JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

FAst MEtabolizer (FAME): A Rapid and Accurate Predictor of Sites of Metabolism in Multiple Species by Endogenous Enzymes

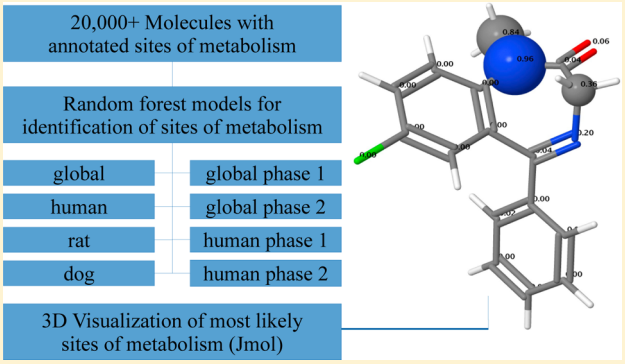
Johannes Kirchmair,[†] Mark J. Williamson,[†] Avid M. Afzal,[†] Jonathan D. Tyzack,[†] Alison P. K. Choy,[†] Andrew Howlett,[†] Patrik Rydberg,[‡] and Robert C. Glen*[,†]

[†]Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, United Kingdom

[‡]Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2100 Copenhagen, Denmark

S *Supporting Information*

ABSTRACT: FAst MEtabolizer (FAME) is a fast and accurate predictor of sites of metabolism (SoMs). It is based on a collection of random forest models trained on diverse chemical data sets of more than 20 000 molecules annotated with their experimentally determined SoMs. Using a comprehensive set of available data, FAME aims to assess metabolic processes from a holistic point of view. It is not limited to a specific enzyme family or species. Besides a global model, dedicated models are available for human, rat, and dog metabolism; specific prediction of phase I and II metabolism is also supported. FAME is able to identify at least one known SoM among the top-1, top-2, and top-3 highest ranked atom positions in up to 71%, 81%, and 87% of all cases tested, respectively. These prediction rates are comparable to or better than SoM predictors focused on specific enzyme families (such as cytochrome P450s), despite the fact that FAME uses only seven chemical descriptors. FAME covers a very broad chemical space, which together with its inter- and extrapolation power makes it applicable to a wide range of chemicals. Predictions take less than 2.5 s per molecule in batch mode on an Ultrabook. Results are visualized using Jmol, with the most likely SoMs highlighted.



INTRODUCTION

Metabolic transformations can have a decisive impact on the performance and safety of chemical products, including drugs, nutritional supplements, cosmetics, and agrochemicals. In recent years, emerging advanced analytical methods for investigating the metabolic fate of xenobiotics have led to a significant decrease of attrition rates related to problematic metabolic properties.[1,2] However, such experimental approaches for studying metabolic processes can involve a very high demand in scientific equipment, human expertise, and time. This is one of the reasons why computational methods for predicting metabolic processes are of great interest; within the last five years in particular, an array of novel predictive methods have become available.[3]

Computational approaches aim to predict metabolic processes and, hence, to provide decision support and guidance to experimentalists. They can make substantial contributions to the understanding of fundamental metabolic processes. Methods have been developed to investigate and predict inhibition and induction of metabolic enzymes, sites of metabolism (SoMs), and structures of metabolites.[3] For the latter, comparably few methods have been reported and evaluated[4] so far. One of the difficulties in predicting the structure of likely metabolites is that this task implies both correct identification of SoMs and metabolic reaction types. Current metabolite structure prediction methods are generally prone to a high false positive rate (i.e., a large number of metabolite structures are generated).[3]

Identifying SoMs of a molecule may give decisive hints in the development of a medicinal chemistry strategy to optimize metabolic properties and, consequently, crucial parameters such as toxicity, bioavailability, and bioactivity. Chemical reactivity and orientation of the ligand bound to the catalytic site of the enzyme are key for determining the location of SoMs. Many SoM predictors are available, and they have generally been shown to yield decent prediction rates.[5] However, there are also several limitations or disadvantages to consider. Some SoM predictors:

- involve complex quantum chemical calculations or simulation of changes to the protein conformation, resulting in long processing times (up to several hundred CPU hours for small molecules[6]), which is prohibitive for the interactive optimization of ligands

Received: August 28, 2013
Published: November 12, 2013

- include a large number of chemical descriptors, making these models noninterpretable
- are only applicable to the prediction of reactions of a specific metabolic enzyme, enzyme family (usually cytochrome P450), or metabolic phase (few methods support phase II metabolism)
- are focused on a specific species (usually human metabolism)
- are focused on a specific chemical domain (mostly drugs) and, within that chemical domain, are trained on a comparably small chemical space
- are based on a proprietary method or are not accessible to the public due to licensing issues
- do not have their underlying methodology published (e.g., for commercial reasons), which can limit their interpretability.

In this work we introduce FAst MEtabolizer (FAME) as a pragmatic approach to SoM prediction, aiming to overcome these limitations. FAME is a collection of random forest[7] models trained on a comprehensive and highly diverse data set of small molecules annotated with their experimentally determined SoMs. FAME is not constrained to a specific family of metabolizing enzymes (such as cytochrome P450s) or metabolic phase. A global model of metabolism has been generated as well as specific models for human, rat, and dog metabolism. In addition, dedicated models are also available to predict SoMs of phase I and II metabolism.

The method described here encodes 2D molecular structures using only six interpretable atomic descriptors (encoding the element type, hybridization state, and electronic configuration of each atom) and one molecular descriptor (encoding the topological size of a molecule). All descriptors can be rapidly calculated using open source software. In comparison studies, FAME performs as well as or better than currently available methods. It is available from the authors free of charge to academia and nonprofit organizations.

METHODS

Data Sources. Training and test data were extracted from the Metabolite Database version 2011.2.[8] Single-step metabolic reactions were used exclusively (this is to avoid data redundancy, since all multistep reactions are included in this database as a summary of constituent single-step reactions). Only reactions with a valid substrate and metabolite were processed (i.e., molecules must have a complete chemical structure, which, e.g. excludes molecules with "R" substituents). This preprocessing resulted in a data set of 79 238 reactions.

Annotation of Sites of Metabolism, Metabolic Phase, and Reaction Types. The Java- and Chemistry Development Kit[9,10] (CDK) based software framework MetaPrint2D[11,12] version 1.0 was adopted to identify and annotate the sites of metabolism (SoMs) of all molecules of the preprocessed Metabolite Database. To do so, a modified version of the recursive backtracking search[13] was used to derive the maximum common substructure of a substrate and its metabolite and consequently to identify all atoms modified during the course of a metabolic reaction (i.e., SoMs). This modified backtracking algorithm is explained in detail in ref 12.

In a second step, all SoMs identified for a specific molecule were merged and stored with a single representation of the molecule. Molecules comprising more than 100 heavy atoms were removed from the data set because calculation of some CDK descriptors (see below) can be disproportionally slow for some very large molecules. At the same time, statistical analysis also suggests that metabolites are generally under-reported for large substrates.[12] This data set is referred to as the global data set, since it is not constrained to a specific species (the vast majority of data comes from mammals).

MetaPrint2D-React[11,12] version 1.0 was used to derive the reaction types for all biotransformations and to categorize them into phase I and phase II reactions, according to the protocol published in ref 14.

Preparation of Molecular Structures. The molecules were prepared using the "Wash" function in MOE[15] version 2011.10. All settings were kept as their default values. During this process, hydrogens are added, with strong acids deprotonated and strong bases protonated.

Descriptor Calculation and Selection. *Group A: CDK Atomic Descriptors.* All descriptors were calculated using CDK version 1.4.18.

Group B: Span-Derived Descriptors. The Span2End descriptor and its components were calculated as reported in ref 16, but with the consideration of hydrogen atoms.

Group C: Sybyl Atom Type Descriptor. Sybyl atom types[17] were determined using the CDK SybylAtomTypeMatcher.

