DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
DMD Fast Forward. Published on March 15, 2006 as doi:10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Title page

Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions

Diansong Zhou, Lovisa Afzelius, Scott W. Grimm, Tommy B. Andersson, Randy J. Zauhar,

Ismael Zamora

Department of Drug Metabolism and Pharmacokinetics, AstraZeneca Pharmaceuticals, Wilmington, DE, USA (D.Z., S.W.G); Department of Chemistry & Biochemistry, The University of Sciences in Philadelphia, Philadelphia, PA, USA (D.Z., R.J.Z.); Department of Drug Metabolism and Pharmacokinetics & Bioanalytical Chemistry, AstraZeneca R&D, Molndal, Sweden (L.A., T.B.A.); Division of Molecular Toxicology, Institute of Environmental Medicine, Karolinska Institutet, Stocholm, Sweden (T.B.A.); Lead Molecular Design, S.L. Valles 96-102 (local27) 08190 Sant Cugat del valles, Barcelona, Spain (I.Z.); and GRIB-IMIM, Dr Aiguader, 80 08003 Barcelona, Spain (I.Z.)

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

Copyright 2006 by the American Society for Pharmacology and Experimental Therapeutics.

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Running title

Running title: Metabolic Sites Prediction of CYP3A4 substrates

Corresponding author:
Diansong Zhou
Department of Drug Metabolism and Pharmacokinetics
AstraZeneca Pharmaceuticals
1800 Concord Pike
Wilmington, DE, USA 19810
Tel 1-302-8862007
Fax 1-302-8865345
Email Diansong.zhou@AstraZeneca.com

Number of text pages: 24 (excluding references, figures and tables)
Number of Tables: 2
Number of Figures: 6
Number of References: 24

Number of words in Abstract: 225
Number of words in Introduction: 712
Number of words in Discussion: 1114

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Abstract

Predictions of the metabolic sites for new chemical entities, synthesized or only virtual, are important in the early phase of drug discovery to guide chemistry efforts in the synthesis of new compounds with reduced metabolic lability. This information can now be obtained from *in silico* predictions and therefore a thorough and unbiased evaluation of the computational techniques available is needed. Several computational methods to predict the metabolic hot spots are emerging. In this study, metabolite identification using MetaSite® and a docking methodology, GLUE, were compared. Moreover, the published CYP3A4 crystal structure and computed CYP3A4 homology models were compared for their usefulness in predicting metabolic sites. A total of 227 known CYP3A4 substrates reported to have one or more metabolites adding up to 325 metabolic pathways were analyzed. Distance-based fingerprints and 4-point pharmacophore derived from GRID molecular interaction fields were used to characterize the substrate and protein in MetaSite® and the docking methodology, respectively. The CYP3A4 crystal structure and homology model with the reactivity factor-enabled achieved a similar prediction success (78%) using the MetaSite® method. The docking method had a relatively lower prediction success (~57% for the homology model), although it still may provide useful insights for interactions between ligand and protein, especially for uncommon reactions. The MetaSite® methodology is automated, rapid, and has relatively accurate predictions compared to the docking methodology used in this study.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Introduction

The cytochrome P450s (CYP450s) belong to a superfamily of heme-containing enzymes that metabolize a wide range of therapeutic agents. CYP3A4 is the most abundant human hepatic CYP450 isoform and is responsible for the metabolism of about 50% of known drugs with wide structural diversity (Guengerich, 1999). From a drug development perspective, a new drug candidate that is susceptible to CYP3A4 metabolism may not be able to reach the target at an effective concentration or may be involved in drug-drug interactions when co-administered with other CYP3A4 substrates or inhibitors. *In silico* prediction of the primary metabolic site from the molecular structure may assist in the identification of the metabolite formed through the metabolic reaction, or could be used to guide the synthesis of metabolically stable compounds or direct the metabolism to certain CYP450 enzymes. Experimental studies at early drug discovery stage on metabolic reactions of new chemical entities can usually allocate the metabolic site(s) to a certain fragment of the molecule. In order to guide chemists to synthesize new compounds with improved metabolic properties, information on the exact metabolic site is preferred. This information may be obtained from *in silico* predictions and therefore we need a thorough and unbiased evaluation of the computational techniques available.

Mainly four approaches are considered in literature to address the prediction of the metabolic sites:

(a) Molecular orbital calculations (De Groot et al. 1999; Singh et al, 2003) can be used to estimate the energy necessary to abstract a hydrogen atom from the substrate. The energy for each possible metabolic site in combination with surface area exposure of the hydrogen atom

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

was used to rank the different sites for the CYP3A4 metabolic liability (Singh et al, 2003). However, this approach does not consider the specific ligand-enzyme interactions, therefore overlooks the specificity of CYP450 enzymes. Hydrogen abstraction is the only metabolic mechanism considered in this method.

(b) Molecular docking in combination with scoring functions has been applied to predict the metabolic sites for CYP2D6 substrates, like codeine and MPTP (Kirton et al., 2002). Though this method theoretically optimizes favorable interactions between ligand and enzyme to estimate the binding mode, predictions heavily rely on a well-defined pharmacophore of the active site. Therefore, the availability of a validated crystal structure or homology model of the CYP450 enzyme is a prerequisite for reliable predictions. Moreover, there are other factors like protein flexibility and water molecule interactions, which could make these methods difficult to use.

(c) Data mining. Recently, a probabilistic scoring method based on fragment analysis of the substrate structures corresponding to the fragments observed in databases has also yielded useful results in prediction of metabolic site (Hasselgren−Arnby et al., 2005). In this case no guidance on major metabolic pathways involved are needed, but this also limits interpretation.

(d) GRID Molecular Interaction Fields (MIFs) (Goodford, 1985) has been used to derive parameters that describe the ligand-protein interactions and this methodology has been implemented in MetaSite®. In MetaSite®, the active site of the protein is characterized using MIF interactions with different probes and a set of distance-based descriptors is calculated by taking the oxygen atom in the heme as reference. The different atoms of the substrate are ranked in agreement to the similarity to the protein and the top ranked positions are selected as the possible

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

metabolic sites (Zamora et al., 2003; Cruciani et al, 2005). A well-defined active site is also essential for this methodology. However, when this methodology was first developed, the human CYP450 crystal structures were not publicly available and predictions were derived using homology models based on alignments of bacterial CYP450 structural models. The recent available wild type CYP3A4 crystal structures (Williams et al., 2004; Yano et al. 2004) provide the opportunity to improve this methodology.

The aim of the present work was to analyze the use of the available CYP3A4 crystal structures in the prediction of metabolic sites of structurally diverse drugs with known metabolic profiles using both the docking/scoring and the MetaSite® methodologies, and comparing the results with the ones obtained from CYP450 homology models. It is not within the scope of this study to predict the metabolic rate, or to differentiate the major or minor metabolic pathways. Other information such as empirically determined data (Km or/and Vmax) would probably be needed to develop such a model.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Material and Methods

Equipment, software and Database

All calculations were done on the Linux (Red Hat 8.0) operating system on a 1.8 GHz Pentium IV. The programs used in the molecular modeling were GRID, GLUE, PENGUINS, MetaSite® (Molecular Discovery Ltd., Middlesex, U.K.), Golpe (Multivariate Infometric Analysis S.r.l., Perugia, ITALY), and CONFORT (Tripos Inc., St. Louis, MO, USA). Chemical structures were imported from the ISIS-BASE database or drawn using ISIS-Draw (MDL information Systems Inc., San Leandro, CA, USA).

Proteins and substrates

A CYP3A4 crystal structure with a resolution of 2.8 Å was used in this study (PDB 1W0E). This is a wild type enzyme, except that the N-terminal membrane insertion peptide has been removed to increase solubility for crystallization. There was no substrate or inhibitor bound in the active site of this crystal structure. A CYP3A4 homology model was generated by comparative modelling based on multiple bacterial CYP450s (PDB codes are 2bmh, 3cpp, 1cpt and 1oxa) using Modeller (De Rienzo et al, 2000).

