Computational Prediction of Binding Affinity for
CYP1A2-Ligand Complexes Using Empirical Free
Energy Calculations

Poongavanam Vasanthanathan, Lars Olsen, Flemming Steen Jørgensen, Nico P. E. Vermeulen and Chris Oostenbrink

Leiden/Amsterdam Center for Drug Research, Division of Molecular Toxicology, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands (P.V., N.P.E.V., C.O.).

Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, 2100, Copenhagen, Denmark (P.V., L.O., F.S.J.).

Institute of Molecular Modeling and Simulation, University of Natural Resources and Applied Life Sciences, Muthgasse 18, 1190 Vienna, Austria (C.O.).
Running title: Binding Free Energy Calculation for CYP1A2 Ligands

Address correspondence to: Dr. Chris Oostenbrink, Institute of Molecular Modeling and Simulation, BOKU University, Muthgasse 18/6-58, A-1190, Vienna, Austria. Tel: +43-1-47654-8302, Email: chris.oostenbrink@boku.ac.at

Manuscript statistics

Number of text pages: 23
Number of tables: 5
Number of figures: 5
Number of references: 44
Number of words in Abstract: 184
Number of words in Introduction: 594
Number of words in Materials and Methods: 1225
Number of words in Results: 1073
Number of words in Discussion: 727

Abbreviations

ADMET: Absorption, Distribution, Metabolism, Excretion and Toxicity; CYP1A2: Cytochrome P450 1A2 Isoform; FEP: Free Energy Perturbation; GOLD: Genetic Optimization for Ligand Docking; LIE: Linear Interaction Energy; MD: Molecular Dynamics; MOE: Molecular Operating Environment; rms: Root Mean Square; SOM: Site of Metabolism; TI: Thermodynamic Integration.
Abstract

Predicting binding affinities for receptor-ligand complexes is still one of the challenging processes in computational structure-based ligand design. Many computational methods have been developed to achieve this goal, such as docking and scoring methods, the linear interaction energy (LIE) method and methods based on statistical mechanics. In the present investigation, we started from an LIE model to predict the binding free energy of structurally diverse compounds of cytochrome P450 1A2 ligands, one of the important human metabolizing isoforms of the cytochrome P450 family. The data set includes both substrates and inhibitors. It appears that the electrostatic contribution to the binding free energy becomes negligible in this particular protein and a simple empirical model was derived, based on a training set of 8 compounds. The root mean square error for the training set was 3.7 kJ/mol. Subsequent application of the model to an external test set gives an error of 2.1 kJ/mol, which is remarkably good, considering the simplicity of the model. The structure of the protein-ligand interactions are further analyzed, again demonstrating the large versatility and plasticity of the cytochrome P450 active site.
Introduction

The cytochromes P450 (P450s) are ubiquitous heme-containing enzymes, which play a vital role in the detoxification of xenobiotics and pro-carcinogen activation. This family shares a common structural feature of the enzymatic active site, including a planar ferric-protoporphyrin IX heme complex, in which the iron is attached through a cysteine link to the protein.

The metabolism of a drug is commonly a process of converting a hydrophobic compound into a hydrophilic one by introducing hydrophilic moieties, very often hydroxyl groups, to the parent compounds. More than 50 human isoforms of P450s have been characterized and only seven of them are responsible for > 90% of metabolism of currently marketed drugs (Williams et al., 2004; Nelson, 2009). The most important isoforms are CYP1A2, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and CYP3A4. Some of the P450s isoforms also show polymorphism (e.g. CYP1A2, CYP2C9 and CYP2D6), which leads to a poor pharmacokinetic profile. Early prediction of pharmacokinetic properties such as absorption, distribution, metabolism, elimination and toxicity of new compounds in drug discovery has become essential these days (DiMasi, 2001). This has been achieved by in vivo, in vitro and in silico methods, and should lead to a reduction of failure rates in the drug development process (van de Waterbeem and Gifford, 2003; Baranczewski et al., 2006; Afzelius et al., 2007; Stjernschantz et al., 2008).