Group D: Atomic Fragment-Based Descriptors. A revised implementation of these recently reported descriptors[18] was used, which is available upon request from PR.

Information gain analysis was performed using the InfoGainAttributeEval class of Weka version 3.6.9.[19]

Generation of Training and Test Sets. On the basis of the global data set of molecules annotated with their SoMs (see above), a training set and three test sets were generated (Figure 1, Table 1) according to the following protocol:

1. The global data set was split into a training set (70%) and a test set (30%; test set 1).
2. Test set 2 was the subset of test set 1 consisting of all molecules with a maximum Tanimoto similarity coefficient of 0.8 when compared against all the molecules present in the training set. The similarity score was derived based on CDK extended fingerprints.
3. In a similar fashion, test set 3 was the subset of test set 1, consisting of molecules with a maximum Tanimoto similarity coefficient of 0.5.

This process was repeated to also generate species-specific data sets for human, rat, and dog metabolism, by filtering the metabolic reactions using metadata, as annotated in the Metabolite Database.

Model Generation. All models were generated with the random forest[7] implementation of Weka. Weka was also used for testing the effect of resampling the data (SpreadSubsample and SMOTE).

Model Evaluation. By definition, all molecules of the Metabolite Database contain at least one experimentally confirmed SoM. Molecules may, however, not include a SoM for both the metabolic phases. During the statistical evaluation of the performance of phase I and phase II-specific models, molecules were ignored if no SoM was reported for the respective metabolic phase.

Predictions with MetaPrint2D. SoMs were predicted with MetaPrint2D[11,12] version 1.0 using the default settings: the weights of the six fingerprint levels were set to 1.0, 1.0, 1.0, 0.75, 0.5, and 0.25: a similarity cutoff of 0.5 was used. Atom environments not covered by the model are assigned a

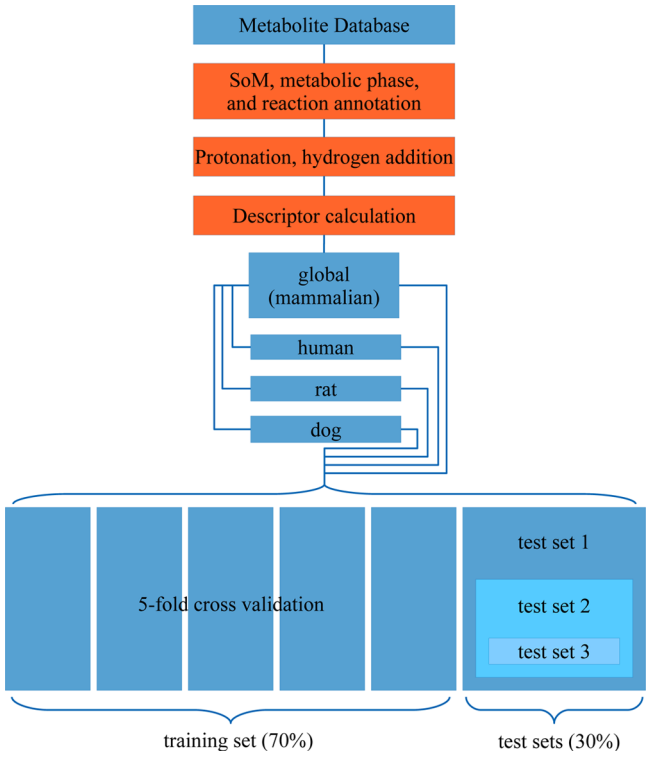


Figure 1. Workflow of this study and data sets used. Data processing steps are indicated in orange, and data sets in blue. The Metabolite Database served as the primary data source. After preprocessing the database, all remaining biotransformations were analyzed to derive the SoMs for all molecules. SoMs were also annotated with the underlying reaction type(s) and metabolic phase membership(s). This was followed by the preparation of the 2D chemical structures and descriptor calculation, resulting in the global data set. Species-specific subsets were derived for human, rat, and dog metabolism. Test set 1 and the training set were generated by random split (repeated for each individual species). The latter was also used to derive model parameters during 5-fold cross-validation. Test set 2 and test set 3 are subsets of test set 1, consisting solely of molecules that are structurally dissimilar to any of the molecules of the training data.

Table 1. Size of Data Sets Used in This Study

species	metabolic phase	no. mol[a] training set	no. mol[a] test set 1	no. mol[a] test set 2	no. mol[a] test set 3
all	phase 1 + 2	21098	9057	1889	181
human	phase 1 + 2	10347	4434	1065	149
rat	phase 1 + 2	13107	5622	1381	136
dog	phase 1 + 2	2929	1260	408	86

[a]Number of molecules.

mean_no_data_score of 0.159 (this value is derived from the mean normalized occurrence ratio of all non-novel atoms of a database sample[12]).

Predictions with SMARTCyp. SoMs were predicted with SMARTCyp[16,20] version 2.4.2.

Software and Hardware Configurations. All calculations were run on a machine with two Intel Xeon E5506 CPUs, 24 GB RAM, Ubuntu version 12.04.3, X86_64 architecture, and Oracle JRE version 1.7.0_25-b15.

RESULTS AND DISCUSSION

FAME is a site of metabolism (SoM) predictor based on a collection of random forest models trained on a preprocessed version of the Metabolite Database.[8] An overview of the workflow and all data sets is provided in Figure 1.

A SoM is defined as the atom where a metabolic reaction is initiated.[20] In order to define a SoM, the exact reaction mechanism underlying a biotransformation needs to be known or at least determinable. However, many reaction mechanisms of metabolic reactions are still poorly understood and information on SoMs is generally very sparse, in particular for enzymes other than cytochrome P450s.[21] For the purpose of modeling, we approximate SoM annotations by an atom mapping approach (initially reported in ref 22 and explained in more detail in ref 12). In short, any positions of a molecule where a heavy atom is added, removed, or altered (i.e., bond order change, ring closure, ring-opening) are mapped. The end user may actually benefit from this SoM mapping procedure: FAME is likely to flag one or two neighboring atoms of potential SoMs, and this contextually richer signal can provide valuable hints about which metabolic reaction may be taking place. For example, for atom positions susceptible to N-dealkylation, FAME generally flags the nitrogen and carbon atoms involved in a biotransformation (Figure 2).

Quality of the Training and Test Sets. The Metabolite Database[8] is a result of continuous data collection from the primary literature, conference proceedings, and new drug applications. It comprises more than 100 000 metabolic reaction entries, each of which consists of a single substrate and a single metabolite, plus metadata. In a recent study,[14] we have shown that the Metabolite Database offers a good coverage of drug-like chemical space, as well as the chemical space of human metabolites and natural products (molecules related to traditional Chinese medicine compounds in particular). As with other major databases,[23] a certain level of inaccuracy exists, partly because of errors in the primary literature. In the case of SoM prediction, the level of noise in the data is unlikely to be the limiting factor for model building, and appears to be a minor issue compared to the challenges arising from the incomplete coverage of metabolic processes:

- Reactive metabolites are difficult if not impossible to identify, but can be highly relevant from a biological perspective (e.g., toxic effects of reactive metabolites).
- Metabolites not excreted via main routes (e.g., desquamation, expiration, or perspiration) are often not captured by in vivo studies.
- Investigations are generally focused on the identification of specific metabolites of a parent molecule (usually, the major metabolite(s) or metabolites suspected to cause side effects).

A statistical analysis of the Metabolite Database found a tendency for molecules reported in earlier literature to have more metabolites reported on average.[12] It is plausible that more established molecules are better characterized than those more recently reported ones. However, it also needs to be considered that more recently reported molecules benefit from modern, more sensitive analysis methods employed for the initial characterization of metabolites. This may balance the different influences to a certain extent.