A total of 227 CYP3A4 substrates were collected from the literature (Rendic and Di Carlo, 1997) and MDL metabolite database. Most of the well-known CYP3A4 substrates, such as midazolam, nifedipine and testosterone, were included in this data set. The substrates are reported to have one or more CYP3A4 mediated metabolites adding up to 325 metabolic pathways. Most of the compounds (165) have only one metabolite catalyzed by CYP3A4, whereas 62 compounds have two or more metabolites. The 2D chemical structures of the 227 compounds were drawn or

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

imported from several databases and converted to 3D structures using PENGUINS. A maximum of 50 diverse conformers within an energy window of 10 kcal/mol were generated for each substrate using CONFORT.

Principal Component Analysis (PCA)

PCA (M Pastor, et al.) was used to compare the active sites of the CYP3A4 crystal structure, the homology model based on 4 cytochrome bacterial structures and the 4 proteins used to build the homology model. The PDB codes for these bacterial crystal structures are 2bmh, 3cpp, 1cpt and 1oxa for $CYP450_{BM3}$, $CYP450_{cam}$, $CYP450_{terp}$ and $CYP450_{eryF}$ respectively. The molecular interaction fields generated in GRID force field were used to characterize the active sites of all proteins. A GRID box (25x25x25 Å) was used to include the superimposed active sites of all 3D structures. In our model, the heme was considered as part of the protein and was not treated differently except where the ferryl oxygen was defined as a dummy atom that is unable to form hydrogen bonds with any probe. Similar treatment was applied in the later calculations. The active site of each structure was characterized using 10 GRID probes (DRY, C_3, $N1^+$, N1, NH=, N:, NM_3, O, O^- and OH) in flexible mode (MOVE=1) with a grid step size of 1 Å. Hydrophobic interactions were calculated with the DRY probe, while the steric interactions were calculated with C_3 and NM_3 probes. $N1^+$ and O^- probes are charged and N1, N:, NH=, O and OH probes are polar.

GRID force field calculations were then imported into GOLPE and the following pre-treatment was done before PCA analysis: (1) The maximum cutoff was set to zero to consider only the favorable interactions (negative energy values); (2) Block unscaled weights (BUW) were used to

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

normalize the interaction energies between the different probes; (3) Variables with values smaller than 0.01 kcal/mol and those with a standard deviation below 0.02 kcal/mol were removed in order to increase signal to noise ratio. PCA analysis provided loading and score plots with the insight into the topology and chemical identity difference between the structures.

MetaSite® methodology using GRID descriptors

MetaSite® was first developed and applied to CYP2C9 and its substrates (Zamora et al., 2003). This methodology is improved by the inclusion of a reactivity factor in the current study and used to characterize both the CYP3A4 crystal structure and the homology model. Both the protein active site and the ligand are represented by selected distance-based descriptors using the molecular interaction fields computed by GRID. The best match between active site and metabolic site is chosen based on similarity and optionally with atom reactivity (Figure 1).

(1) Protein Treatment. The Molecular Interaction Fields (MIFs) in the active site of CYP3A4 were generated using four probe types: hydrophobicity (DRY probe), hydrogen-bond donor (amide nitrogen N1 probe), hydrogen-bond acceptor (carbonyl oxygen O probe) and electrostatic property (positively charged N1+, N2+, N3+ probes and negatively charged O-, COO-, N-: probes) with a grid step size of 1 Å. A grid box of 25x25x25 Å (active site) was defined and heme was located at the bottom of the box.

The MIFs were generated using either flexible (MOVE=1) or rigid mode (MOVE=0) in GRID (Braiuca et al., 2004). In the rigid mode, the structure of the protein is considered as fixed and the atomic coordinates from the protein structure are used directly in the interaction energy calculation. In the flexible mode, the side chains of amino acids in the active site are allowed to

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

react to the presence of the probe and position themselves at the most energetically favourable distance from the probe. In this way, the flexible GRID fields can accommodate different substrates, based on their shape, size and interactions. Nevertheless, the protein backbone is not allowed to move in either case, and therefore, the flexible mode could not be considered as describing the overall protein dynamics.

Twenty-nine crystallographic water molecules have been determined in the CYP3A4 crystal structure used in this study. In the process of CYP450 mediated metabolism, one water molecule is usually generated (Guengerich, 1999). Water molecules may also play an important role by forming certain hydrogen bonds to hold the substrate with proper orientation towards the heme (Wester et al., 2003). Therefore, CYP3A4 crystal structures with or without these crystallographic water molecules were considered separately. Overall, six CYP3A4 protein structures or models were explored in this study: a) homology model in GRID rigid mode, b) homology model in GRID flexible mode, c) crystal structure with water in GRID rigid mode, d) crystal structure with water in GRID flexible mode, e) crystal structure without water in GRID rigid mode and f) crystal structure without water in GRID flexible mode.

The following MIF treatment was similar to the one previously published (Zamora et al., 2003). Briefly, the regions close to the binding site, but not accessible to the substrates, were removed from the analysis in an automatic cut out process. Finally the distances were calculated between the selected GRID points and the fixed ferryl oxygen at the reactive center of the enzyme. The distances were grouped at a resolution of 0.8 Å and plotted as correlograms, which were compared to the distance-based descriptor of the substrate.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

(2) Substrate Treatment. The compounds were built or imported as 2D structures. Substrate conformation sampling is critical in order to simulate flexible interaction between the substrate and CYP3A4 enzyme; therefore, conformation search followed by energy minimization was performed in CONFORT for each CYP3A4 substrate. The atoms of a CYP3A4 substrate were then classified into four categories according to their hydrophobic, hydrogen bond donor, hydrogen bond acceptor and electrostatic interaction capabilities using the Tripos force-field atom type definitions. The distances between the possible metabolic sites, such as hydrogen or nitrogen atoms, and the different pre-classified atoms were computed and transformed into grouped variables with a resolution of 0.8 Å. Four sets of fingerprints for each possible metabolic site in a substrate are generated.

(3) Substrate-Protein Comparison. Once the protein interaction profiles were transformed into distances from the reactive center of the enzyme to the interaction points in the protein and the structure of substrate was described as a distance based fingerprint for each possible metabolic site, both sets of descriptors were compared using Carbó similarity index (Amat L and Carbó-Dorca R, 1999). Four similarity indexes were obtained for each possible metabolic site in a substrate according to hydrophobic interactions, hydrogen bond donor/acceptor interactions and charge-charge interactions. The atomic position with the highest similarity score will be the one that has the best complementary with the protein and theoretically the enzyme will orient the compound with this atom towards the heme of CYP3A4.

(4) Reactivity. In addition to the similarity comparison, a substrate fragment recognition factor called ‘reactivity’ has been implemented in this methodology. A database of different small

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

fragments with pre-calculated reactivity values has been generated and applied in MetaSite®. Each fragment was considered as a participant in oxidative reactions and a reactivity value was assigned to each atom regarding the liability towards the oxidative reaction. When a fragment in the molecule under study is recognized as one in this database, all atoms in the fragment are assigned to that reactivity value.

The final ranking for potential metabolic site is the product of protein effect (computed on the basis of the similarity analysis) and atomic reactivity effect (computed using the fragment based approach).

Docking/Scoring Methodology

All the substrates were docked into the active site of CYP3A4 using the homology model and the available crystal structure (PDB 1W0E), with and without crystallographic water molecules. GLUE, a GRID based docking program, was used to analyze the ligand-receptor interaction and to perform the docking experiments.

In GLUE, the active sites were mapped using hydrophobic, hydrogen-bond donor/acceptor and electrostatic probes. All possible tetrahedra obtained from 4 minima energy points from GRID are computed. These 4-point pharmacophores derived from the interaction for the active sites were then used as templates to compare with the ligand.