Cytochrome P450 1A2 is an inducible member of the CYP superfamily. It is induced by some polycyclic aromatic hydrocarbons (PAH) and heterocyclic amines, some of which are found in cigarette smoke and charred food (Fontana et al., 1999). Activation and overexpression of this isoform is linked with high risk of various cancers (Seow et al., 2001; Jernstrom et al., 2008).

The importance of CYP1A2 in drug discovery is due to its metabolic contribution to endogenous
compounds (e.g. estradiol) and ~5% of the currently marketed drugs such as caffeine, naproxen, paracetamol, theophylline etc (Flockhart, 2007).

Substrates of CYP1A2 are characterized as planar neutral, aromatic (two to four aromatic rings) and lipophilic molecules with at least one putative hydrogen bond donor. This also matches the topology of the active site which is well characterized using X-ray crystallography (Sansen et al., 2007) (Figure 1) and computational models (Smith et al., 1997; Ekins et al., 2001).

Molecular modeling methods have been used over the years to study and predict the CYPs active site accessibility (Harris, 2004), ligand binding affinity (de Groot et al., 2004; Vasanthanathan et al., 2009b), site of metabolism (SOM) (Cruciani et al., 2005; Vasanthanathan et al., 2009a), selectivity and specificity (Terfloth et al., 2007; Jung et al., 2008), toxicity (Hansch et al., 2004), virtual screening and rule based methods for CYP mediated metabolic reactions (Rydberg et al., 2009).

In previous studies, we have shown that docking and scoring approaches may be used to predict the site of metabolism of substrates or to screen for binders from a database of compounds, but that these methods are not suitable to rank binders according to their affinity (Vasanthanathan et al., 2009a). Here, we explore the possibilities to predict binding affinities for a set of structurally diverse ligands of CYP1A2, starting from the linear interaction energy (LIE) method (Aqvist et al., 1994). We will use a set of compounds that can act either as a substrate or as an inhibitor and for which the affinity is determined using a fluorimetric assay (Cohen et al., 2003). The models will be trained on one dataset and validated using a second set of compounds (Kobayashi et al., 1998). In addition to this, we will analyze the dynamic behavior of the protein-ligand complexes.
Materials and Methods

Computational methodology

Predicting binding affinities for structurally diverse compounds is still a challenging task in computer aided drug design (Gilson and Zhou, 2007). Very often docking and scoring functions have been used to evaluate the binding free energy for large databases, because these are computationally inexpensive approaches. However, these methods often only allow for limited protein flexibility and desolvation and the prediction of affinities from a score is usually rather poor. On the other hand, free energy perturbation (FEP) and thermodynamic integration (TI) (Beveridge and DiCapua, 1989) are derived from statistical mechanics and are correct in the limit of infinite sampling and infinitely accurate force fields. However, these methods are computationally very expensive and difficult to apply to structurally diverse compounds. The LIE method may be positioned in between, having a theoretical basis in linear response theory and including protein flexibility and solvation effects.

The LIE method requires the calculation of average interaction energies between the ligand and its surroundings from molecular dynamics (MD) or monte carlo (MC) simulations of the ligand free in solution and when bound to the protein. The binding free energy is subsequently estimated using the following equation

$$
\Delta G_{\text{bind}} = \alpha \langle V_{l-s}^{\text{vdw}} \rangle_{\text{Bound}} - \alpha \langle V_{l-s}^{\text{vdw}} \rangle_{\text{Free}} + \beta \langle V_{l-s}^{\text{el}} \rangle_{\text{Bound}} + \beta \langle V_{l-s}^{\text{el}} \rangle_{\text{Free}} + \gamma
$$

In this equation, the brackets (\langle \rangle) denote an ensemble average of the intermolecular electrostatic (el) and van der Waals (vdw) energies of the ligand with its surroundings (l-s) in the bound and free states. Parameters of this equation are the coefficients \( \alpha \) and \( \beta \) for the non-polar
and polar contributions to the binding free energy, respectively, and \( \gamma \) as an additional constant. In the original derivation, \( \gamma \) was not included, the value of \( \beta \) was derived from linear response to be 0.5 and \( \alpha \) was empirically determined to be 0.16 (Aqvist et al., 1994). Later studies have shown that \( \beta \) may deviate from 0.5 depending on the system and \( \alpha \), \( \beta \) and \( \gamma \) are often used as empirical parameters to train the LIE equation on a given set of binding affinities (Aqvist et al., 2002). In the current work, we will investigate various forms of the LIE equation on CYP1A2 inhibition data. Previously, this method was applied on CYP1A1 (Liu et al., 2003) and CYPcam (Paulsen et al., 1996; Almlöf et al., 2004), although some of that work was based on significantly shorter simulations.