The underreporting of metabolite data, however, does not necessarily imply a capping of the performance and usefulness of predictors trained on the data that is available. The inherent

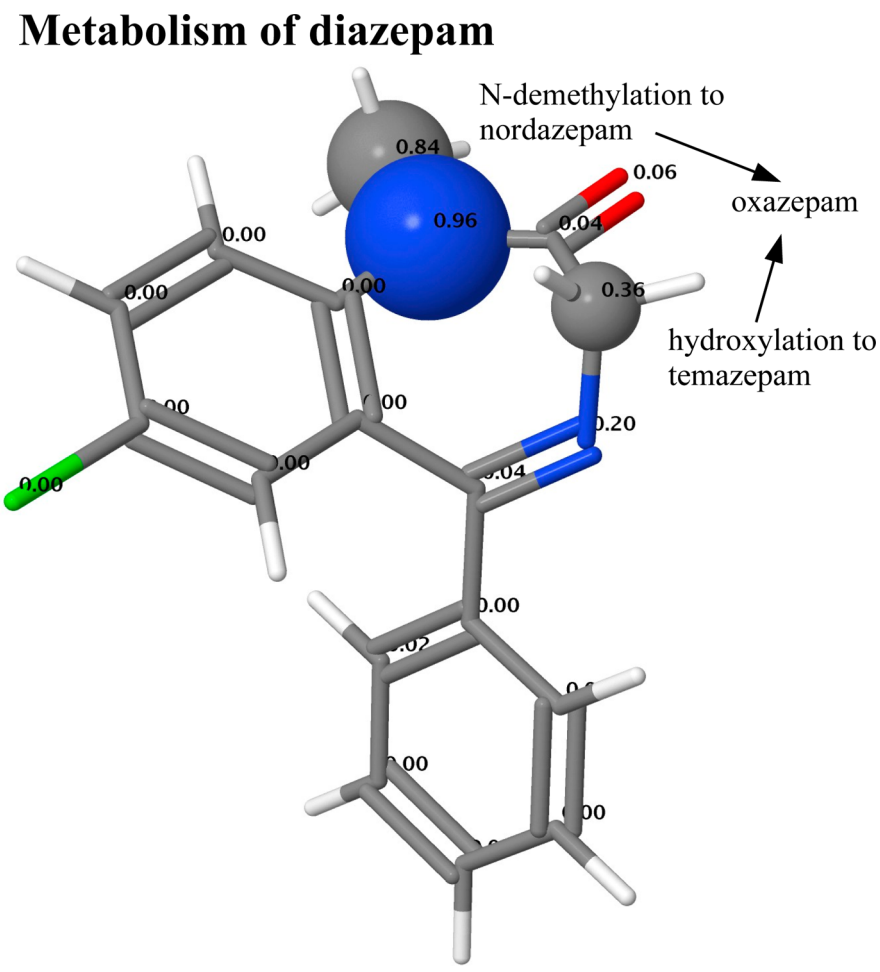


Figure 2. Visualization of FAME results for diazepam (human metabolism model) using Jmol. All heavy atoms have their probability of being involved in a metabolic reaction annotated. The three most likely SoMs are highlighted with spheres (radius decreasing with probability). The prediction corresponds to the experimentally determined metabolites, the N-demethylated nordazepam, and the hydroxylated temazepam (reactions indicated by arrows).

Table 2. Overview of Descriptors Used in this Study

descriptor name	descriptor group	definition	information gain
[a]PartialTChargeMMFF94	A	total partial charges of a heavy atom as derived from the MMFF94 model	0.0741
[a]PartialSigmaCharge	A	Gasteiger−Marsili sigma partial charges in sigma-bonded systems	0.0661
[a]PiElectronegativity	A	pi electronegativity	0.0608
[a]SigmaElectronegativity	A	Gasteiger−Marsili sigma electronegativity	0.0576
[a]SybylAtomType	C	Sybyl atom type for a specific atom, encoding element type and hybridization state	0.0411
[a]EffectiveAtomPolarizability	A	effective atom polarizability of a heavy atom	0.0180
[a]MaxTopDist	B	maximum topological distance between two atoms of a molecule	0.0149
AtomValence	A	atom valence	0.0140
Span2End	B	maxTopDist − topDist	0.0139
RelativeSpan	B	topDist/maxTopDist	0.0106
AtomHybridizationVSEPR	A	hybridization state of an atom based on the VSEPR model	0.0097
AtomHybridization	A	hybridization state of an atom	0.0097
TopDist	B	maximum topological distance between an atom and the end of a molecule	0.0096
StabilizationPlusCharge	A	stabilization of the positive charge	0.0034
AtomDegree	A	number of non-hydrogens attached to an atom	0.0019

[a]Descriptor was used for model generation.

bias toward a characterization of the most important metabolites may actually help to pinpoint the most relevant SoMs. This can, however, have a significant impact on the evaluation of SoM predictors, and it is to be expected that a performance plateau will be observed.

In this study, the SoMs of all molecules of the preprocessed Metabolite Database were derived using MetaPrint2D,[11,12] resulting in data sets of up to more than 20 000 unique molecules annotated with their SoMs (Table 1). To our best knowledge, these SoM data sets are the most comprehensive

dx.doi.org/10.1021/ci400503s | J. Chem. Inf. Model. 2013, 53, 2896−2907

Table 3. Performance of the Human Metabolism Model[a]

top-k	no. trees 5	no. trees 10	no. trees 25	no. trees 50	no. trees 75	no. trees 100	no. trees 50 balanced 1:1 with SpreadSubSample	no. trees 50 balanced ~1:1 with SMOTE
1	0.64	0.66	0.68	0.69	0.69	0.69	0.60	0.68
2	0.76	0.78	0.80	0.80	0.80	0.81	0.77	0.80
3	0.82	0.84	0.86	0.86	0.87	0.87	0.85	0.86
average	**0.74**	**0.76**	**0.78**	**0.79**	**0.79**	**0.79**	**0.74**	**0.78**

[a]Values reported for 5-fold cross-validation.

sets that have been used to date for model building and evaluation.

Descriptors. The Java-based Chemistry Development Kit (CDK) was selected as an open source library for calculating atomic descriptors. CDK supports the calculation of 30 atom-based descriptors. After removal of (i) all 3D descriptors, (ii) descriptors that are not or not wholly applicable to the problem, and (iii) descriptors that can require a comparably long calculation time, a group of ten descriptors remained (group A). For more detail see Supporting Information Table 1.

Group B comprises the Span2End descriptor and its components.[16] Span2End is an accessibility descriptor which encodes the steric exposure of an atom to the catalytic center of a metabolizing enzyme. It measures whether an atom is located at the center or edge of a molecule, where in the latter case it is assumed that the atom is more easily transformed by the metabolic enzyme. This and related descriptors are used in other SoM predictors such as SMARTCyp.[16,20]

Sybyl atom types[17] were added to encode the element type and hybridization state (group C). They have been proven useful in MetaPrint2D,[11,12] a fingerprint-based SoM predictor (see below).

Group D comprises 18 atomic fragment-based descriptors encoding the chemical properties of fragments separating a specific atom from the closest end of a branch or the molecule.[18] Properties include volume, number of rotatable bonds, heavy atoms, hydrogen bond donors/acceptors, etc.

Groups A, B, and C were reduced down to a total of seven descriptors based on an information gain analysis of the human training set (Table 2). Feature selection was preceded by iterative model training and cross-validation using different combinations of descriptors. It was found that the feature ranking obtained from information gain reflected well the observed relationship between descriptor sets and model performance.

Five of the selected descriptors encode the electronic state of heavy atoms (i.e., total partial charges, sigma partial charges, pi electronegativity, sigma electronegativity, and polarizability). According to information gain, the most important attribute is the PartialTChargeMMFF94 (Table 2). The remaining two descriptors are Sybyl atom type and the maximum topological distance descriptor (i.e., maximum number of chemical bonds between any pair of atoms of a molecule).[16]

Descriptors discarded were AtomValence, AtomHybridizationVSEPR, AtomHybridization, and AtomDegree, which are in part implicitly described by SybylAtomType. In an analogous fashion, MaxTopDist was selected as the representative of the group of span-related descriptors: Span2End, RelativeSpan, and TopDist. StabilizationPlusCharge was mostly zero for both classes and hence discarded.