Similarly, GLUE identified the polar and hydrophobic heavy atoms of the ligand and calculated all possible tetrahedra between these atoms. The atomic positions of the different conformers for each potential substrate were compared to the pharmacophores based on the hydrophobic, hydrogen-bond donor/acceptor and electrostatic interaction capabilities and geometry. When a

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

pharmacophore was recognized, the ligand was aligned in the enzyme cavity and an energy computation followed. If there were any conflict contacts between the ligand and the protein, an induced fit process was started to accommodate the substrate in the protein cavity. The same process was repeated for all possible 4-point pharmacophore templates.

The metabolic potential of each atom in the substrate was estimated using (1) a probability function based on the distance between each atom and the reactive center of the protein and (2) a probability function based on the energy of interaction as computed by GLUE. Distance based probability was calculated considering a Gaussian distribution for the difference in the distance between each substrate atom and the fixed ferryl oxygen (2 Å above heme) at the active center of the enzyme, with the optimal distance found in some crystallographic structures, 2.6 Å (Schlichting et al, 2000).

The energy probability was calculated based on the Boltzman distribution for each docking solution. The product of distance probability for each atom in each docking solution and energy probability for each docking solution yield a value for each atom that was used to rank the probable metabolic site. The predicted metabolic site was identified using the ranking position for each atom evaluated from the best docking conformer (lowest energy) and from an ensemble of all docking solutions.

Results

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Principle Component Analysis (PCA)

PCA was used to evaluate the topological difference of the active sites of the CYP3A4 crystal structure, homology model and four bacterial crystal structures. Crystal structures with water molecules were excluded for this analysis as the corresponding active sites were smaller compared to the other models and would skew the results. The PCA scores plot is shown in Figure 2 and the relative distance between individual data points along each principal components is an indication of the extent of the difference between the active sites of the CYP3A4 crystal structure, the homology model and the four bacterial structures. There are four groups of these models: CYP3A4 crystal structure (group 1); CYP3A4 homology model and CYP_{terp} (group 2); CYP_{cam} and CYP_{eryF} (group 3); and CYP_{BM3} (group 4). The CYP3A4 crystal structure differs from the other structures analyzed as indicated by its PC scores in the scores plot. The calculated root-mean-square deviation (RMSD) between the CYP3A4 crystal structure and the homology model is 5.52 Å, which also reflect the dissimilarity between these two structures. CYP_{BM3} was separated from CYP_{cam} and CYP_{eryF} by the second principal component. The active site of the CYP3A4 homology model is more similar to that of the CYP_{terp} structure, and also more similar to the active site of CYP3A4 crystal structure than to the other three bacterial forms (CYP_{BM3}, CYP_{cam} and CYP_{eryF}), as indicated by their proximity along the PC1 score in the scores plot.

MetaSite® methodology using GRID descriptors

The validation of the prediction of the metabolic site was based on counting the number of experimentally reported metabolic pathways that were found among the first, second and third

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

sites as ranked by the methodology used. A total of 227 reported CYP3A4 substrates with 325 metabolic pathways were studied, 165 compounds had only one reported metabolite, while 62 compounds that had multiple metabolic sites. When considering the multiple-metabolite substrates, two schemes were used to evaluate the prediction methods: (a) each reaction was considered independently, so that 325 pathways were considered as 325 different substrates. (b) Substrates with multiple metabolic sites were analyzed separately and the substrate was classified as well predicted if at least one pathway was among the first, second or third-ranked site proposed by this methodology. Since there is no data available about the rate of formation for each of the metabolite reported, the prediction for one of them was enough to classify the compound.

The predictions for all 325 metabolic pathways using MetaSite® methodology and different protein models are shown in Figure 3 and the prediction success using either the top-ranked site or the top three sites for the different models are presented in Table 1. In the cases of considering both protein similarity and atomic reactivity, the crystal structure with crystallographic water molecules model yielded the lowest prediction success of 58 % when using the top three ranking positions and when applied to predict 325 metabolic pathways, while all the other models had similar predictive success, an average of 70%. Of these 29 crystallographic water molecules, 15 waters were inside the defined grid box, but only 5 water molecules can be considered as being inside the binding site. Using the crystal structure with these five water molecules included always led to the least successful performance and will not

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

be considered when computing average prediction success. When the reactivity factor was not considered the average prediction success decreased from 70% to 39%.

The MetaSite® methodology yielded an average prediction success of 75% when using the first 3 ranked positions and when applied to the 165 substrates that had only one metabolite reported, (and where both protein complentarity and reactivity were considered). This methodology also achieved an average prediction success of 86% when at least one metabolic site was well predicted by one of the first 3 ranked positions, and applied to the 62 substrates with multiple sites. The overall prediction success for all compounds when at least one metabolic site was predicted among the top three ranked sites by this methodology was 78 % when reactivity was enabled.

For the 325 reactions under study, 43% (139 reactions) were aliphatic or aromatic hydroxylation, 27% (89 reactions) were N-dealkylation and 7% (22 reactions) were O-dealkylation. These are three major metabolic pathways for CYP3A4 substrates and they were well predicted using the flexible mode for the homology model or crystal structure with reactivity option enabled (Figure 4A). Other reaction types such as reductions (6 reactions), epoxidations (5 reactions), N-hydroxylations (2 reactions) were not well predicted. This could be due to the fact that these reactions are uncommon; therefore the reactivity factor disfavours these kinds of reactions. In the model with crystal structure, flexible GRID mode and atomic reactivity option enabled, the method predicted correctly (among the top three ranked sites): 70 out of 98 aliphatic hydroxylations (71%), 23 out of 41 aromatic hydroxylations (56%), 76% of N-dealkylation and 100% of O-dealkylation.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

The prediction success corresponding to various types of reactions for the 165 substrates with only one metabolic site by the homology and crystal models in flexible mode and with reactivity option enabled are shown in Figure 5. The prediction success for 57 hydroxylated substrates, 53 N-dealkylated substrates, and 12 O-dealkylated substrates were 75%, 85% and 100%, respectively by using the homology model.

Sixty-two multi-metabolite substrates have 160 metabolic pathways, including 82 hydroxylations, 36 N-dealkylations, and 10 O-dealkylations. The prediction success using the crystal structure model were 60%, 67% and 100% for hydroxylation, N-dealkylation and O-dealkylation, respectively.

The rigid and flexible GRID modes were evaluated for all substrates. The reactions that were accurately predicted by the CYP3A4 crystal structure without added water molecules in rigid or flexible modes and with reactivity enabled were compared, and no significant difference was observed. The CYP3A4 crystal structure in its rigid mode and with reactivity enabled could correctly predict the metabolic sites of 183 substrates while in its flexible mode, 173 out of a total of 227 substrates were well predicted. The calculated active sites of CYP3A4 crystal structure and the homology model are presented in Figure 6. Little size difference in the active sites of homology model in either flexible or rigid modes was observed, while the size of the active site in the crystal structure in flexible mode is much larger than that seen in its rigid mode. The apparent primary reason for this was because the hydrogen bonding between Glu308 and Arg212, which blocked the extension of the active site in rigid mode for the crystal structure. However, the size of the active site region in the vicinity of the heme was similar for both

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

flexible and rigid modes. This also might be the reason why the use of the crystal structure in either the flexible or rigid mode had no clear impact on the accuracy of predictions.

Generally, when the atomic reactivity ranking was not considered, the prediction success decreased for all models. In order to evaluate the importance of reactivity in the methodology for the metabolic site prediction, the prediction success of the model constructed using the crystal CYP3A4 structure without water in GRID flexible mode were compared either with reactivity enabled or not. When reactivity was enabled 173 out of 227 substrates were well predicted compared with 103 of 227 when reactivity was disabled. Seventeen substrates were only well predicted in the model without the reactivity factor used, and 6 of these 17 substrates are metabolized through uncommon reactions, such as epoxidation, reduction, or dehydrogenation. The model without reactivity tends to have better prediction for uncommon reactions and reactivity factor may disfavour the rank of uncommon reactions in this method. However, the uncommon reactions are only a small portion of all reactions mediated by CYP3A4, and the models with reactivity factor always yield better overall prediction success than the models without reactivity.