**Preparation of CYP1A2 and ligands**

A consistent set of drugs (training compounds, Figure 2A) with IC\(_{50}\) values was collected from the literature (Cohen et al., 2003). The IC\(_{50}\) values were all determined in a common phenacetin assay, in which the formation of the fluorescent 4-acetaminophenol was measured. A second set of 5 compounds (test set, Figure 2B) for which IC\(_{50}\) values were determined using the same assay, was obtained from a different study (Kobayashi et al., 1998). Known suicide inhibitors binding irreversibly to the enzyme were not included in the data set to avoid uncertainties in the model building. All IC\(_{50}\) values were converted to K\(_i\) values using the Cheng-Prusoff equation to correct for different substrate concentrations. The binding free energy (\( \Delta G_{\text{bind}} \)) was estimated using \( \Delta G_{\text{bind}} = -k_B T \ln K_i \) and compared to the calculated free energies. Here \( k_B \) is the Boltzmann constant and \( T \) is the absolute temperature.

The crystal structure of human cytochrome P450 1A2 in complex with \( \alpha \)-naphthoflavone (\( \alpha \)NF) was obtained from the Protein Data Bank (1.95 Å, PDB entry 2H14) (Sansen et al., 2007). Hydrogen atoms were added using the MOE software (version 2008.10) 3D structures of the
ligands were built in MOE. All compounds used in the study were modeled in their neutral form. The structures were energy minimized using the MMFF94s force field (Halgren, 1996).

**Initial ligand conformations**

GOLD version 3.2 (Jones et al., 1997) (Genetic Optimization for Ligand Docking) was used to obtain initial conformations and orientations of the ligands in the active site of CYP1A2 prior to the molecular dynamics simulations. All compounds were docked into the enzyme with the active site water molecule that bridges an interaction between the enzyme and αNF included in the active site (Sansen et al., 2007). The Chemscore scoring function was used and other parameter settings were described in detail elsewhere (Vasanthanathan et al., 2009a). For the substrates in both sets of compounds, the major and minor sites of metabolism are indicated in Figure 2. This information was used to select proper starting configurations with the site of metabolism pointing towards the heme iron. Details are given in Table 1. For mexillitine and reluzole an additional distance restraint (length 6 Å and a force constant of 50 kJ/mol/Å²) was added during the docking procedure between the experimentally observed site of metabolism and the heme iron atom.

**Molecular dynamics simulations**

All molecular dynamics simulations were performed and analyzed using the GROMOS05 simulation package (Christen et al., 2005). The GROMOS 45A4 force field was used. Interaction parameters for the ligands were determined by analogy to similar groups in the force field and are available in the supplementary material. In GROMOS, the parameterization of new ligands involves careful manual selection of all interaction parameters. In CYPcam, it was shown that LIE is relatively robust with respect to various force fields (Almlöf et al, 2004). The initial structures of each ligand and CYP1A2-ligand complex were energy minimized separately and
placed into a pre-equilibrated rectangular box with SPC water (Berendsen et al., 1981). The minimum solute to solvent distance and minimum solute to box wall distance were set to 0.23 nm and 1.4 nm, respectively. The protein systems now contained about 39,000 water molecules and were energy minimized using steepest descent minimization. During minimization and equilibration, solute atoms were positionally restrained in order to relax any unfavorable solute-solvent interactions. In the process of thermalization and equilibration, initial velocities are generated from a Maxwell-Boltzmann distribution at 60 K and gradually increased to 300 K. An additional simulation at a constant pressure of 1 atm was performed for 100 ps. Temperature and pressure were kept constant using the weak coupling scheme (Berendsen et al., 1984) with relaxation times of 0.1 ps and 0.5 ps, respectively. The SHAKE algorithm was used to constrain the lengths of all bonds, allowing for a time steps of 2 fs (Ryckaert et al., 1977).

