ADMEWORKS Predictor (Fujitsu Kyushu Systems, Ltd., Fukuoka, Japan).

146 Smithing, M. P. & Darvas, F. HazardExpert. An expert system for predicting chemical toxicity. *ACS Symp. Ser.* **484**, 191-200 (1992).

147 Leadscope toxicity models (Leadscope, Columbus, OH, 2013).

148 (M)CASE ultra & modules (Beachwood, OH, Multicase, 2013).

149 Patlewicz, G., Jeliazkova, N., Safford, R. J., Worth, A. P. & Aleksiev, B. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ. Res.* **19**, 495-524 (2008).

150 TOPKAT (Accelrys, San Diego, CA, 2013).

151 SYMMETRY (Prous Institute for Biomedical Research, Barcelona, Spain, 2013).

152 Campagna-Slater, V., Pottel, J., Therrien, E., Cantin, L. D. & Moitessier, N. Development of a computational tool to rival experts in the prediction of sites of metabolism of xenobiotics by p450s. *J. Chem. Inf. Model.* **52**, 2471-2483 (2012).

153 ADME database (Fujutsu Kyushu Systems, Fukuoka, Japan, 2014).

154 Liu, T., Lin, Y., Wen, X., Jorissen, R. N. & Gilson, M. K. Bindingdb: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **35**, D198-D201 (2007).
155 Gaulton, A. *et al.* ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100-D1107 (2011).
156 Gostar drug database (GVK Biosciences Private Limited, Hyderabad, India, 2013).
157 Wishart, D. S. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668-D672 (2006).
158 Gao, J. F., Ellis, L. B. M. & Wackett, L. P. The University of Minnesota Biocatalysis/Biodegradation Database: Improving public access. *Nucleic Acids Res.* **38**, D488-D491 (2010).
159 Wishart, D. S. *et al.* HMDB: The human metabolome database. *Nucleic Acids Res.* **35**, D521-D526 (2007).
160 Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109-D114 (2012).
161 Metabase (Thomson Reuters, 2013).
162 Hachad, H., Ragueneau-Majlessi, I. & Levy, R. H. A useful tool for drug interaction evaluation: The University of Washington Metabolism and Transport Drug Interaction Database. *Human Genomics* **5**, 61 (2010).
163 Metabolite (Accelrys, San Diego, CA, 2013).
164 Smith, C. A. *et al.* Metlin: A metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747-751 (2005).
165 Hewett, M. *et al.* PharmGKB: The Pharmacogenetics Knowledge Base. *Nucleic Acids Res.* **30**, 163-165 (2002).
166 Cao, D. *et al.* ADMET evaluation in drug discovery. 11. PharmacoKinetics Knowledge Base (PKKB): A comprehensive database of pharmacokinetic and toxic properties for drugs. *J. Chem. Inf. Model.* **52**, 1132-1137 (2012).
167 Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242 (2000).
168 *The PubChem Project*, http://pubchem.ncbi.nlm.nih.gov/ (accessed 2015-01-07).
169 Preissner, S. *et al.* SuperCYP: A comprehensive database on cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. *Nucleic Acids Res.* **38**, D237-243 (2010).
170 *Wikipedia*, http://www.wikipedia.org (accessed 2015-01-07).

ELEMENTS

Box 1. The many implications of drug metabolism

Categories of drugs with respect to their metabolic behaviour

Soft drugs have (or, are designed for) fast and predictable metabolic transformation into inactive, nontoxic metabolites (often via conjugation or hydrolysis reactions)[106]. Examples include succinylcholine and esmolol. Metabolically stable (or highly resistant) drugs, on the other hand, are excreted without undergoing metabolism. Examples include lisinopril and bisphosphonates.

Drugs with active or highly active metabolite(s) include cisplatin (→ monoaqua and diaqua species), diazepam (→ nordiazepam), encainide (→ O-desmethylencainide and 3-methoxy-O-desmethyl encainide), morphine (→ morphine-6-O-glucuronide), tamoxifen (→ 4-hydroxytamoxifen and endoxifen), and tramadol (→ O-desmethyltramadol)

Metabolites with comparable or improved therapeutic properties marketed as drugs include oxazepam (← diazepam), cetirizine (← hydroxyzine), desloratadine (← loratadine), paracetamol (← phenacetin), and fexofenadine (← terfenadine).

Prodrugs are medicinal compounds whose complete, or nearly complete therapeutic potential is based on the activity of the metabolite(s)[107-111]. Chemical groups liable to metabolism are often introduced to improve ADME (absorption, distribution, metabolism and excretion) properties or to reduce toxicity (e.g. chemotherapeutics). Examples include enalapril (→ enalaprilat), fenofibrate (→ fenofibric acid), levodopa (→ dopamine), oseltamivir (→ oseltamivir carboxylate), and valaciclovir (→ aciclovir).