Docking/Scoring approach

The first observation when analyzing the docking solutions is that not all substrates could be docked into the CYP3A4 active site. This also depended on the protein model used. Two out of 227 substrates that could not be docked into the CYP3A4 homology model, while 18 substrates could not be docked into the CYP3A4 crystal structure without the crystallographic water molecules, and 78 could not be docked into the CYP3A4 crystal structure when water molecules

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

were included. Therefore, only 323, 298 and 205 metabolic pathways were evaluated for homology, crystal structure without water, and crystal structure with water, respectively.

A number of metabolic pathways can be well predicted using this docking approach. For example, the 4-hydroxylation is the major pathway for CYP3A4-mediated metabolism of alprazolam whereas 1'-hydroxylation is relatively minor (Williams et al., 2002). The lowest energy docking result showed that distance between the C4 position of alprazolam and the ferryl oxygen was 3.35 Å whereas the distance between the C1' position and oxygen was 4.47 Å, and both metabolites are possible.

Multiple docking solutions were available for most of the CYP3A4 substrates. Similar to the MetaSite® procedure, the evaluation of the docking results was based on the number of solutions that predicted correctly the metabolic site in the 1st, 2nd and 3rd ranked positions.

Two kinds of analysis were performed depending upon whether the best docking solution or all docking solutions included. In the case of considering the best-docking (lowest energy) solution for each substrate, only 18% to 27 % of the metabolic reactions were correctly predicted as possible metabolic sites among the first three selections (Table 2). When all docking solutions were analyzed, it is considered a successful prediction if at least one of the docking solutions exhibited correct orientation. In this later case, homology model achieved the best prediction (57%), whereas the docking based on the crystal structure without water molecules yielded 47% prediction success and the structure with water molecules gave the lowest success (27%). Again, the crystal structure with water molecule model won't be considered in the following analysis.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

When the substrates that only had one metabolite were analyzed, the prediction success were 63% and 53% for the CYP3A4 homology model and crystal structure when all docking solutions was considered. In the case of multiple metabolites substrates, the prediction success increased to 82% and 74% respectively when at least one metabolite was correctly predicted.

The prediction success corresponding to the different reaction types found by docking into the CYP3A4 homology model and the crystal structure are also presented in Figure 4B. Three major metabolic pathways were analyzed: hydroxylation yielded a prediction success of 52% for the crystal structure based model and 58% for the homology based one, N-dealkylation reactions gave a prediction success of 38% and 54% for the crystal and homology based docking respectively and finally the O-dealkylation reactions yield 62% for the crystal structure docking and 73% for the homology docking. Some reactions such as reductions (2 of 3 reactions), epoxidations (3 of 3 reactions), N-hydroxylations (1 of 2 reactions) which were not well predicted with MetaSite® methodology could be reasonably predicted with the docking approach using the CYP3A4 homology model.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Discussion

CYP3A4 plays a major role in drug metabolism in humans and understanding the CYP3A4-mediated metabolism is especially relevant in drug discovery. Two computational techniques, the MetaSite® and docking methodologies, for the prediction of the metabolic sites have been successfully applied and evaluated for a set of CYP3A4-mediated substrates using both the CYP3A4 crystal structure and a homology model. The predicted metabolic sites using the docking methodology in conjunction with the CYP3A4 homology model matched, in 69% of the cases, with the experimentally reported metabolic sites. Both the CYP3A4 crystal structures and the homology models achieved similar prediction success by using the MetaSite® method, i.e., the predicted metabolic sites agreed with experimental results in 78% of the cases. The MetaSite® methodology correctly predicted 1' and 4-hydroxylmidazolam, 6β-hydroxytestosterone and dehydronifedipine in its top three rankings, while the docking methodology also identified metabolic sites of midazolam and testosterone but failed to predict dehydronifedipine. Additionally, the MetaSite® methodology can be considered as an automated method after the active site properties have been mapped. Another disadvantage of the docking method is that some compounds could not be docked into the enzyme active site. This might be because of deficiencies of the docking algorithm or the flexibility of the ligands. Overall, the MetaSite® methodology can be used as a metabolic site prediction tool at the early drug discovery stage in terms of both speed and accuracy.

One aspect to be considered in the evaluation of any method for predicting the metabolic site is the approach used to establish an objective criterion to measure its predictive power. In this

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

study, the number of metabolic reactions that are well predicted considering the 1st, 2nd and 3rd ranked atomic positions was used as a measurement of the predictive capabilities of the method. Nevertheless, this prediction success will depend on the number of metabolic pathways reported for one substrate. When each reaction is considered independently of the substrate, the ranking position could underestimate the prediction power. For example, a substrate with 4 metabolic sites would always have one site that will be misclassified since we considered only the first 3 ranked positions. Therefore, the first three ranked positions for each reaction independent of substrate and for at least one reaction reported for a substrate were both used as a predictive measurement in this study. The first measurement represents the lower predictive power limit and the second one gives the higher predictive power limit of each methodology.

Both methodologies in this study evaluate the interactions between protein and ligands; therefore, the protein model is critical in determining the quality of predictions. The wild type CYP3A4 crystal structure (PDB code: 1W0E) without bound ligand used in this study is very similar to a CYP3A4 crystal structure published by another group (PDB code: 1TQN). Both structures have a cluster of phenylalanine residues that lies above the active site, which makes a relatively small active site for CYP3A4. CYP3A4 sometimes displays cooperative behavior with the binding of substrates (Domanski et al., 2001; Galetin et al., 2003; He et al., 2003; Tang and Stearns, 2001), and this is generally rationalized by a flexible and large CYP3A4 active site, which can accommodate multiple substrates. However, published CYP3A4 crystal structures, in contrast to CYP2B6 (Scott et al., 2004), have little conformational change in the ligand-free and ligand-bound forms. One possible reason might be that the ligand in the crystal is relatively small

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

and does not cause dramatic conformational change of CYP3A4. Large ligands when used in crystallization either stay outside of the active site pocket (progesterone) or have yet to be co-crystallized with the protein (erythromycin) (Williams et al., 2004; Yano et al. 2004). All information suggests that the current CYP3A4 crystal structure might be only one of many available conformations for the enzyme. This might be the reason that the CYP3A4 crystal structure did not exhibit any advantage over the homology model in predicting the metabolic site. With the MetaSite® methodology, this could also be related to the protein treatment where the interaction profile was compressed to a distance based descriptors and the impact of the protein/substrate complementarity decreased. As the docking technique depends much more on the protein structure than MetaSite® methodology, the difference of the prediction success from docking between the homology model and the crystal structure were larger (~10% in the prediction using top three ranking positions).

In an effort to model the flexibility of CYP3A4, we tried to take advantage of a special mode of GRID field calculation (MOVE =1), also termed the flexible mode in this study. Theoretically, the protein is more flexible, and can accommodate different substrates, therefore, flexible mode model could provide more accurate predictions. However, in contrast to the big impact on the prediction success for CYP2C9 substrates (Zamora et al., 2003), there is no clear effect of this mode for the CYP3A4 substrates. One possible reason might be that the GRID flexible mode can only capture the side chain movement of certain amino acids, but not the major conformational changes, such as helix movement. More extensive calculations, perhaps involving backbone

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

motion, might be required to simulate the flexibility of the protein and the interaction between CYP3A4 and the ligands.

The water molecules can play a significant role in ligand-protein bindings (Wester et al., 2003), but it is still challenging to include waters during the automated docking process. The docking studies have shown that including the fixed water molecules can either increase or decrease the docking accuracy (Osterberg et al., 2002; De Graaf et al., 2005). Even if only 5 water molecules are inside the defined binding site, we have shown that the models including these fixed waters generated the lowest prediction success in both methodologies. Water molecules may mediate the substrate-enzyme interaction differently compared to the ones obtained from the crystal structures. The positions of water molecules might depend on the specific substrate in the active site. The use of fixed positions for the water molecules obtained from a crystal structure would therefore not represent all the possibilities. As in this study, it is not practical to specify the water molecule position(s) in the active site for each individual substrate when attempting to examine a large number of compounds. The structures excluding these waters would thus be better models in the metabolic sites prediction.