In dedicated experiments group D was added to this set of seven descriptors in order to investigate whether this leads to an increase in performance.

Model Generation. The random forest approach is an ensemble learning method that can be applied to classification and regression problems.[7] Underlying this methodology is the combination of classification/regression trees trained with bootstrapped samples of a data set. Random forests were chosen for their high accuracy, efficiency, and robustness proven for the problem of SoM prediction[24] and in the chemical context in general.[25]

Performance of the individual models was measured with the top-k (standard) accuracy metric, the de facto standard for evaluating SoM predictors.[5,20,26,27] The metric denotes the percentage of molecules for which at least one experimentally confirmed SoM is present among the k top-ranked atom positions of the individual molecules.

A suitable value for the number of trees used for the random forest models was identified during a series of individual 5-fold cross-validation runs based on the human data set and the reduced set of seven descriptors from descriptor groups A, B, and C (Table 3). Performance increases with the number of trees until reaching a plateau level with the number of trees equal to 50, which is top-1, top-2, and top-3 scores of 0.69, 0.80, and 0.86, respectively. No gain in accuracy can be observed for models based on more than 50 trees.

We also investigated the effect of data balancing techniques on the model performance. When reducing the number of instances in the majority class in order to obtain a balanced data set a drop rather than an increase in prediction accuracy was observed (Table 3). Also resampling of the data set using the synthetic minority oversampling technique[28] (SMOTE) did not lead to an increase in performance (Table 3). For this reason we used the original data sets for model generation.

Consideration of further descriptors, such as the 18 atomic fragment-based descriptors of descriptor group D, did not yield better results (Table 4). Hence, for all further experiments the seven descriptors selected from groups A, B, and C were used, and the number of trees was set to 50.

Overall Performance. A global SoM model and specific SoM models for human, rat, and dog metabolism were generated. The performance of all models was assessed using 5-fold cross-validation. In addition, the models were evaluated using three test sets, with different degrees of similarity between the training and the test set:

- Test set 1: Generated by a random split of the global data set into a training and test set.
- Test set 2: Subset of test set 1, comprising only molecules with a maximum fingerprint-based Tanimoto similarity coefficient of 0.8 with any of the molecules present in the training set (see Methods).
- Test set 3: Same as test set 2, but the similarity criterion is lowered to 0.5.

Table 4. Evaluation of Different Descriptor Sets[a]

	I[b]	II[c]	III[d]	IV[e]
top-1	0.686	0.689	0.691	0.693
top-2	0.798	0.796	0.804	0.801
top-3	0.863	0.861	0.864	0.867
AUC	0.865	0.866	0.864	0.870

[a]Prediction rates for 5-fold cross-validation of models based on human metabolism data. [b]All 15 descriptors of descriptor groups A, B, C. [c]All 15 descriptors of descriptor groups A, B, C, plus all 18 descriptors of descriptor group D. [d]Seven selected descriptors from descriptor groups A, B, and C. This descriptor set was found most suitable for model building and all final models of FAME are based on this descriptor set. [e]Seven selected descriptors from descriptor groups A, B, and C, plus 18 atomic fragment-based descriptors (descriptor group D).

By comparing the prediction rates for these three test sets, we aimed to learn how well these models are able to extrapolate to new chemical space.

The global model obtains top-1, top-2, and top-3 prediction rates of 71%, 81%, and 87% on test set 1 (Table 5). The human and rat model obtain comparably good prediction rates, even though their training sets are only about half the size of the global model's. A noticeable performance drop is observed for the model on dog metabolism, which loses about 3−7 percentage points compared to the global model. This is related to the fact that there is less data available on dog metabolism, both in terms of the number of molecules (about 15% of the training set for the global model) as well as in the number of SoMs reported per parent molecule. The latter is reflected by the random success rates, which are also reported in Table 5.

For compounds more dissimilar to molecules present in the training data (contained in test set 2), the top-1 prediction rates drop by about 8−12 percentage points, respectively. Performance is less affected when taking into account more (i.e., three) of the top-ranked atom positions (performance drop of 4−5 percentage points). From test set 2 to test set 3 (data set consisting of molecules most dissimilar to the training set) prediction rates drop by 1% point on average; however, it needs to be considered that random success rates for test set 3 are higher (Table 5).

Performance differences between the individual models and the various test sets are reflected by ROC curves and AUC values (Figure 3). The ROC curves also indicate that FAME models obtain good early enrichment rates. For example, for test set 1, about half of all experimentally determined SoMs can be recovered with a false positive rate of 0.05, and close to 70% of all SoMs with a false positive rate of 0.10 (with the exception of the model for dog metabolism, which performs slightly worse).

Overall, the moderate decrease in prediction rates for molecules very dissimilar to instances of the training set indicates that FAME has good extrapolation power, making the method applicable to the assessment of molecules of very different chemical spaces.

Significance of the Size of the Training Data. We also investigated the dependency of prediction performance on the size of the training data set (by random resampling of the training data). This analysis can be of interest to those with specific in-house data e.g. pharmaceutical companies, allowing the deduction of an estimate of how much (in-house) data (covering specific areas of chemistry) is required for building specific models and to improve prediction accuracy. As shown in Figure 4, there appears to be a linear relationship between the size of the training set and prediction accuracy, and, unsurprisingly, the more data the better the model performance. It is also apparent from this graph that the top-3 prediction rate is less affected by the size of the training sets than the top-1 prediction rate.

Significance of Molecular Size. The problem of identifying a SoM becomes increasingly difficult as the size of a molecule increases (as reflected by the number of heavy atoms). An ideal model would be able to identify SoMs independently of the size of a molecule. Figure 5 shows that there is a gradual decline in the percentage of successfully predicted molecules with respect to the number of heavy atoms. Good prediction accuracy is largely maintained for drug-sized molecules (i.e., up to a molecular weight of 500 Da, equivalent to about 35 heavy atoms), with the exception of the top-1 prediction rate on test set 2.

Prediction Rates for Phase I and II Metabolism. Metabolic reactions can be classified into phase I and phase II reactions. Phase I reactions are catalyzed by a wide range of metabolic enzymes and include modifications such as oxidation, hydrolysis, reduction, ring closing, and opening, etc. Cytochrome P450 enzymes are the most important enzymes catalyzing phase I reactions. Phase II reactions are conjugation

Table 5. Performance of FAME Measured with the Top-*k* Metric

metabolic phase	species	*k*[a]	5×CV[b]	test set 1	test set 2	test set 3	random test set 1[c]	random test set 2[c]	random test set 3[c]
phase I + II	all	1	0.70	0.71	0.61	0.57	0.18	0.22	0.28
phase I + II	all	2	0.81	0.81	0.76	0.78	0.30	0.37	0.44
phase I + II	all	3	0.87	0.87	0.83	0.85	0.40	0.48	0.56
phase I + II	human	1	0.69	0.70	0.58	0.55	0.16	0.20	0.26
phase I + II	human	2	0.80	0.80	0.74	0.73	0.28	0.34	0.42
phase I + II	human	3	0.86	0.87	0.82	0.83	0.38	0.45	0.53
phase I + II	rat	1	0.68	0.70	0.61	0.53	0.19	0.22	0.26
phase I + II	rat	2	0.80	0.81	0.75	0.72	0.32	0.37	0.42
phase I + II	rat	3	0.86	0.87	0.83	0.79	0.42	0.48	0.54
phase I + II	dog	1	0.60	0.64	0.56	0.56	0.14	0.17	0.20
phase I + II	dog	2	0.74	0.75	0.69	0.72	0.26	0.30	0.36
phase I + II	dog	3	0.82	0.84	0.80	0.82	0.36	0.41	0.48

[a]Number of top-ranked atom positions considered for the determination of prediction success. A prediction is deemed correct if at least one known SoM is among the *k* top-ranked atom positions. [b]5-fold cross-validation. [c]Random success rate.