Changes in pharmacological activity

Drug metabolites can range from intrinsically inactive to highly active, on the identical pharmacological target as the parent drug or any other biomolecule[4,112-114].

Gain in pharmacological activity:

Metabolism can produce active species that contribute to the therapeutic activity of drugs, and this actually is not a rare event. In the past drugs were often developed and almost solely guided by animal models, and it could happen that the observed biological response was to a significant part (or even completely) a result of an active metabolite and not the parent[113].

Gain of therapeutic activity may be observed for metabolites with favourable physicochemical properties for disposition, assisting transfer to their pharmacological target. Also chiral inversion as a result of metabolism may lead to severe changes in biological activity/toxicity. Metabolites may be (preferred) substrates of influx transporters.

Loss of pharmacological activity:

Extensive biotransformation of metabolically unstable molecules into inactive or rapidly excreted metabolites can result in a substantial drop or complete loss of therapeutic efficacy. This is in principle predictable and assays are available that can identify problematic molecules. Metabolic instability is particularly challenging if drug metabolism is induced as part of a resistance mechanism (most commonly observed for anti-infectives and anti-cancer drugs). Metabolites may also be substrates of efflux transporters.

Toxicity

Biotransformation bears the risk of unwanted toxification, resulting in adverse drug reactions (ADRs)[4].

On-target ADRs (in the context of metabolism) are a result of exceedingly high concentrations of a drug or an active metabolite, sometimes in a non-target tissue. They are generally dose-dependent and in principle predictable.

Off-target ADRs result from the interaction of a drug or its metabolite(s) with a non-intended target. A relevant example is cardiotoxicity caused by several (often but not exclusively) lipophilic drugs belonging to various pharmacological classes by blocking the human ERG potassium channel at therapeutic doses[115]. Off-target ADRs are generally dose-dependent and in principle predictable.

ADRs involving reactive metabolites are a reason for concern because they can involve covalent binding to biomolecules (adduct formation) and/or oxidative

stress following the formation of reactive oxygen species[116,117]. They are in principle predictable (or rationalisable) and dose-dependent.

Idiosyncratic drug reactions are rare (< 1 case in 5000) but with potentially severe consequences (anaphylaxis, blood dyscrasias, hepatotoxicity, and skin reactions). They are not currently predictable and apparently dose-independent. Their underlying mechanisms are poorly understood but are often thought to be associated with reactive metabolites[118].

Drug-drug interactions

Drug-drug interaction refers to the situation where the drug concentration present at (anti-) target sites is changed dramatically by the interference of another substance with drug-metabolising enzymes or related biomolecules. Potential outcomes of drug-drug interactions are (i) loss of pharmacological efficacy due to enhanced clearance, (ii) toxic effects caused by accumulation, and (iii) increase of the rate of reactive, toxic intermediates formed. They can be life threatening. For example, monoamine oxidase inhibitors are well known for potentially lethal dietary[119] and drug[120,121] interactions, and hence they are currently used only as a last resort for the treatment of atypical depression.

Drug resistance

Induction of metabolic pathways is a major route of drug resistance, in particular for anti-infectives and cancer drugs, but this also has huge implications for the effectiveness of pesticides. Multidrug resistance can also be driven by active efflux facilitated by transporters[122].