Non-bonded interactions were calculated using a triple range scheme. Interactions within a short range cutoff of 0.8 nm were calculated every time step from a pairlist that was generated every fifth step. At these time points, interactions between 0.8 and 1.4 nm were also calculated and kept constant between updates. A reaction-field contribution was added to the electrostatic interactions and forces to account for a homogeneous medium outside the long-range cutoff (Tironi et al., 1995), using a relative dielectric constant of 61. After the equilibration phase the MD production run was started at constant temperature and pressure. No restraints were applied during this step. Electrostatic and van der Waals ligand-surrounding energies were stored from simulations of the ligands bound to the protein and when free in solution every 0.2 ps. All simulations were initially run for 1 ns and convergence of the ligand-surrounding energies was checked by visual inspection. Whenever necessary, the simulations were extended by another 1 ns. Coordinates of the protein ligand complexes were stored every 5 ps for further analysis. All
simulations were performed on a cluster and the total time required for 1 ns simulation was approximately 2 days using 4 CPUs in parallel.
Results

Ligand orientation in the active site

Binding affinities for two sets of structurally diverse substrates and inhibitors from the literature were predicted (Figure 2). All ligands were docked into the CYP1A2 active site to obtain the correct binding orientation (see above). Initially, the fifty best docking poses were analyzed, if necessary, dockings with constraints were performed. Here, we have used the approach that if experimental information on the binding orientation was available in terms of observed product formation, this information was used to select an appropriate initial pose. If such information was not available, the first ranked pose was used in the simulation. Table 1 summarizes the docking results and initial pose selection for the 1A2 substrates of which 6 were part of the training set and 2 were part of the test set. Figure 3 compares the first ranked poses to the selected poses for the substrates propranolol (PPD), tacrin (TCN), mexillitine (MXL) and reluzole (RZL). All in all, the first ranked pose was selected for 8 of the 13 compounds.

MD and affinity prediction from interaction energies

Molecular dynamics simulations were performed for 13 compounds, both free in solution and bound in the active site of CYP1A2, respectively. The convergence of the simulations was checked periodically by monitoring the interaction energies between the ligand and the surrounding molecules (CYP1A2 and water). Whenever necessary, the simulations were extended by 1 or 2 ns, and ensemble averages were obtained from at least 1 ns of simulation. Representative examples of time series of the ligand-surrounding interaction energies are given in the supplementary material. Table 2 gives an overview of the average of the van der Waals and electrostatic interaction energies in the free and bound forms. The van der Waals (non-polar)
energy seems to favor binding more than the electrostatic interactions. As was shown in Figure 1, a number of non-polar aromatic amino acids are present in the active site accommodating the aromatic rings of the ligands.

Using the average interaction energies, LIE models were created using equation 1. The 8 compounds in Figure 2A were used as training set and the 5 compounds in Figure 2B as independent test set. The quality of the models was evaluated by means of the root mean square (rms) error. A summary of various LIE models developed in this work is presented in Table 3.

Application of a standard LIE model, with parameters $\alpha$ and $\beta$ only, leads to model 1, with an rms error of 6.0 kJ/mol for the training set and of 7.8 kJ/mol for the test set. As most compounds were underestimated, we included a constant parameter $\gamma$ in model 2, which showed an rms error of 3.7 kJ/mol for the training set and 2.2 kJ/mol for the test set. However, fitting three parameters for eight compounds can be considered as overfitting. We note from both models 1 and 2 that the value of $\beta$ is quite low, leading on average to unfavorable electrostatic contributions of 0.5 kJ/mol to the binding free energy. This indicates that for this system, the electrostatic interaction energies do not seem to correlate with the binding affinity. We therefore decided to parameterize model 3, which can be no longer be considered as an LIE model, but is a purely empirical correlation of the average non-polar interaction energies with the experimental binding affinities.