dx.doi.org/10.1021/ci400503s | J. Chem. Inf. Model. 2013, 53, 2896−2907

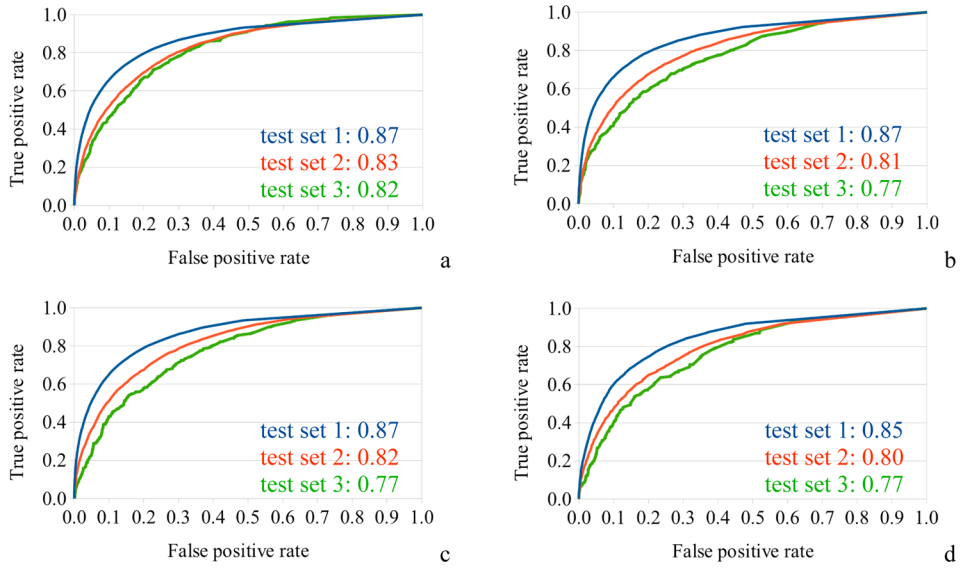


Figure 3. ROC curves for the (a) global, (b) human, (c) rat, (d) and dog models, for test set 1 (blue), 2 (orange), and 3 (green).

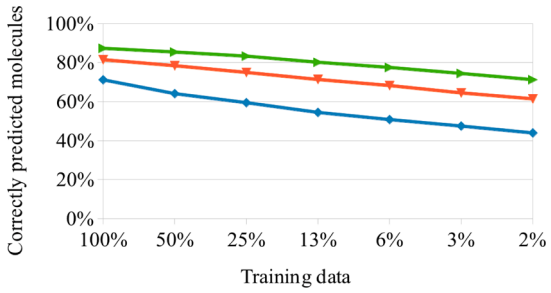


Figure 4. Dependency of the model performance on the size of the training data set, as exemplified by the global metabolism model. Top-1 (blue), top-2 (orange), and top-3 (green) prediction rates. 100% of training data correspond to 21 098 molecules.

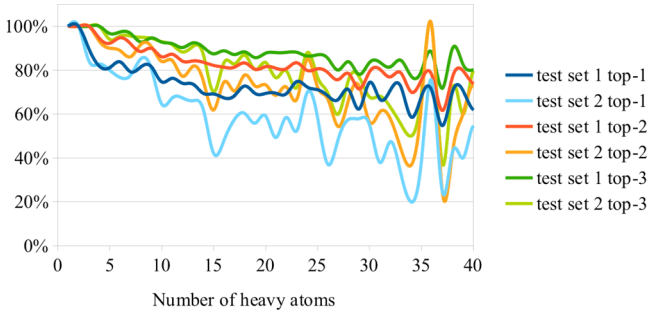


Figure 5. Percentage of correctly predicted molecules with respect to their size for test set 1 and test set 2 (global model). A gradual decline with increasing molecular size can be observed.

reactions and often follow phase I reactions. Common groups attached during phase II biotransformations are glucuronic acid, glutathione, glycine, acetyl, and methyl. The physicochemical property shifts introduced to small molecules by phase I and II reactions differ greatly.[14]

MetaPrint2D-React,[11,12] an extension of MetaPrint2D, allows the prediction of biotransformation reaction types and their categorization into phase I and II reactions. This tool was used to assign the reaction type to each SoM (this process is described in detail in ref 14.). By this processing, it was possible to generate phase-specific data sets, which were used for training phase-specific models. The performance of the phase I models compares well to the general models for metabolism, with top-3 prediction rates for phase I metabolism ≥80%, even for test set 3 (Table 6). Top-3 prediction rates are about 85% for all phase II models on test set 1. However, some of these models appear to be weaker in their ability to inter- and extrapolate, with some prediction rates below 80% for test set 2 and 3. At the same time, early enrichment is better for phase II than phase I models (see Figure 6).

Prediction Rates for Specific Metabolic Reaction Types. It is also important to analyze the performance with respect to the underlying metabolic reaction type in order to avoid bias toward specific, possibly dominant reaction types. Figure 7 reports the top-3 prediction rates for the most commonly observed metabolic reactions (global data set). The numbers need to be interpreted with some caution, because for a significant number of metabolic reactions included in the Metabolite Database the explicit reaction type cannot be determined by MetaPrint2D. This is often related to the fact that reactions, even though annotated as single step reactions in the database, are actually a product of two or more biotransformations. For this reason we selected test set 1 for the analysis of reaction types, since it contains a large number of reactions (compared to the test set 2 and test set 3) to average out some of the noise in the data. It is apparent from Figure 7 that the prediction rates are not correlated with the propensity of a specific metabolic reaction (which reflects the proportion of data points available for training the models), and this is expected for several reasons. Some metabolic reactions occur only at very specific atom environments, and if present in a molecule, with a high probability. Such reactions are comparably easy to predict, even if only a few instances are available in the training set. An example of this type of reaction is epoxide hydrolysis. Also the chemical environment susceptible to glucuronidation can be easily determined (even though in this case the propensity is relatively high): glucuronidation reactions often occur at hydroxy groups that were introduced during a phase I reaction. In contrast, despite their frequent occurrence, specific hydroxylation reactions are comparably difficult to predict, mainly because the number of possible hydroxylation sites is often high and a near miss is

dx.doi.org/10.1021/ci400503s | J. Chem. Inf. Model. 2013, 53, 2896−2907

Table 6. Performance of FAME Metabolic Phase-Specific Models Measured with the Top-*k* Metric

metabolic phase	species	k^a	5×CV^b	test set 1	test set 2	test set 3	random test set 1^c	random test set 2^c	random test set 3^c
phase I	all	1	0.71	0.72	0.65	0.62	0.19	0.24	0.28
phase I	all	2	0.81	0.81	0.77	0.80	0.32	0.40	0.44
phase I	all	3	0.87	0.87	0.85	0.89	0.43	0.51	0.56
phase I	human	1	0.69	0.71	0.61	0.58	0.17	0.22	0.27
phase I	human	2	0.79	0.81	0.75	0.79	0.30	0.37	0.44
phase I	human	3	0.86	0.87	0.83	0.85	0.40	0.48	0.55
phase I	rat	1	0.69	0.71	0.63	0.59	0.20	0.23	0.27
phase I	rat	2	0.80	0.81	0.76	0.73	0.34	0.39	0.44
phase I	rat	3	0.87	0.87	0.83	0.80	0.44	0.50	0.55
phase I	dog	1	0.62	0.68	0.62	0.61	0.15	0.18	0.22
phase I	dog	2	0.74	0.77	0.75	0.79	0.27	0.32	0.38
phase I	dog	3	0.82	0.85	0.81	0.81	0.38	0.43	0.51
phase II	all	1	0.71	0.70	0.58	0.46	0.09	0.11	0.14
phase II	all	2	0.81	0.81	0.72	0.64	0.17	0.21	0.26
phase II	all	3	0.85	0.85	0.79	0.73	0.24	0.30	0.36
phase II	human	1	0.73	0.71	0.59	0.48	0.08	0.11	0.13
phase II	human	2	0.81	0.80	0.73	0.62	0.16	0.21	0.25
phase II	human	3	0.85	0.84	0.77	0.67	0.23	0.29	0.33
phase II	rat	1	0.69	0.70	0.57	0.38	0.09	0.12	0.14
phase II	rat	2	0.80	0.81	0.71	0.53	0.18	0.22	0.27
phase II	rat	3	0.84	0.85	0.78	0.58	0.25	0.31	0.36
phase II	dog	1	0.66	0.69	0.59	0.73	0.06	0.08	0.09
phase II	dog	2	0.79	0.79	0.67	0.75	0.13	0.17	0.19
phase II	dog	3	0.82	0.84	0.72	0.76	0.19	0.25	0.28

^{a}Number of top-ranked atom positions considered for the determination of prediction success. A prediction is deemed correct if at least one known SoM is among the k top-ranked atom positions. ^{b}5-fold cross-validation. ^{c}Random success rate.