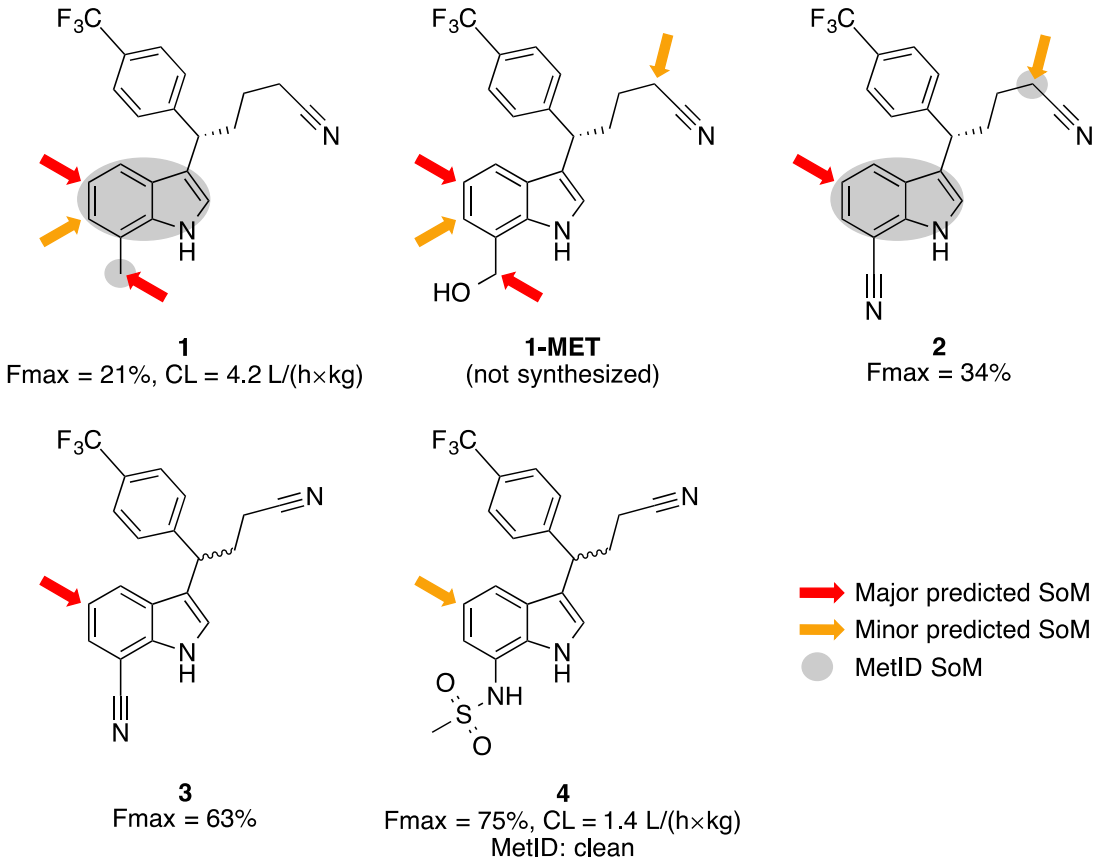
Changes in physicochemical properties

Biotransformation has an impact on physicochemical properties[6] and hence pharmacokinetics, in particular with respect to distribution and excretion. It also affects parameters such as absorption, passive membrane permeation, transport, binding to macromolecules etc.[9,12].

<< END OF BOX 1>>

Box 2. Combination of experimental and theoretical approaches to analyse and predict drug metabolism: An example from real life

The fruitful combination of experimental and computational approaches to optimize metabolic compound stability in a pharmaceutical industry setting can be exemplified by the work on a series of potent and selective non-steroidal mineralocorticoid receptor antagonists from a lead optimization project at Bayer Healthcare[123]. Since metabolite identification (MetID) assays are limited in throughput, most of the optimization of metabolic properties was done *via in vitro* and *in vivo* stability assays combined with sites of metabolism (SoM) prediction with the Bayer in-house software CypScore[89]. Additionally, *in silico* prediction allowed profiling of alternative designs before synthesis. In the course of the project, hundreds of compounds and metabolites were predicted *in silico*, whereas only a small amount of MetID experiments could be performed.



1
Fmax = 21%, CL = 4.2 L/(h×kg)

1-MET
(not synthesized)

2
Fmax = 34%

3
Fmax = 63%

4
Fmax = 75%, CL = 1.4 L/(h×kg)
MetID: clean

Mineralocorticoid receptor antagonist **1** showed low metabolic stability in hepatocytes with an Fmax (percentage of compound remaining) of 21% and also high blood clearance (CL) in rat of 4.2 L/(h×kg). *Via* MetID assay experiments in human and rat liver microsomes, oxidation of the of the 5- (or 6-) position of the indole and two-step oxidation of the methyl group leading first to the alcohol and then the acid were identified as the main metabolites. CypScore SoM predictions for **1** and the (not synthesized) metabolite **1-MET** were in accordance with experiment, but for **1-MET** an additional weaker SoM was predicted at the carbon in the α-position to the cyano group. There was no experimental in-house evidence for such a metabolic reaction.

In accordance with target SAR, the methyl group was replaced with a nitrile in **2**, which as expected resulted in an increase of stability for rat hepatocyte, Fmax to 34%, but not as much as expected. The major SoM was now the aromatic edge, but again accompanied by the weaker predicted SoM at the α-carbon of **2**, which was now also identified *via* MetID. Based on target fit and SoM prediction, the chain length was successfully reduced by one carbon leading to **3**. This increased the Fmax value to 63%. Further metabolic stabilization was consistently achieved by introducing derivatives with fluorination in the 5-position.

The metabolically labile methyl group in compound **1** was also replaced by various other substituents of different size, hydrophobicity and polarity. Several active sulphonamides were prepared that turned out to reduce the liability of the aromatic fragment, as exemplified by the low-clearance compound **4** with an Fmax = 75% and rat plasma clearance of 1.4 L/(h×kg).