In summary, this study has presented the docking and the MetaSite® methodologies to successfully predict the metabolic site for CYP3A4 substrates under various conditions. Each method has its own advantages, and with proper application alone or in combination with each other, these methods should be of great help in identifying the metabolic sites, elucidating metabolite structures and guiding chemical programs to synthesis compounds with improved metabolic properties.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

References

Amat L and Carbó-Dorca R (1999) Fitted electronic density functions from H to Rn for use in quantum similarity measures: cis-diamminedichloroplatinum II complex as an application example. *J Comput Chem* **20**:911-920.

Braiuca P, Cruciani G, Ebert C, Gardossi L, and Linda P (2004) An innovative application of the "flexible" GRID/PCA computational method: study of differences in selectivity between PGAs from *Escherichia coli* and a *Providentia rettgeri* mutant. *Biotechnol Prog* **20**:1025-1031.

Cruciani G, Carosati E, De Boeck B, Ethirajulu K, Mackie C, Howe T and Vianello R (2005) MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist. *J Med Chem* 48: 6970-6979.

De Graaf C, Pospisil P, Pos W, Folkers G, Vermeulen NP (2005) Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. *J Med Chem* **48**:2308-18.

De Groot MJ, Ackland MJ, Horne VA, Alexander AA and Barry CJ (1999) A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite prediction using a combined protein and pharmacophore model for CYP2D6. *J Med Chem* **42**:4062-4070.

De Rienzo F, Fanelli F, Menziani MC and De Benedetti PG (2000) Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4. *J Comput-Aided Mol Des* **14**:93-116.

Domanski TL, He YA, Khan KK, Roussel F, Wang Q, Halpert JR (2001) Phenylalanine and tryptophan scanning mutagenesis of CYP3A4 substrate recognition site residues and effect on substrate oxidation and cooperativity. *Biochemistry* **40**: 10150-10160.

Galetin A, Clarke SE, Houston JB (2003) Multisite kinetic analysis of interactions between prototypical CYP3A4 subgroup substrates: midazolam, testosterone, and nifedipine. *Drug Metab Dispos* **31**:1108-1116.

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

Guengerich FP (1999) Cytochrome P-450 3A4: Regulation and role in drug metabolism. *Annu Rev Pharmacol Toxicol* **39**:1–17.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Hasselgren−Arnby C, Smith J, Glen RC and Boyer S (2005) SPORCalc − Fingerprint based probabilistic scoring of metabolic sites. *The 7th International Conference on Chemical Structures*.

He YA, Roussel F and Halpert JR (2003) Analysis of homotropic and heterotropic cooperativity of diazepam oxidation by CYP3A4 using site-directed mutagenesis and kinetic modeling. *Arch Biochem Biophys* **409**: 92-101.

Kirton SB, Kemp CA, Tomkinson NP, St.-Gallay S and Sutcliffe MJ (2002) Impact of incorporating the 2C5 crystal structure into comparative models of Cytochrome P450 2D6. *Proteins Struct Funct Genet* **49**:216–231.

Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **46**:34-40.

Rendic S and Di Carlo FJ (1997) Human Cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers and inhibitors. *Drug Metab Rev* **29**:413-580.

Schlichting I, Berendzen J, Chu K, Stock AM, Maves SA, Benson DE, Sweet RM, Ringe D, Petsko GA and Sligar SG (2000) The Catalytic pathway of cytochrome P450cam at atomic resolution. *Science* **287**: 1615-1622.

Scott EE, White MA, He YA, Johnson EF, Stout CD and Halpert JR (2004) Structure of Mammalian Cytochrome P450 2B4 Complexed with 4-(4-Chlorophenyl)imidazole at 1.9-Å Resolution. *J Biol Chem* **279**: 27294–27301.

Singh SB, Shen LQ, Walker MJ and Sheridan RP (2003) A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. *J Med Chem* **46**:1330-1336.

Tang W, and Stearns RA (2001) Heterotropic cooperativity of cytochrome P450 3A4 and potential drug-drug interactions. *Curr Drug Metab* **2**:185-198.

Wester MR, Johnson EF, Marques-Soares C, Dijols S, Dansette PM, Mansuy D and Stout CD (2003) Structure of Mammalian Cytochrome P450 2C5 Complexed with Diclofenac at 2.1 Å Resolution: Evidence for an Induced Fit Model of Substrate Binding. *Biochemistry* **42:** 9335-9345.

Williams JA, Ring BJ, Varon EC, Jones DR, Eckstein J, Ruterbories K, Hamman MA, Hall SD and Wrighton SA (2002) Comparative metabolic capabilities of CYP3A4, CYP3A5, and CYP3A7. *Drug Metab Dispos* **30**:883–891.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Williams PA, Cosme J, Vinkovic DM, Ward A, Angove HC, Day PJ, Vonrhein C, Tickle IJ and Jhoti1 H (2004) Crystal structures of human Cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **305**:683-386.

Yano JK, Wester MR, Schoch GA, Griffin KJ, Stout CD and Johnson EF (2004) The structure of human microsomal Cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution. *J Biol Chem* **37**:38091–38094.

Zamora I, Afzelius L and Cruciani G (2003) Predicting drug metabolism: A site of metabolism prediction tool applied to the Cytochrome P450 2C9. *J Med Chem* **46**:2313-2324.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Legends for figures

Figure 1. Interface of MetaSite® program that explains the procedure of MetaSite® methodology.

Figure 2. GOLPE-generated PCA scores plot to illustrate the difference between the CYP3A4 crystal structure, the homology model, and four bacterial crystal structures (CYP_{terp}, $CYP_{cam,}$ CYP_{eryF} and CYP_{BM3}).

Figure 3. Atom ranking for all CYP3A4 substrates, one-metabolite substrates, or multiple-metabolite substrates in all models evaluated using MetaSite® methodology, homology model in rigid mode, homology model in flexible mode, crystal structure with water in rigid mode, crystal structure with water in flexible mode, crystal structure without water in rigid mode and crystal structure without water in flexible mode. Rank orders in all models with reactivity factor enabled or disabled are presented separately.

Figure 4. (A) The prediction success corresponding to various types of reactions for all 325 CYP3A4 mediated metabolic pathways by the homology model (grey bar) and the crystal structure (white bar) in flexible mode and reactivity factor enabled using MetaSite® methodology. The number in the parentheses is the number of each reaction evaluated in this study. (B) The prediction success corresponding to various types of reactions for all CYP3A4 mediated metabolic pathways by the homology model (grey bar) and the crystal structure (white bar) using docking methodology.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 5. The prediction success corresponding to various types of reactions for 165 one-metabolite CYP3A4 substrates by the homology model (grey bar) and the crystal structure (white bar) in flexible mode and reactivity factor enabled using MetaSite® methodology.

Figure 6. Calculated active site illustrating difference in shape and volume of the active sites of the CYP3A4 homology model and the CYP3A4 crystal structure: (A) the homology model in flexible mode (B) the crystal structure in flexible mode (C) the homology model in rigid mode and (D) the crystal structure in rigid mode.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Table 1 Prediction success (%) for all CYP3A4 metabolic pathways found using either the top-ranked site or the top three sites for the different models using MetaSite® methodology, the homology model in rigid mode, the homology model in flexible mode, the crystal structure in rigid mode and the crystal structure in flexible mode. Results for reactivity factor enabled or disabled are presented separately.

Models	Prediction success for top-ranked site (%)		Prediction success for top three sites (%)	
	Reactivity on	Reactivity off	Reactivity on	Reactivity off
Crystal flexible	41	20	68	38
Crystal rigid	41	21	72	40
Homology flexible	42	22	69	39
Homology rigid	42	20	69	41

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Table 2 Total metabolic pathways evaluated by docking methodology and prediction success using top three sites for the homology model, the crystal structure and the crystal structure with water models corresponding to best docking and all docking solutions.