Using only $\alpha$ and $\gamma$ as empirical parameters and $\beta=0$, we obtain an rms error of 3.7 kJ/mol for the training set and 2.1 kJ/mol for the test set. We subsequently derived models 4 and 5, which do not employ a constant, and in which the value of $\beta$ was forced to be 0.5 and 0.0 respectively. It is clear that the errors increase significantly in these models. As a consistency check against models 2 and 3, we have also derived model 6, using all 13 compounds as a training set to fit the
parameters $\alpha$, $\beta$ and $\gamma$. Again, we find an extremely low value of $\beta$ while $\alpha$ and $\gamma$ remain roughly the same and the overall error is comparable.

The correlation between calculated and experimental binding free energies is shown in Figure 4. Compounds mexilitine, riluzole, tacrine, aprindine, bepridil, lidocaine and propafenone are all predicted with this model with deviations from the experimental binding free energy smaller than 2.5 kJ/mol, the value of $k_B T$ at room temperature. The prediction for compounds phenacetin, propranolol and quinidine is reasonable, with an error of less than 4.0 kJ/mol. Average deviations $\sim 4.2$ kJ/mol (1 kcal/mol) are commonly reported for empirical free energy calculations (Aqvist et al., 2002). Ellipticine and $\alpha$-naphthoflavone are reported as strong binders for CYP1A2 with binding affinities of -41.7 kJ/mol and -42.5 kJ/mol, respectively. Model 3 predicts them to be poorer binders, by at most 5.8 kJ/mol, which translates to roughly a factor 10 in the $K_i$ or $IC_{50}$. Even though this is only a small dataset, we stress that the compounds are structurally diverse. These results are similar to previously reported LIE calculations on much less diverse compounds for CYPcam (Paulsen et al., 1996; Almlöf , et al., 2004). However, we note that the character of the active site of CYPcam is significantly different from the one of CYP1A2. The $\Delta G$ values as predicted by the Chemscore scoring function do not correlate with the experimental values ($r^2 = 0.03$) and show a rms error of 13.8 kJ/mol.

Analyzing the MD trajectory of ligands in the bound state provides interesting structural and dynamic information regarding the interactions between the ligands and CYP1A2. For all compounds the atom-positional root-mean-square deviation for the backbone remains below 0.25 nm throughout the simulations. For CYP1A2, most of the relevant dynamics are expected to take place at the level of the amino acid side chains and water molecules in the active site. This can well be captured on a 1 ns simulation scale. For the substrates, the distance between the heme-
iron ion and the major site of metabolism as indicated in Figure 2 was monitored (Table 4). For most substrates, the site of metabolism stays within 6 Å, which we consider to be the upper bound for a reactive pose of the substrate. For ellipticine and mexilitine the SOM moves further away from the heme during the simulation. This also happens for reluzole repeatedly for periods of maximally 50 ps. The NH\textsubscript{2} group of reluzole, however, always returns to a distance of about 4 Å from the heme iron.
Discussion

The fact that a predictive free energy model can be constructed that does not include electrostatic interaction energies does not mean that these do not play a role in binding. All ligands were seen to form hydrogen bonds within the active site to some extent. A summary of the occurrences of hydrogen bonds between active site residues and ligands is shown in Table 5. Some residues are involved in hydrogen bonds with most of the 13 ligands in the simulation, while other residues are more selective. For example, Thr124 forms a hydrogen bond with 9 out of 13 compounds (α-naphthoflavone, ellipticine, furafylline, mexilitine, propranolol, riluzole, tacrine, lidocaine, and quinidine (S)) and similarly, Phe125, Asp320, and Gly316 form hydrogen bonds to five ligands each. Overall propafenone, quinidine (S), riluzole, tacrine, furafylline and α-naphthoflavone interact via hydrogen bonds with the enzyme for a large part of the simulation time. Most compounds also show a significant number of hydrogen bonds with water molecules in the active site. Only compounds aprindine, bepridil and lidocaine are involved in hydrogen bonds for less than 30% of the time. It seems that hydrogen bond donors are better accommodated in the active site than hydrogen bond acceptors.