This example shows the successful combination of various *in vitro* and *in vivo* experiments with *in silico* models to optimize the metabolic stability and in parallel target activity of a compound series.

<< END OF BOX 2>>

Figure 1

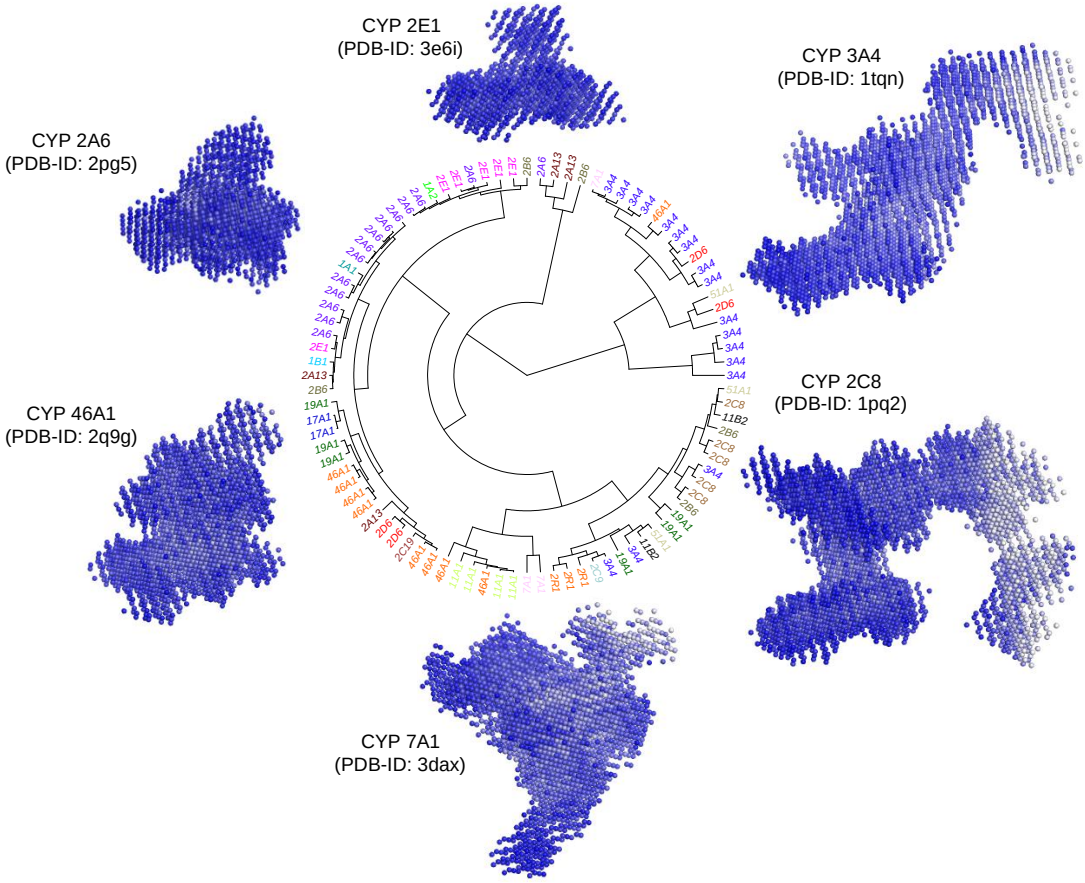


Figure 1. Shapes of cytochrome P450 binding pockets differ according to the class. The cluster diagram in the centre of the figure represents a grouping of the enzymes based on the shape and "buriedness" of their heme-containing active sites. Cartoons of substrate-free pockets are shown for selected cytochrome classes. The pockets were automatically extracted using the software PocketPicker[124]. Note that CYP2 and CYP3 families have evolved as essentially xenobiotic metabolising enzymes, whereas CYP11 and CYP46 are two among the many CYP families playing essential physiological roles. Colour intensity indicates the local buriedness (dark colour: Far from the protein surface, light colour: Surface exposed).

Table 1. Experimental Methods for Characterising Metabolic Stability and Identifying Sites of Metabolism.