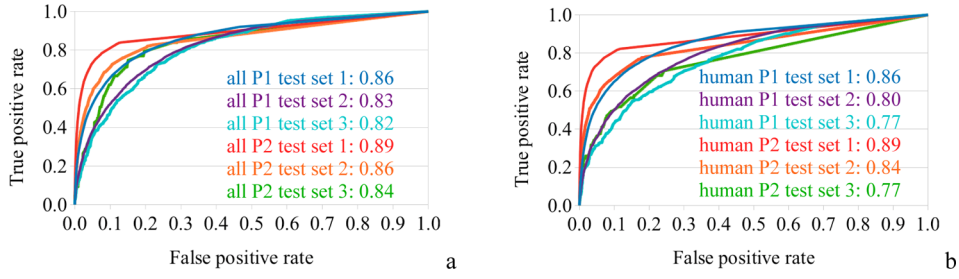


Figure 6. Performance of metabolic phase-specific FAME models: (a) global models, (b) human models.

more likely. For this important reaction type, performance may be improved by the introduction of additional rules or descriptors.

Comparison of FAME and MetaPrint2D. MetaPrint2D is one of the few available SoM predictors not limited to the prediction of reactions for a specific family of metabolizing enzymes (such as cytochrome P450s). The method derives a statistical model from mining databases of metabolic reactions such as the Metabolite Database. It screens the individual biotransformations to calculate the likelihood (occurrence ratio) of each atom of each substrate being involved in a metabolic reaction. An atom is encoded as a circular fingerprint of up to six levels. Once the model is built, a query molecule can be provided to MetaPrint2D. MetaPrint2D then encodes all atoms of this molecule as fingerprints, searching for matching instances in its database (the statistical model). An occurrence ratio is reported for the query fingerprint. In many cases no exact match can be obtained for the fingerprint of a query atom. For this reason MetaPrint2D calculates fingerprint similarities in order to extrapolate and calculate the likelihood of an atom being metabolised.

MetaPrint2D was selected as a reference for the performance of holistic SoM predictors for three reasons: (i) the method is conceptually related to FAME and, hence, allows direct comparison, (ii) the source code of the program is available and models can be trained by the user (which was done in this study), assuring that the models are not biased toward the testing data, and (iii) the software is designed to also run in batch mode.

Comparison of MetaPrint2D with FAME shows that the prediction rates of the latter are about 10 to 20 percentage points higher (Figure 8). With FAME, prediction rates for test set 3 (highly challenging) are as good as or better than with MetaPrint2D for test set 1 (can contain molecules similar to those in the training data). Besides the generally better performance, it is also apparent that FAME is noticeably stronger in inter- and extrapolating the metabolic reactivity of molecules that are structurally unrelated to instances in the training data. This is plausible, as FAME aims to encode metabolic reactivity with chemical descriptors, while MetaPrint2D is a fingerprint-based method, which encodes specific atom environments (structural patterns) particularly well but is

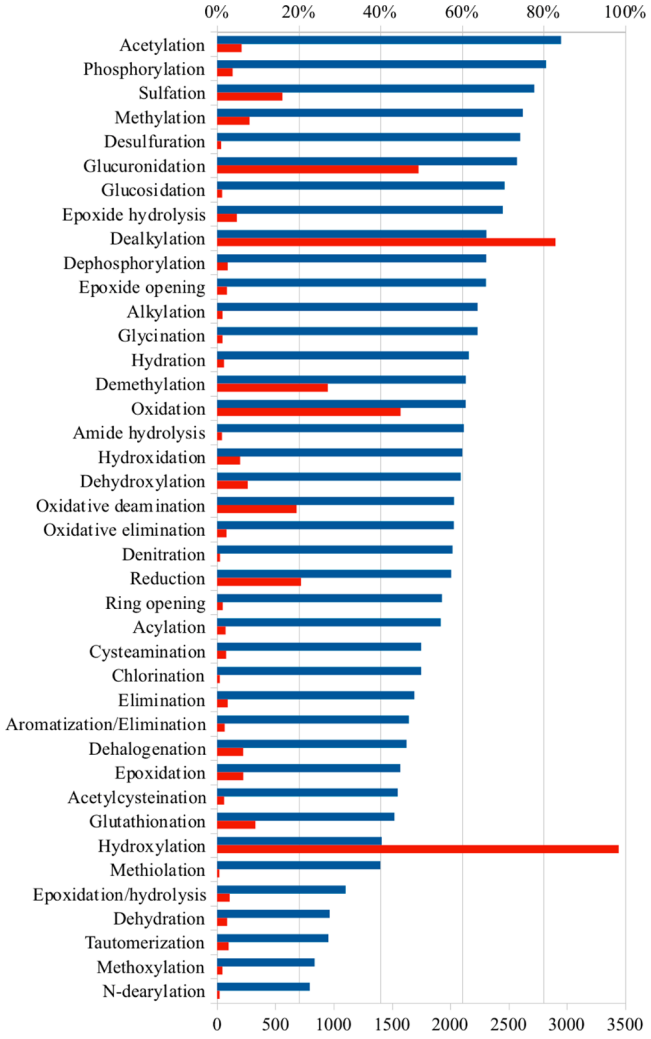


Figure 7. Top-3 prediction rates for the global metabolism model on test set 1, itemized with respect to the reaction type. No correlation between propensity of a reaction type (the number of reactions: red bars) and prediction rates (blue bars) is observed. The graph reports on all reaction types with at least 20 instances in test set 1.

limited in its capability to apply this to new chemical environments. Caution must be exercised regarding the statistical significance of the observed performance differences: FAME produces a density distribution while MetaPrint2D reports occurrence ratios, and in cases where no occurrence ratio for a an atom with a specific atom environment can be determined (i.e., if not present in the training data) a rate of 0.159 is assigned to this atom (i.e., the mean normalized occurrence ratio for the Metabolite Database).

Comparison of FAME to Cytochrome P450 SoM Predictors. A recent publication has summarized and compared the performance of most SoM predictors available to date.[5] This study shows that state of the art methods obtain top-2 prediction rates of 70−85% approximately. FAME reaches a comparable level of accuracy (Tables 5 and 6), but several caveats need to be considered when attempting to directly compare these numbers. First, the methods compared in this study are predictors of cytochrome P450 reactions only, while FAME is dealing with the increased complexity of predicting the function of a wide range of metabolic enzymes. The test sets and definitions of SoMs (SoMs as defined by FAME basically represent a superset of the mechanism-based SoMs as defined by SMARTCyp, see above) differ among the various studies. As a matter of fact, molecules annotated with SoMs of cytochrome P450 reactions have fewer SoMs on average,[26] and, hence, the random guess rate is lower (see Table 5 for the random success rates for the FAME test sets). This brings an advantage to FAME, which is expected to balance the increased complexity of multienzyme data.