	Reactions evaluated	Prediction using best docking (%)	Prediction using all dockings (%)
Homology model	323	26	57
Crystal structure	298	27	47
Crystal with water	205	19	27

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

Figure 1

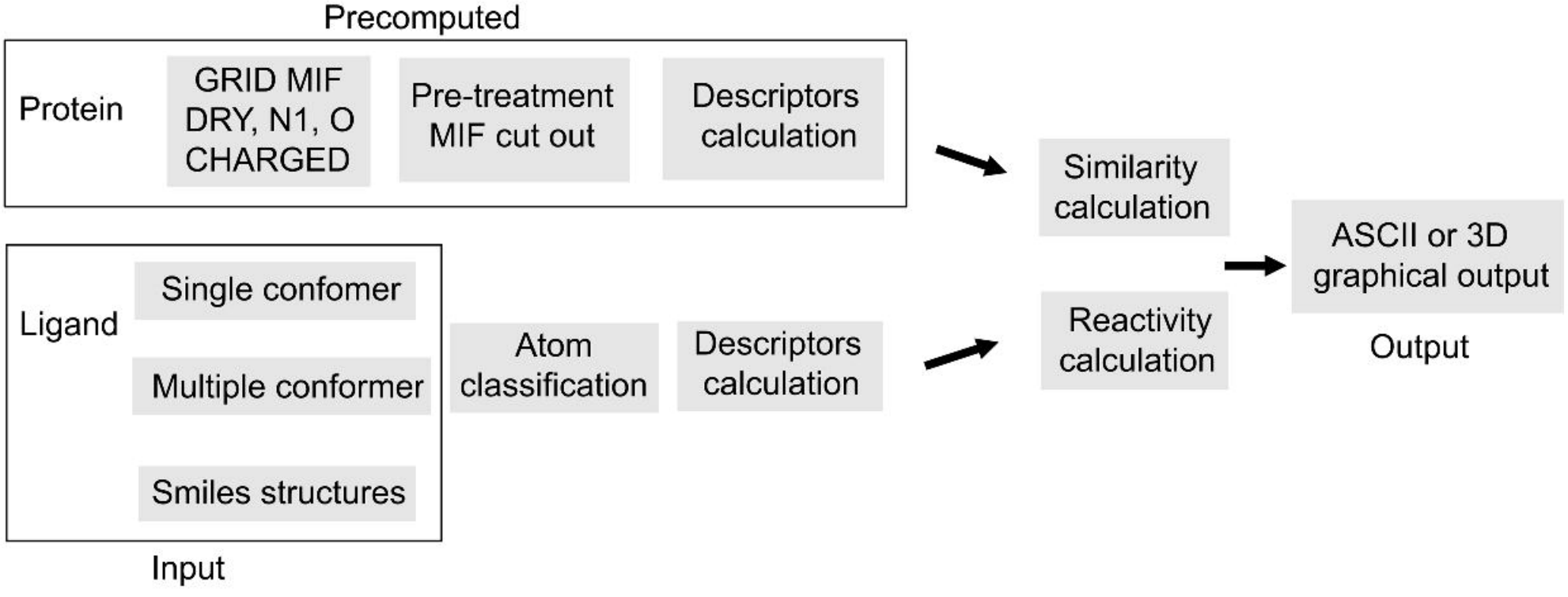

Precomputed
Protein
GRID MIF
DRY, N1, O
CHARGED
Pre-treatment
MIF cut out
Descriptors
calculation
Ligand
Single confomer
Multiple conformer
Smiles structures
Atom
classification
Descriptors
calculation
Input
Similarity
calculation
Reactivity
calculation
ASCII or 3D
graphical output
Output

Figure 2

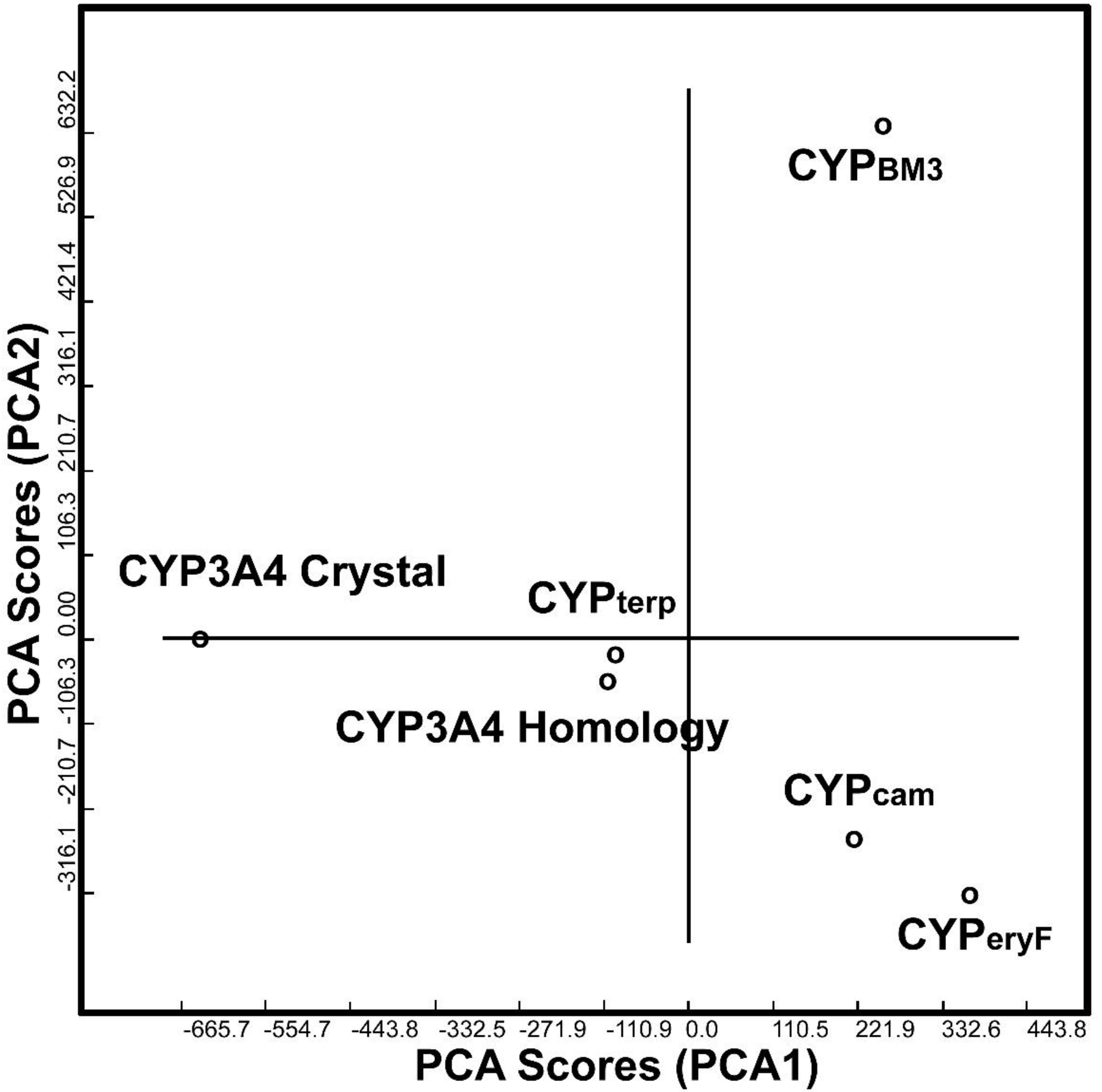
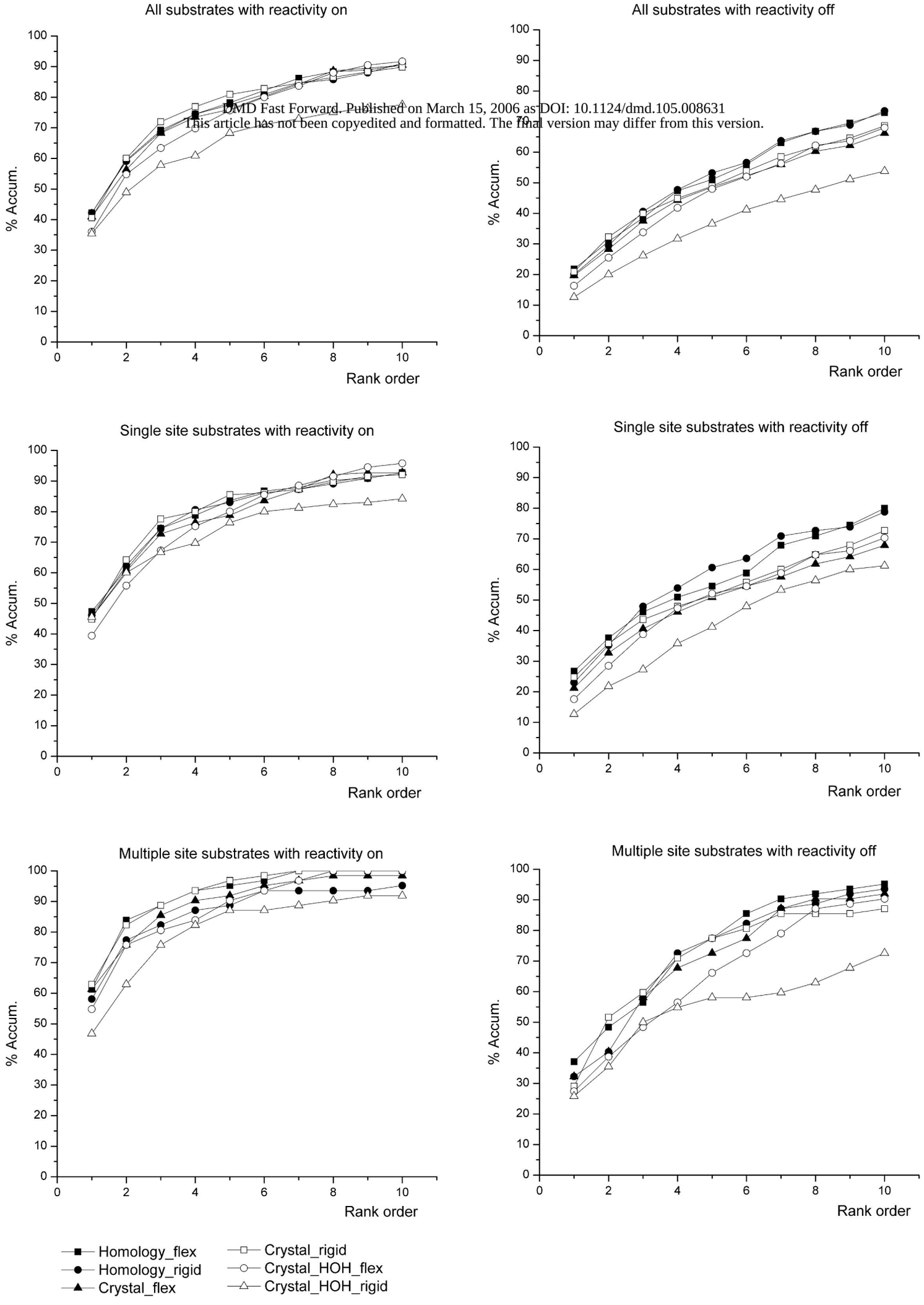