Experimental approach	Scope
Incubations with individual drug-metabolising enzymes e.g. CYPs, UGTs	Method for determining enzyme involvement and specificity, mechanism–based inhibition and drug-drug interactions.
Microsomal incubations + NADPH	More detailed determination of oxidative metabolism, metabolic stability.
Microsomal incubations + UDPGA	Determination of glucuronidation.
Incubations with hepatocytes (from a range of species including human)	Determination of hepatic metabolite profiles including conjugations and reactive metabolite trapping, and metabolic stability.
Specific reactive metabolite trapping in microsomal incubations, trapping of soft nucleophiles using e.g. glutathione or cysteine, and hard nucleophiles with e.g. cyanide.	Identification of hot spots leading to reactive intermediates, using high resolution MS or, in complicated cases, NMR for precise structures. If reactive intermediates are detected, LC-MS-based proteomics can be used for further identification of sites of binding and nature of the reactive intermediate. If radiolabelled compounds are available then quantitative studies of covalent binding of reactive intermediates to proteins can be performed.
Animal models: Most commonly rodents (mouse and rat), including complex models designed to look at e.g. biliary elimination. Newer models include genetically modified and “humanized” strains (usually mice) as well as “chimeric” animals containing e.g. human hepatocytes.	Used to solve problems where e.g. *in vivo* pharmacokinetics are poorly predicted by *in vitro* studies, e.g. due to unchanged excretion, drug-drug interactions or more information on compound distribution is needed.

Table 2. Computer Software Used in Drug Metabolism Prediction.[a,b]

Predicting regioselectivity	Core component(s)	Type	Coverage	Licensing	Description
MEXAlert[125]	rules	LB, 2D	phase II	commercial	Quick screening tool to identify metabolically unstable molecules.
QikProp[126]	rules	LB, 2D	~20 phase I reactions	commercial	Fast SMARTS pattern matcher for predicting SoMs for phase I reactions.
MetaSite[59]	molecular interaction fields + reactivity estimator	SB, 3D	variety of CYPs	commercial	Uses MIFs derived from protein structures plus molecular orbital calculations to identify likely SoMs.
P450 Site of Metabolism Predictor[127]	induced-fit docking + reactivity estimator	SB, 3D	CYPs 2C9, 2D6, 3A4	commercial	Induced fit docking approach in combination with a quantum chemical model.
SMARTCyp[128,129]	DFT-derived reaction energies	LB, 2D	CYPs 1A2, 2A6, 2B6, 2C8, 2C19, 2E1, 3A4	free, open source	Utilizes a set of pre-computed activation energies in combination with topological accessibility descriptors. Available also as free online service.
StarDrop P450 Metabolism Prediction module[130]	semi-empirical method	LB, 3D	CYPs 2C9, 2D6, 3A4	commercial	Combines quantum chemical analysis with a ligand-based model of CYP substrates to identify SoMs. Takes into account calculated logP values.
CypScore[89]	surface electrostatics + semi-empirical method	LB, 3D	Individual CYP reactions	free reimplementation available; requires commercial software components	Collection of six MLR models to cover the major reaction types of CYPs.
Metaprint2D[48]	atom mapping; statistical model	LB, 2D	phase I+II	free, open source	Derives likelihoods of metabolic transformation for atoms with a defined atom environment by mining large biotransformation databases. Available also as free online service.

ADMET Predictor Metabolism module[131]	ANN ensemble	LB, 2D	CYPs 1A2, 2A6, 2B6, 2C8, 2C19, 2C9, 2D6, 2E1, 3A4	commercial	Derives likelihoods of metabolic reactions using ANN ensembles on a large, curated dataset.
Percepta P450 Regioselectivity module[132]	PLS (GALAS)	LB, 2D	human liver microsomal metabolism and CYPs 1A2, 2C9, 2C19, 2D6, 3A4	commercial	Global partial least squares-based QSAR model for calculating baseline regioselectivity; local corrections according to training data. Predicts and ranks major reaction types.
RS-WebPredictor[133]	MIRank (SVM)	LB, 2D	CYPs 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4	free	Array of pre-trained SVM models using topological descriptors and SMARTCyp reactivities for predicting SoMs. Intuitive web service.
FAME[91]	random forest	LB, 2D	phase I+II	free for academic use	Set of random forest models for predicting phase I and II metabolism in different species. Trained on drugs, drug-like molecules, endogenous metabolites and natural products.
Predicting metabolites	**Core component(s)**	**Type**	**Coverage**	**Licensing**	**Description**
MetabolExpert[134]	knowledge-based system	LB, 2D	phase I+II	commercial	Knowledge base containing rules and lists of substructures that inhibit or promote the reaction. Uses logP for filtering metabolites likely to be directly excreted. Predicts pathways in animals, plants or through photodegradation.
Meta-PC[135]	knowledge-based system	LB, 3D	phase I+II	commercial	Uses a large biotransformations dictionary. Analyses metabolite stability using quantum mechanical calculations and predicts pathways in mammals, through aerobic and anaerobic biodegradation.
Meteor Nexus[21]	knowledge-based system	LB, 2D	phase I+II	commercial	Employs a collection of knowledge-based biotransformation rules defined using a dedicated structure representation language. User-accessible knowledge base to aid the decision-making process. Considers calculated logP values for predictions. Latest version includes SMARTCyp.
MetaDrug[136]	knowledge-based system	LB, 2D	phase I+II	commercial	Derived from a large knowledge base (MetaBase). Generates metabolites from a dictionary of 160 rules. Predicted metabolites are rank-ordered and can be directly