The test sets for human phase I metabolism used in this study come closest to the applicability domain of SMARTCyp.[16,20] SMARTCyp is a reactivity-based model for human cytochrome P450-based metabolism. It computes likely SoMs by estimating transition state energies from a library of precalculated fragments. SMARTCyp also uses span descriptors accounting for steric accessibility. A web server is available.[29]

Aware of the caveats mentioned above, we were interested in obtaining an estimate of how well FAME compares to SMARTCyp, or more accurately, how many of the phase I reactions are covered by SMARTCyp. As reported in Table 7, SMARTCyp covers about 20 to 30 percentage points fewer phase I SoMs than FAME. The numbers reported here should be considered as rough estimates. It is more interesting to confirm that prediction rates for SMARTCyp do not drop between test set 1 and test set 3 and are correlated with the random success rates. This is expected since the test sets used here are designed to be increasingly challenging for FAME and MetaPrint2D. Some molecules of the SMARTCyp calibration set or analogs thereof may be present in these phase I metabolism test sets.

We also performed the reverse experiment, testing FAME on the cytochrome P450 SoM data set of Zaretzki et al.[21] This data set (the nonisoform-specific, merged version) contains 680 molecules, each with their annotated SoMs. Table 8 reports that the top-1, top-2, and top-3 prediction rates of SMARTCyp are 14, 5, and 1 percentage points higher than for FAME. The reason for the larger difference in the top-1 prediction rate when compared to the top-2 and top-3 prediction rates is the discrepancy in the definition of SoMs between FAME (atom mapping-based) and SMARTCyp (mechanism-based). For example, in the case of N-/O-dealkylation this means that FAME ranks the heteroatom higher than the carbon atom and the carbon atom is the mechanism-based SoM of this reaction annotated in the Zaretzki data set. Hence, when using FAME on the Zaretzki data set, it is likely that FAME gives a high probability value to both atoms, but the carbon atom (the mechanism-based SoM) will be ranked second (see Figure 2 for an example). This is a major reason why we see closer top-2 and top-3 prediction rates between FAME and SMARTCyp compared to top-1 values. Again, it needs to be considered that molecules used for training FAME and SMARTCyp or their analogs can be present in the (Zaretzki) test set. The main caveat remains however that SMARTCyp was explored partly outside its applicability domain.

The conclusion to draw from these experiments is that FAME offers competitive accuracy, also on cytochrome P450 metabolism. Results from this study and the many other reports on SoM predictors[5] indicate a plateau level for top-2 and/or top-3 prediction rates at about 85%. As mentioned earlier, the performance of SoM predictors is primarily limited by the availability of data, its quality and completeness. In particular the latter can be responsible for the observed plateau, and, hence, the SoM prediction method appears to be generally underestimated.

dx.doi.org/10.1021/ci400503s | J. Chem. Inf. Model. 2013, 53, 2896−2907

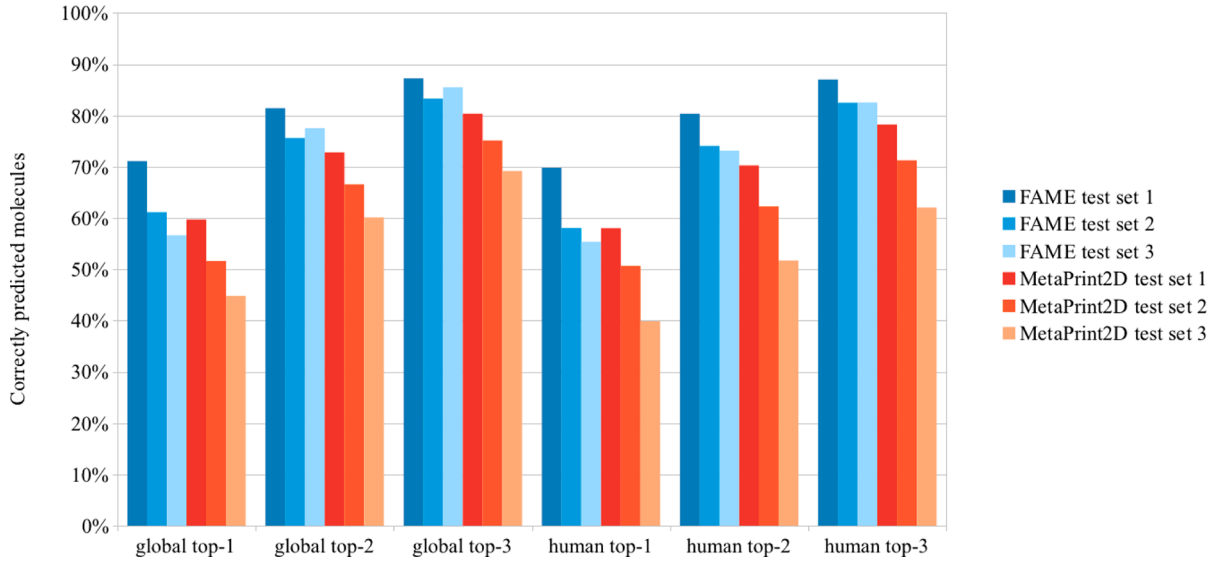


Figure 8. Prediction rates of FAME and MetaPrint2D. Note that the prediction rates for molecules that are structurally more dissimilar from the training data (i.e., test sets 2 and 3) decrease much less with FAME when compared to MetaPrint2D. This indicates a better inter- and extrapolation power of the chemical descriptor-based approach.

Table 7. Coverage of Human Phase I SoMs by FAME and SMARTCyp[a]

	FAME test set 1	SMARTCyp test set 1	FAME test set 2	SMARTCyp test set 2	FAME test set 3	SMARTCyp test set 3
top-1	0.71	0.37	0.61	0.39	0.58	0.40
top-2	0.81	0.50	0.75	0.52	0.79	0.54
top-3	0.87	0.58	0.83	0.60	0.85	0.65

[a]These performance indicators give a rough estimate of the proportion of phase I reactions covered by SMARTCyp. The results cannot be used for drawing conclusions on the performance of the individual predictors since SMARTCyp is run on data that is partly not within its applicability domain (see discussion).

Table 8. FAME Prediction Rates for the Zaretzki Cytochrome P450 Data Set

	FAME on Zaretzki data set	SMARTCyp on Zaretzki data set
top-1	0.50	0.64
top-2	0.73	0.79
top-3	0.83	0.84

Calculation Time. With today's computational power, it is unlikely that computing performance will be prohibitive for the application of a specific SoM or metabolite predictor; however there are some important use cases where fast prediction is essential. For example, metabolism predictors can be used interactively by medicinal chemists to elaborate new strategies for lead optimization in an interactive fashion allowing proposals to be tested on the fly. FAME is very fast. Measured for a batch of 200 druglike molecules calculated on a 2011 MacBook Air with an Intel i7 dual-core CPU, descriptor calculation takes about 2.2 s per molecule (110 ms per heavy atom), and the SoM prediction process itself only takes about 80 ms per molecule.

Software Usage. FAME allows large libraries of molecules (SD file input) to be processed. For correct interpretation, molecular structures must include hydrogens (FAME has an option to add these automatically using a CDK component), with strong acids deprotonated and strong bases protonated. An easy way to prepare molecules for FAME is using the "Add Hydrogens appropriate for pH" option (called with "-p") of Open Babel[30] (free software).

Results are reported in a single CSV file. There is an option available to visualize the results with Jmol.[31] For each heavy atom, the probability of each atom being a SoM is reported, and the three most likely SoMs are highlighted as spheres (Figure 1). There is an option available for generating 3D structures of molecules for the purpose of visualization.