In the reported crystal structure of the human CYP1A2 isoform the inhibitor α-naphthoflavone, which contains only hydrogen bond acceptors, forms a hydrogen bond with one water molecule in the active site. It is known that water molecules may affect the outcome of e.g. docking and virtual screening experiments (de Graaf C. et al., 2005; Santos et al., 2009; Vasanthanathan et al., 2009a). We have calculated the hydrogen bonds between ligands and the active site water from the MD trajectories. During the simulations, different water molecules seem to play a role at different points in time. This is exemplified in Figure 5A, where the occurrence of various hydrogen bonds with water (Y-axis) is given as a function of time. In the case of propranolol the
same hydrogen bond is after some initial switches maintained throughout the simulation. For α-naphthoflavone, it can be clearly seen that one hydrogen bond is being exchanged by another one at several points in time.

Further, we have monitored the position of the various water molecules that form hydrogen bonds with the ligands (Figure 5B). From this analysis it is clear that the position of the water molecule that is observed in the X-ray structure is indeed occupied for most of the time, but not necessarily by the same water molecule.

In summary, the electrostatic interactions between ligands and CYP1A2 do not seem to be needed to explain the difference in binding affinity between the compounds. Rather, this seems to be governed by shape complementarity as represented by the Van der Waals interaction energies. However, electrostatic interactions do play a role in binding, but with varying partners in the protein active site. This leads us to believe that this is another demonstration of the versatility and plasticity, previously observed for other CYPs (Guengerich, 2006; Hritz et al., 2008). If the ligand can be accommodated by the active site based on shape complementarity, then the protein is able to offer the necessary hydrogen bonding partners using various amino acids and water molecules.

**Conclusion**

We have investigated the applicability of an LIE model to predict the binding free energy to one of the most important bio-transforming cytochrome P450 isoforms, CYP1A2, with two sets of structurally and experimentally diverse compounds. It appears that the electrostatic contribution to the binding affinity becomes negligible and that a predictive model can be obtained using only the van der Waals interaction energies and a constant. This model yielded a rms error from the
experimental values of 3.7 kJ/mol for the training set and of 2.1 kJ/mol for an independent test set. This is a remarkably good agreement, considering the simplicity of the model. The model indicates that non-polar interactions are mainly governing the affinity of the CYP1A2 ligands. When bound, the ligands form hydrogen bonds to various residues, possibly mediated through various (exchanging) water molecules.

It may very well be that this is another demonstration of the extreme versatility and plasticity of CYP active sites. While the affinity is mostly governed by the non-bonded interaction energies and shape complementarity, the protein is able to accommodate a small number of hydrogen bonding moieties through interactions with various residues and water molecules.
Acknowledgments

Authors gratefully acknowledge the financial assistance provided by the Drug Research Academy, Faculty of Pharmaceutical Sciences, University of Copenhagen, Copenhagen, Denmark (PV), the Carlsberg Foundation, Copenhagen, Denmark (LO) and Netherlands Organization for Scientific Research, VENI Grant No. 700.55.401 (CO).
References


Beveridge DL and DiCapua FM (1989) Free energy via molecular simulation: applications to


pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios.

*Drug Metab Dispos* **32**: 1201-1208.
Figure captions

Figure 1.
Active site of CYP1A2 characterized with molecular interaction fields (MIF), left panel: active site cavity (C3 probe with energy level of -0.01 kcal/mol); right panel: hydrophobic interactions (DRY probe with energy level of -0.07 kcal/mol). Molecular MIF were generated using the GRID program (Goodford, 1985).

Figure 2.
CYP1A2 ligands used in this study. Major and minor metabolic positions for substrates are indicated using full and dashed spheres, respectively. Compounds without any mark are inhibitors A: Training compounds, B: Test compounds.

Figure 3.
Initial docking pose used for MD simulations, left panel shows the first ranked docking solution for propranolol (A), tacrine (C), mexiltine (E) and reluzole (G) and right panel shows alternative/constrained docking solution for the above mentioned compounds (B, D, F, H) used as initial pose for MD simulations (see text for details).