					assessed regarding their pharmacological potential and toxicity.
TIMES[137]	knowledge-based system	LB, 2D	phase I+II	commercial	Employs a biotransformation library and a heuristic algorithm to generate metabolic maps. Dedicated models for skin metabolism, rat *in vitro* (S9) and *in vivo* metabolism.
SyGMa[22]	knowledge-based system	LB, 2D	phase I+II	available to academia	Predicts structures of likely metabolites based on rules derived from statistical analysis of several thousand biotransformations. Assigns probability scores to each metabolite.
EAWAG-BBD Pathway Prediction System[138]	knowledge-based system	LB, 2D	phase I+II	free	Rule-based system specialized in microbial catabolic metabolism of environmental pollutants. Classification of metabolites with respect to their likelihood. Intuitive web service.
JChem Metabolizer[139]	knowledge-based system	LB, 2D	phase I (can be extended to phase II)	commercial	Enumerates all possible metabolites of a given compound. Prognosis on metabolic pathways, major metabolites and metabolic stability. Species-specific predictions of metabolites.
Metaprint2D-React web server[48]	atom mapping; statistical model	LB, 2D	phase I+II	free	Generates structures of likely metabolites based on the MetaPrint2D data mining approach (using SMIRKS patterns). Intuitive web service.
MetaSite[59]	molecular interaction fields	SB, 3D	CYP	commercial	Produces a comprehensive set of metabolites from a collection of metabolic reactions. Calculates exact mass and relative retention times. Generated metabolites can be used for automated metabolite identification in Mass-MetaSite.
Predicting interactions of drugs with metabolising enzymes	**Core component(s)**	**Type**	**Coverage**	**Licensing**	**Description**
Percepta P450 Specificity module[132]	PLS (GALAS)	LB, 2D	CYPs 1A2, 2D6, 2C9, 2C19, 3A4	commercial	Collection of CYP models for predicting inhibitors and substrates. Based on same training data and modelling technique as the Percepta regioselectivity predictor (see above). Reports reliability measure for individual predictions.
Percepta Microsomal	random forest	LB, 2D	human liver microsomal	commercial	Random forest model for classifying small molecules into

Stability module[132]			metabolism		metabolically "stable", "unstable" or "undefined".
isoCyp[140]	SVM	LB, 3D	CYPs 2C9, 2D6, 3A4	commercial	Classification model based on 2D descriptors. Works under the assumption that any input molecule is substrate of one of the three isoforms.
MetaDrug substrate models[136]	recursive partitioning	LB, 2D	CYPs 1A2, 2B6, 2D6, 3A4	commercial	Collection of models for identifying substrates and inhibitors of important CYPs. Use recursive partitioning in combination with atom-centred fragment descriptors. Fully integrated into the MetaDrug pharmacology platform (web interface).
MetaDrug inhibitor models[136]	recursive partitioning	LB, 2D	CYPs 1A2, 2C19, 2C9, 2D6, 3A4	commercial	
MetaPred web server[141]	SVM	LB, 3D	CYPs 1A2, 2C9, 2C19, 2D6, 3A4	free	SVM model for predicting CYPs metabolising drug-like molecules.
ADMET Predictor Metabolism module[131] **for substrates**	ANN ensemble	LB, 2D	CYPs 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4	commercial	Predictor based on large curated dataset. Covers the most important CYPs and includes a dedicated model for predicting inhibitors of human liver microsomes. Predicts K_m and V_{max} values for hydroxylation reactions and CL_{int} resulting from the action of five CYPs.
ADMET Predictor Metabolism module[131] **for inhibitors**	ANN ensemble	LB, 2D	human liver microsomes and CYPs 1A2, 2C19, 2C9, 2D6 and 3A4	commercial	
WhichCyp[142] **web server**	SVM	LB, 2D	CYPs 1A2, 2C9, 2C19, 2D6, 3A4	free	Web service for predicting the selectivity of CYP inhibitors using SVM models derived from high-throughput screening data. Uses molecular signatures[143] as descriptors.
VirtualToxLab[144]	docking + QSAR	LB + SB, 3D	CYPs 1A2, 2A13, 2C9, 2D6, 3A4	commercial; free version available	Uses flexible docking in combination with a multi-dimensional QSAR approach to predict ligand interaction with 16 proteins, including CYP1A2, 2A13, 2C9, 2D6 and 3A4.
ADMEWORKS Predictor[145]	MLR	LB, 2D and 3D	CYP3A4 inhibitors CYP2D and 3A4 substrates	commercial	Collection of MLR-based QSAR models for the prediction of K_i values for CYP3A4 inhibition and K_m values for CYP2D and 3A4. Classification model for CYP3A4 inhibitors.
Predicting toxicological effects of metabolites	**Core component(s)**		**Coverage**	**Licensing**	**Description**