CONCLUSIONS

We have described the development and evaluation of FAME, a site of metabolism (SoM) predictor based on random forest models. FAME aims to overcome many of the limitations of state of the art SoM predictors. The predictor is not constrained to a specific family of metabolic enzymes and supports the prediction of phase I and II metabolism, while performing equally well or better than more specialized predictors (i.e., methods limited to cytochrome P450 reactions,[5] for example). FAME currently offers a global model for metabolism and specific models for human, rat, and dog metabolism. Dedicated models for phase I and phase II reactions are available. The approach relies on only seven descriptors that encode the element type, hybridization state, geometric accessibility, and electronic configuration of each atom of a molecule. This is in striking contrast to many other machine learning methods for metabolism prediction; with RS-Predictor[26] (a support vector machine model) representing the opposite end of the spectrum: RS-predictor uses 540 descriptors in total (148 topological and 392 quantum chemical atom-specific descriptors).

dx.doi.org/10.1021/ci400503s | J. Chem. Inf. Model. 2013, 53, 2896−2907

The implementation of a small number of atom charge descriptors (such as PartialTChargeMMFF94, PartialSigmaCharge) in combination with Sybyl atom types (encoding the element type and hybridization state) and a descriptor for molecular geometry is able to capture a significant part of the underlying recognition and chemistry of biotransformations. This observation is in concordance with earlier reports indicating the high descriptive power of atomic charge and polarizability descriptors in this context.[32] More sophisticated charge descriptors could therefore help to further improve the models.

The data used for training FAME represent a very diverse chemical space and cover marketed drugs, lead molecules, endogenous metabolites, and natural products (see ref 14 for a detailed analysis of the individual chemical domains). Together with improved extrapolation power, this makes FAME applicable to a very broad range of molecules. The methodology of FAME is published in such way that the interested reader will be able to develop new models based on this approach—for example, to extend the application domain to a new chemical space of interest. Predictions work at interactive speed, allowing the user to evaluate changes to the chemical structure of a molecule in real time. Visualization by Jmol facilitates an intuitive interpretation of the results.

While FAME obtains competitive prediction accuracy and many of the limitations of other SoM predictors have been resolved, there is room for further improvement. Certainly, the training data is still a bottleneck, with respect to the (i) coverage of chemical space, (ii) completeness of reported metabolites, (iii) correctness of data, and (iv) the annotation of SoMs (atom mappings vs mechanistic SoMs).

FAME is freely available from the authors to academia and nonprofit organizations.

■ ASSOCIATED CONTENT

Supporting Information

Table reporting on descriptor selection. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rcg28@cam.ac.uk. Phone +44 (1223) 336 432.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.K., M.J.W., J.D.T., A.H., and R.C.G. thank Unilever for funding. P.R. thanks Lhasa Ltd for funding. We thank Dr. Hamse Mussa (Unilever Centre for Molecular Science Informatics), Dr. Guus Duchateau, Leo van Buren, and Prof. Werner Klaffke (Unilever R&D Vlaardingen) for helpful discussion.

■ REFERENCES

(1) Leeson, P. D.; Empfield, J. R. Reducing the risk of drug attrition associated with physicochemical properties. *Annu. Rep. Med. Chem.* **2010**, *45*, 393−407.

(2) Caldwell, G. W.; Yan, Z.; Tang, W.; Dasgupta, M.; Hasting, B. ADME optimization and toxicity assessment in early- and late-phase drug discovery. *Curr. Top. Med. Chem.* **2009**, *9*, 965−980.

(3) Kirchmair, J.; Williamson, M. J.; Tyzack, J. D.; Tan, L.; Bond, P. J.; Bender, A.; Glen, R. C. Computational prediction of metabolism: Sites, products, SAR, P450 enzyme dynamics, and mechanisms. *J. Chem. Inf. Model.* **2012**, *52*, 617−648.

(4) Piechota, P.; Cronin, M. T.; Hewitt, M.; Madden, J. C. Pragmatic approaches to using computational methods to predict xenobiotic metabolism. *J. Chem. Inf. Model.* **2013**, *53*, 1282−1293.

(5) 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.* **2012**, *52*, 2471−2483.

(6) 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.* **2011**, *7*, 3829−3845.

(7) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.

(8) *Accelrys Metabolite Database*, version 2011.2; Accelrys, Inc.: San Diego, CA, 2011.

(9) Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493−500.

(10) Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E. L. Recent developments of the Chemistry Development Kit (CDK) - an open-source java library for chemo- and bioinformatics. *Curr. Pharm. Des.* **2006**, *12*, 2111−2120.

(11) http://www-metaprint2d.ch.cam.ac.uk (accessed Dec 11, 2012).

(12) Adams, S. E. *Molecular similarity and xenobiotic metabolism*. Ph.D. Thesis. University of Cambridge, UK. 2010; http://www.dspace.cam.ac.uk/handle/1810/225225 (accessed Oct 03, 2013).

(13) Krissinel, E. B.; Henrick, K. Common subgraph isomorphism detection by backtracking search. *Software Pract. Exper.* **2004**, *34*, 591−607.

(14) Kirchmair, J.; Howlett, A.; Peironcely, J. E.; Murrell, D. S.; Williamson, M. J.; Adams, S. E.; Hankemeier, T.; van Buren, L.; Duchateau, G.; Klaffke, W.; Glen, R. C. How do metabolites differ from their parent molecules and how are they excreted? *J. Chem. Inf. Model.* **2013**, *53*, 354−367.

(15) *Molecular Operating Environment (MOE)*, version 2011.10; Chemical Computing Group: Montreal, QC, 2011.

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

(17) http://www.tripos.com/mol2/atom_types.html (accessed Aug 20, 2013).

(18) Long, A.; Rydberg, P. Enrichment of true positives from structural alerts through the use of novel atomic fragment based descriptors. *Mol. Inf.* **2013**, *32*, 81−86.

(19) Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA data mining software: An update. *SIGKDD Explor. Newsl.* **2009**, *11*, 10−18.

(20) 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.* **2010**, *1*, 96−100.

(21) Zaretzki, J.; Rydberg, P.; Bergeron, C.; Bennett, K. P.; Olsen, L.; Breneman, C. M. RS-Predictor models augmented with SMARTCyp reactivities: Robust metabolic regioselectivity predictions for nine CYP isozymes. *J. Chem. Inf. Model.* **2012**, *52*, 1637−1659.

(22) Boyer, S.; Arnby, C. H.; Carlsson, L.; Smith, J.; Stein, V.; Glen, R. C. Reaction site mapping of xenobiotic biotransformations. *J. Chem. Inf. Model.* **2007**, *47*, 583−590.

(23) Tiikkainen, P.; Franke, L. Analysis of commercial and public bioactivity databases. *J. Chem. Inf. Model.* **2012**, *52*, 319−326.

(24) Sheridan, R. P.; Korzekwa, K. R.; Torres, R. A.; Walker, M. J. Empirical regioselectivity models for human cytochromes P450 3A4, 2D6, and 2C9. *J. Med. Chem.* **2007**, *50*, 3173−3184.

(25) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947−1958.

(26) Zaretzki, J.; Bergeron, C.; Rydberg, P.; Huang, T. W.; Bennett, K. P.; Breneman, C. M. Rs-predictor: A new tool for predicting sites of

cytochrome P450-mediated metabolism applied to CYP 3A4. *J. Chem. Inf. Model.* **2011**, *51*, 1667−1689.
(27) Tyzack, J. D.; Williamson, M. J.; Torella, R.; Glen, R. C. Prediction of cytochrome p450 xenobiotic metabolism: Tethered docking and reactivity derived from ligand molecular orbital analysis. *J. Chem. Inf. Model.* **2013**, *53*, 1294−1305.
(28) Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321−357.
(29) Rydberg, P.; Gloriam, D. E.; Olsen, L. The SMARTCyp cytochrome P450 metabolism prediction server. *Bioinformatics* **2010**, *26*, 2988−2989.
(30) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33.
(31) *Jmol*, version 13.0; 2011; http://www.jmol.org/.
(32) Kim, D. N.; Cho, K. H.; Oh, W. S.; Lee, C. J.; Lee, S. K.; Jung, J.; No, K. T. EaMEAD: Activation energy prediction of cytochrome P450 mediated metabolism with effective atomic descriptors. *J. Chem. Inf. Model.* **2009**, *49*, 1643−1654.