Figure 4.
Correlation between the calculated and observed binding free energies for training and test compounds, using the empirical model 3 in Table 3.

Figure 5.
Time series of water mediated hydrogen bonds to ligands α-naphthoflavone and propranolol (A). The positions of hydrogen bond forming water molecules during the MD simulations of α-naphthoflavone (left) and propranolol (right). Water molecules are shown as space filling spheres. The crystal water is labeled with ‘c’ (B).
Table 1.

Generation of initial conformation from docking experiments for CYP1A2 substrates for MD simulations.

<table>
<thead>
<tr>
<th>Substrates</th>
<th>Metabolic reaction</th>
<th>Distance to Fe first ranked pose (Å)</th>
<th>Selected initial pose for MD</th>
<th>Distance to Fe Selected pose (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Major</td>
<td>Minor</td>
<td>Major</td>
</tr>
<tr>
<td>Ellipticine</td>
<td>Aromatic-OH C2</td>
<td></td>
<td>Aromatic-OH C3</td>
<td>5.3</td>
</tr>
<tr>
<td>Mexilitine</td>
<td>N-oxidation N3</td>
<td>7.9</td>
<td>8.8</td>
<td>Constrained docking</td>
</tr>
<tr>
<td>Phenacetin</td>
<td>Aliphatic-OH C2</td>
<td>3.7</td>
<td>-</td>
<td>First pose</td>
</tr>
<tr>
<td>Propranolol</td>
<td>N-deisopropylation C2</td>
<td>11.3</td>
<td>4.5, 6.2</td>
<td>Sixteenth pose</td>
</tr>
<tr>
<td>Riluzole</td>
<td>N-oxidation N1</td>
<td>8.9</td>
<td>-</td>
<td>Constrained docking</td>
</tr>
<tr>
<td>Tacrine</td>
<td>Aliphatic-OH C1</td>
<td>9.4</td>
<td>10.3, 9.8, 8.8</td>
<td>Second pose</td>
</tr>
<tr>
<td>Lidocaine</td>
<td>N-deethylation C1,C5</td>
<td>5.3, 6.0</td>
<td>11.4</td>
<td>First pose</td>
</tr>
<tr>
<td>Propafenone</td>
<td>N-depropylation C1</td>
<td>13.0</td>
<td>-</td>
<td>Twelveth pose</td>
</tr>
</tbody>
</table>

Note: OH: Hydroxylation
Table 2.

Experimental and calculated free energies of binding of CYP1A2 compounds and ligand-surrounding interaction energies (Averages of van der Waals ($\Delta V_{vdw}$), electrostatic ($\Delta V_{el}$) energy terms for the free and bound form). $\Delta G_{cal}$ is obtained using the empirical parameters of model 3 in Table 3. Compounds 1-8 formed the training set in this parameterization, while compounds 9-13 were used as test set.