DEREK Nexus[21]	knowledge-based system	LB, 2D	broad range of toxicological endpoints	commercial	Predicts toxicological profiles by evaluating evidence for and against a broad collection of endpoints. Literature lookup functionality for examination of data underlying a prediction.
HazardExpert[146]	knowledge-based system	LB, 2D	7 different toxicity classes	commercial	Identifies toxic molecules based on fragments. Also calculates bioavailability, accumulation, and other parameters.
VirtualToxLab[144]	docking + QSAR	LB + SB, 3D	16 target proteins	commercial; free version available	Docking and QSAR hybrid approach for predicting activity on hERG, hormonal receptors, drug metabolising enzymes and their modulators.
MetaDrug[136]	recursive partitioning	LB, 2D	70 models for compound toxicity, ADME and therapeutic activity	commercial	Systems pharmacology platform built on data from MetaBase. Prediction of mechanism of action, toxicity, and off-target effects.
Leadscope toxicity models[147]	partial logistic regression	LB, 2D	8 sets of models for a broad range of toxicological endpoints	commercial	Collection of models for adverse cardiological, hepatobiliary and urinary tract effects, as well as developmental, genetic, neurotoxic, reproductive toxicity, and carcinogenicity.
(M)CASE Ultra & modules[148]	QSAR hybrid approach	LB, 2D	~450 models covering a wide range of toxicological and pharmacological measures	commercial	Uses a fragment-based approach and continuous parameters for modelling toxicity.
ToxTree[149]	decision tree	LB, 2D	17 modules for predicting toxicological and metabolic properties	free, open source	Free software for predicting various toxicological properties.
TOPKAT[150]	linear regression (different types)	LB, 2D	14 toxicity measures	commercial	Predicts toxicity measures in a variety of *in vitro* assays and animal models.
Percepta Toxicity modules[132]	PLS (GALAS)	LB, 2D	7 models for a variety of toxicity measures	commercial	Prediction of acute toxicity, aquatic toxicity, endocrine disruption, genotoxicity, hERG channel inhibition, irritation and health effects.
ADMET Predictor[131]	ANN ensemble	LB, 2D and 3D	19 models for a variety of toxicity and environmental	commercial	Prediction of endocrine disruption, hERG channel inhibition, skin sensitization, phospholipidosis etc.

			fate measures		
TIMES[137]	collection of QSAR models	LB, 2D	15 models for a variety of toxicity endpoints	commercial	Collection of models for evaluating the toxicity of metabolites, including skin sensitization, acute oral toxicity, phototoxicity, various endpoints related to endocrine disruption and cancerogenicity.
SYMMETRY[151]	collection of QSAR models	LB	>80 models for a variety of toxicity endpoints	commercial	Collection of models for preclinical and clinical predictive toxicology.

[a]Adapted with permission from [43]. Copyright 2012 American Chemical Society. Supplemented with data from [34], revised, and extended with additional software categories and products.

[b]Abbreviations: ADME, absorption, distribution, metabolism, excretion; ANN, artificial neural network; DFT, density functional theory; LB, ligand-based; MLR, multiple linear regression; PLS, partial least squares; SB, structure-based; SoM, site of metabolism; SVM, support vector machine.

Table 3. Scope and Limitations of Computational Methods.

Investigation/prediction of	Computational method(s)	Scope, limitations
Structure, function and mechanisms of metabolic enzymes	Homology modelling, quantum mechanics, molecular dynamics simulations etc.	Analysis of ligand binding events and enzyme mechanisms at a high level of detail and accuracy. Particularly useful for the investigation of unstable reaction intermediates with very short lifetimes.
Sites of metabolism	Knowledge-based systems, data mining, machine learning, QSAR models, reactivity models, ligand docking, molecular interaction fields, shape-based methods etc.	Able to predict the likely SoMs with adequate accuracy: In general at least one SoM is correctly identified among the three highest-ranked atom positions of a molecule in 70-90% of all cases[152] within the domain of chemical applicability.
Metabolites (chemical structure)	Knowledge-based systems, data mining	Dominated by knowledge-based systems. Can produce large numbers of metabolites. Main challenge: finding ways of ranking metabolites accurately.
Metabolic rates	Quantum mechanics, molecular dynamics simulations, (QSAR models)	Prediction generally not possible. Only within extremely narrow chemical space QSAR-like approaches *may* work.
Interactions of drugs with targets related to drug metabolism	QSAR models	Prediction of ligand affinity and inhibitory activity where adequate training data is available. Prediction of mechanism-based inhibitors remains highly challenging.
	Free energy calculations	Accurate prediction of binding affinities without need for extensive training data. Computationally expensive and labour-intense.
Bioactivity and toxicological effects	Various ligand- and structure-based approaches	Target prediction methods have become abundantly available but high false positives rates (i.e. accurate ranking of targets) remain a limiting factor. Prediction of bioactivities for metabolites hampered by lack of training data. Rule-based approaches are able to detect most toxicophores, but prediction of time-dependent inhibitors remains challenging.
Metabolite identification (MetID)	Various metabolite generation and spectra analysis approaches	Has seen major advances in recent years, driven by increasingly available data, data exchange and new algorithms. Major scientific instruments manufacturers offer bespoke MetID software. Vendor-independent and open-source packages are becoming increasingly available.