CYPBM3
CYP3A4 Crystal
CYPterp
CYP3A4 Homology
CYPcam
CYPeryF
PCA Scores (PCA2)
632.2
526.9
421.4
316.1
210.7
106.3
0.00
-106.3
-210.7
-316.1
-665.7
-554.7
-443.8
-332.5
-271.9
-110.9
0.0
110.5
221.9
332.6
443.8
PCA Scores (PCA1)

Figure 3



DMD Fast Forward. Published on March 15, 2006 as DOI: 10.1124/dmd.105.008631
This article has not been copyedited and formatted. The final version may differ from this version.

Downloaded from dmd.aspetjournals.org at ASPET Journals on April 20, 2024

Figure 4

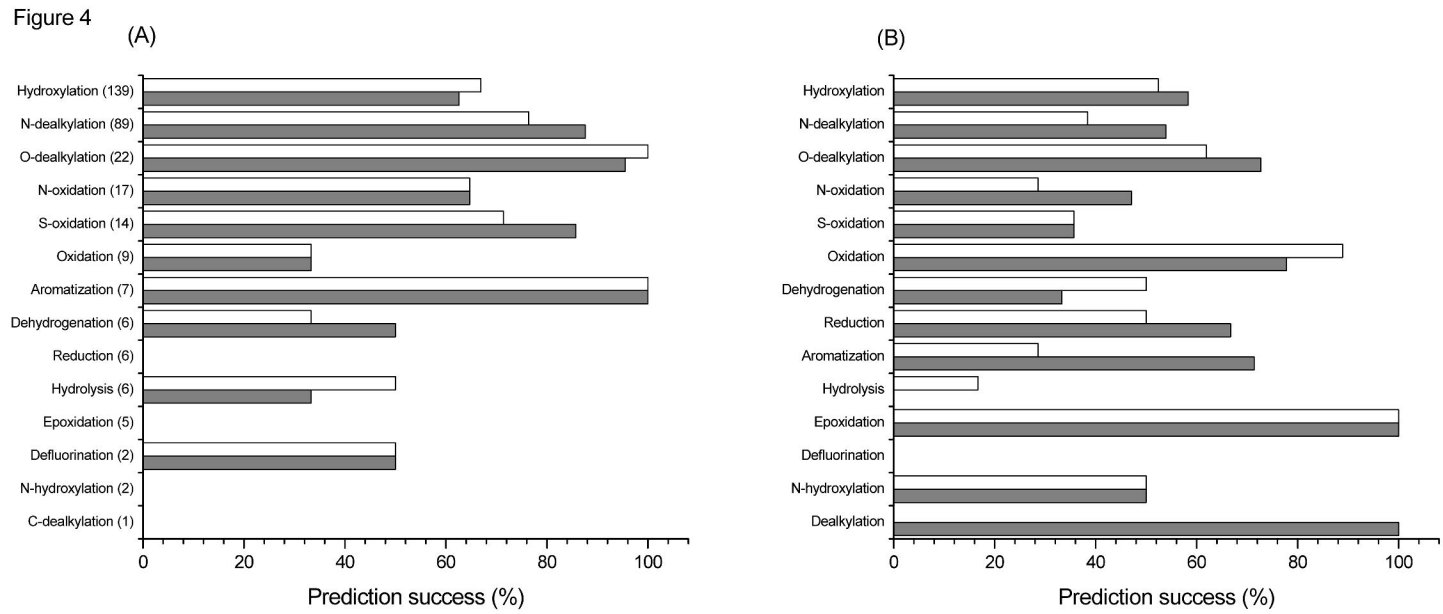

(A)
Hydroxylation (139)
N-dealkylation (89)
O-dealkylation (22)
N-oxidation (17)
S-oxidation (14)
Oxidation (9)
Aromatization (7)
Dehydrogenation (6)
Reduction (6)
Hydrolysis (6)
Epoxidation (5)
Defluorination (2)
N-hydroxylation (2)
C-dealkylation (1)
0
20
40
60
80
100
Prediction success (%)
(B)
Hydroxylation
N-dealkylation
O-dealkylation
N-oxidation
S-oxidation
Oxidation
Dehydrogenation
Reduction
Aromatization
Hydrolysis
Epoxidation
Defluorination
N-hydroxylation
Dealkylation
0
20
40
60
80
100
Prediction success (%)