Table 4. Examples of Data Resources for Drug Metabolism and Overview of Their Information Content.

Resource	Description	Key figures (examples)	Drugs	Substrates, inhibitors, inducers, activators of metabolising enzymes	Metabolites	Metabolic reactions and/or pathways	Metabolizing enzymes	PK parameters	Drug transporters	Bioactivities	Additional ADME/Tox-related data	Structural data on target protein	NMR and/or mass spectra
ADME Database[153]	Database of interactions of small molecules with drug-metabolising enzymes and transporters.	>100,000 protein-ligand interactions	x	x			x	x	x	x			
BindingDB[154]	Bioactivity database focused on drug targets.	>1,000,000 bioactivities for >450,000 compounds	x	x	x	x	x	x	x	x			
ChEMBL[155]	Very large repository for bioactivity data.	>12,000,000 bioactivities for >1,600,000 compounds	x	x	x		x		x	x			
Drug Database (GOSTAR)[156]	Comprehensive resource for metabolites of approved drugs.	PK parameters collected from >50,000 publications.	x	x	x	x	x	x		x	x		
DrugBank[157]	Comprehensive encyclopaedic database on drugs.	>1,500 approved drugs >1,200 drug metabolites	x	x	x	x	x	x	x	x	x		**x**
EAWAG-BBD[158]	Data resource for biodegradation of xenobiotics, mainly environmental pollutants.	~1,400 molecules >1,500 reactions >200 degradation pathways	(x)	x	x	x	x						
Human Metabolome Database (HMDB)[159]	Database focused on human endogenous metabolites.	>41,000 metabolites		x	x	x	x	x	x		x		**x**
KEGG[160]	Large database comprising ~20 different (sub-) collections of data	>17,000 molecules (KEGG COMPOUND)	x	x	x	x	x		x	x	x		

	on metabolism.													
MetaBase[161]	Database of biological properties and effects of drugs and drug-like molecules.	>9,000 metabolic reactions >5,000 endogenous metabolites	x	x		x	x	x				x	x		
Metabolism and Transport Drug Interaction Database (DIDB)[162]	Collection of in vitro and in vivo data on drug-drug interactions in humans.	Data from >8,300 publications >70 new drug applications	x	x					x	x			x		
Metabolite[163]	Collection of metabolic pathways of drugs, drug-like molecules, natural products, and endogenous metabolites.	>100,000 biotransformations ~15,000 metabolic pathways	x	x		x	x	x					x		
METLIN[164]	Large and quickly expanding repository of metabolite information and MS/MS spectra. Powerful online search.	>240,000 metabolites >60,000 MS/MS spectra	x	x		x									**x**
PharmGKB[165]	Encyclopaedia focused on drugs. Rich in pharmacogenomics data.	>3,000 drugs >100 drug metabolism pathways	x	x		x	x	x	x	x	x	x	x		
PKKB[166]	Pharmacokinetic data for drugs.	>1,600 drugs >11,000 data points	x	x		x	x	x	x			x	x		
Protein Data Bank (PDB)[167]	Largest public repository for experimentally determined structures of biomolecules	> 100,000 structures	x	x				x			x	x		x	
PubChem[168]	One of the largest libraries of chemicals. Extensive pharmacological data for drug molecules.	>52,000,000 molecules	x	x		x	x	x	x	x	x	x	x		
SuperCyp[169]	Database for drug-CYP interaction. Contains biological and structural data, and information on CYP polymorphisms.	>1,100 drugs >2,700 drug-CYP interactions	x	x				x			x				
Wikipedia[170]	Largest and most popular general encyclopaedia.	Comprehensive coverage of marketed drugs, drug candidates and failed drugs.	x	x		x	x	x	x	x	x	x	x		