Figure 5

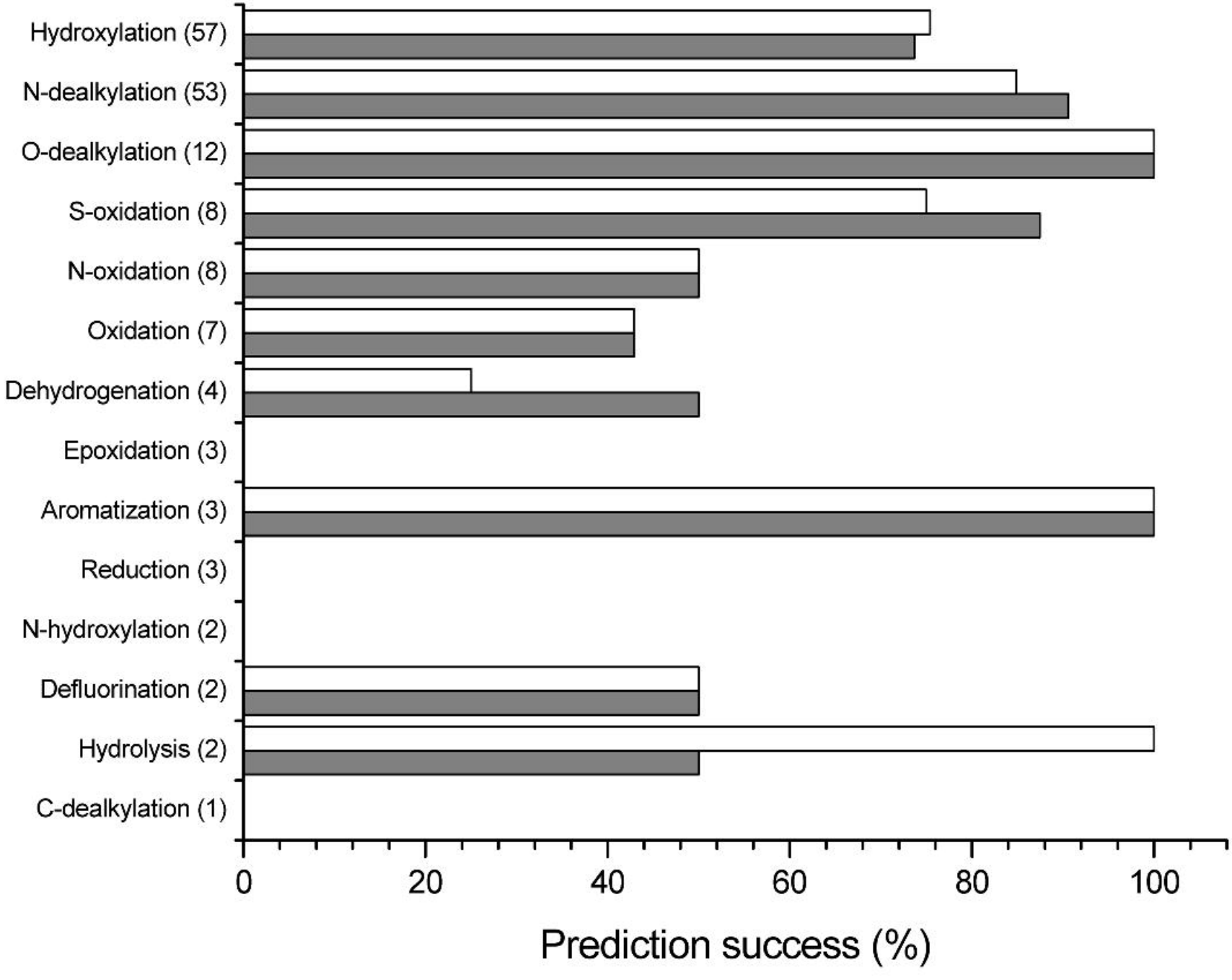
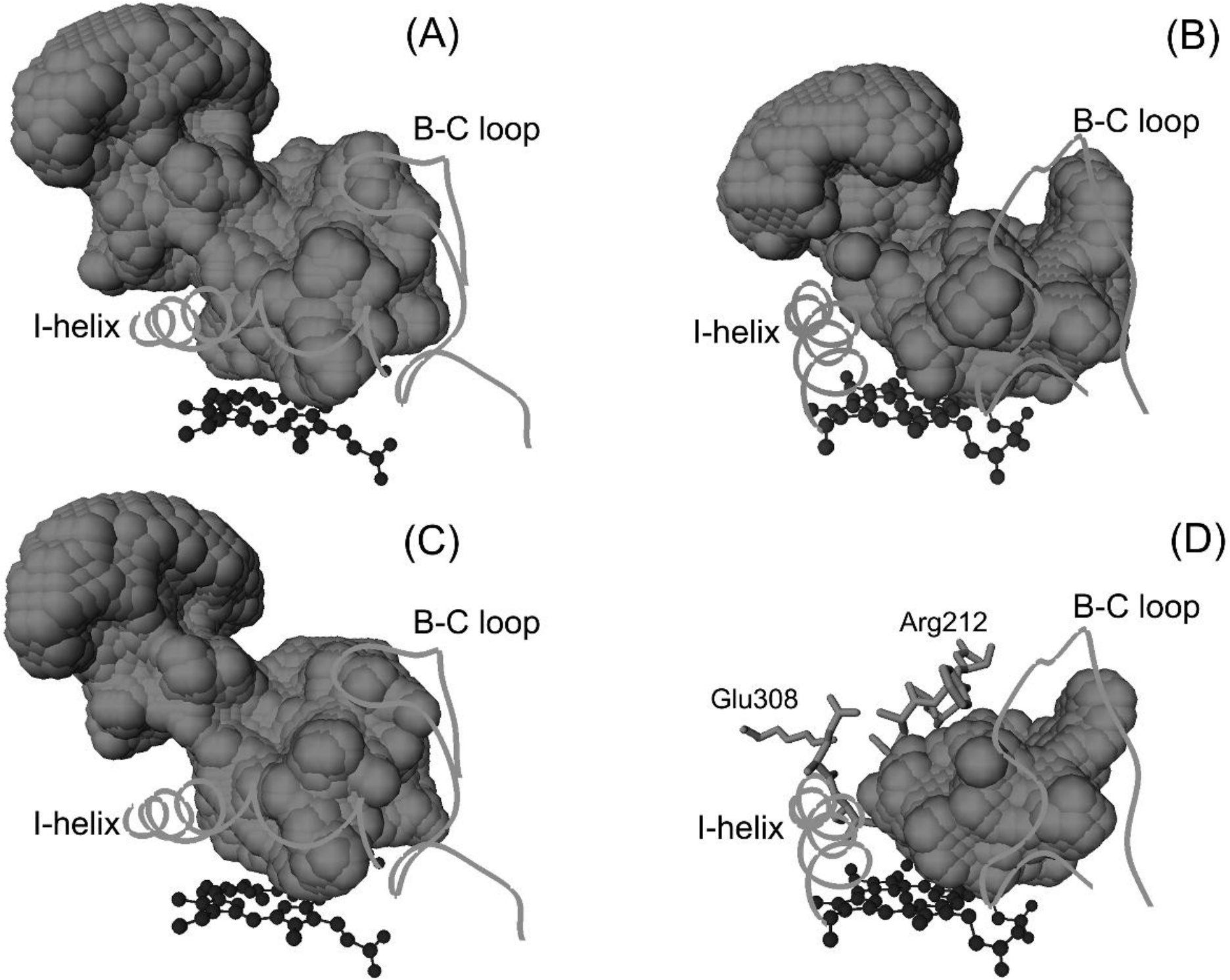

Hydroxylation (57)
N-dealkylation (53)
O-dealkylation (12)
S-oxidation (8)
N-oxidation (8)
Oxidation (7)
Dehydrogenation (4)
Epoxidation (3)
Aromatization (3)
Reduction (3)
N-hydroxylation (2)
Defluorination (2)
Hydrolysis (2)
C-dealkylation (1)
0
20
40
60
80
100
Prediction success (%)

Figure 6


(A)
B-C loop
I-helix
(B)
B-C loop
I-helix
(C)
B-C loop
I-helix
(D)
B-C loop
Arg212
Glu308
I-helix