<table>
<thead>
<tr>
<th>No</th>
<th>Compounds</th>
<th>$\Delta G_{Exp}$ (kJ/mol)</th>
<th>Ligand-surrounding interactions (kJ/mol)</th>
<th>$\Delta G_{Cal}$ (kJ/mol)</th>
<th>Error (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\alpha$-naphthoflavone</td>
<td>-42.5</td>
<td>$-203.12 \quad -123.70 \quad -32.55 \quad -45.93$</td>
<td>-37.3</td>
<td>5.2</td>
</tr>
<tr>
<td>2</td>
<td>Ellipticine</td>
<td>-41.7</td>
<td>$-174.60 \quad -101.61 \quad -25.62 \quad -61.44$</td>
<td>-36.0</td>
<td>5.8</td>
</tr>
<tr>
<td>3</td>
<td>Furafylline</td>
<td>-35.7</td>
<td>$-195.11 \quad -102.04 \quad -62.24 \quad -131.82$</td>
<td>-40.1</td>
<td>-4.4</td>
</tr>
<tr>
<td>4</td>
<td>Mexilitine</td>
<td>-34.7</td>
<td>$-128.24 \quad -71.35 \quad -42.34 \quad -55.15$</td>
<td>-32.6</td>
<td>2.1</td>
</tr>
<tr>
<td>5</td>
<td>Phenacetin</td>
<td>-27.9</td>
<td>$-130.45 \quad -81.37 \quad -53.33 \quad -52.82$</td>
<td>-30.9</td>
<td>-3.1</td>
</tr>
<tr>
<td>6</td>
<td>Propranolol</td>
<td>-32.8</td>
<td>$-177.73 \quad -102.03 \quad -69.17 \quad -81.72$</td>
<td>-36.5</td>
<td>-3.6</td>
</tr>
<tr>
<td>7</td>
<td>Riluzole</td>
<td>-29.3</td>
<td>$-106.79 \quad -64.40 \quad -25.73 \quad -77.79$</td>
<td>-29.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>8</td>
<td>Tacrine</td>
<td>-32.1</td>
<td>$-139.04 \quad -75.63 \quad -69.39 \quad -80.34$</td>
<td>-34.0</td>
<td>-1.8</td>
</tr>
<tr>
<td>9</td>
<td>Aprindine</td>
<td>-39.1</td>
<td>$-229.37 \quad -135.16 \quad -12.67 \quad -33.70$</td>
<td>-40.3</td>
<td>-1.2</td>
</tr>
<tr>
<td>10</td>
<td>Bepridil</td>
<td>-39.5</td>
<td>$-249.73 \quad -151.94 \quad -24.41 \quad -43.56$</td>
<td>-41.1</td>
<td>-1.6</td>
</tr>
<tr>
<td>11</td>
<td>Lidocaine</td>
<td>-37.0</td>
<td>$-159.18 \quad -90.53 \quad -26.42 \quad -61.47$</td>
<td>-35.0</td>
<td>2.0</td>
</tr>
<tr>
<td>12</td>
<td>Propafenone</td>
<td>-43.3</td>
<td>$-225.36 \quad -125.33 \quad -78.06 \quad -105.67$</td>
<td>-41.5</td>
<td>1.8</td>
</tr>
<tr>
<td>13</td>
<td>Quinidine (S)</td>
<td>-35.3</td>
<td>$-198.51 \quad -112.74 \quad -77.57 \quad -114.26$</td>
<td>-38.6</td>
<td>-3.3</td>
</tr>
</tbody>
</table>
Table 3.

Summary of LIE models for training and test compounds (see text for detail).

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>rms error (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>β</td>
</tr>
<tr>
<td>1</td>
<td>0.510</td>
<td>0.034</td>
</tr>
<tr>
<td>2</td>
<td>0.218</td>
<td>0.020</td>
</tr>
<tr>
<td>3</td>
<td>0.208</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>0.698</td>
<td>0.500</td>
</tr>
<tr>
<td>5</td>
<td>0.5026</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0.196</td>
<td>0.014</td>
</tr>
</tbody>
</table>
Table 4.

Distance between the heme Fe ion and the major metabolic position indicated in figure 2, for the substrates, monitored over the MD simulation.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>ELL</th>
<th>MXL</th>
<th>PNC</th>
<th>PPD</th>
<th>RZL</th>
<th>TCN</th>
<th>LDC</th>
<th>PPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial distance</td>
<td>4.1</td>
<td>5.0</td>
<td>3.7</td>
<td>5.1</td>
<td>4.4</td>
<td>5.6</td>
<td>5.3</td>
<td>5.0</td>
</tr>
<tr>
<td>Average</td>
<td>6.3</td>
<td>8.7</td>
<td>4.6</td>
<td>5.1</td>
<td>5.4</td>
<td>4.4</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>Minimum distance</td>
<td>5.0</td>
<td>6.7</td>
<td>3.5</td>
<td>4.0</td>
<td>3.0</td>
<td>3.6</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Maximum distance</td>
<td>7.6</td>
<td>10.8</td>
<td>5.8</td>
<td>6.1</td>
<td>7.9</td>
<td>5.2</td>
<td>5.1</td>
<td>5.1</td>
</tr>
<tr>
<td>RMSD</td>
<td>0.39</td>
<td>0.53</td>
<td>0.41</td>
<td>0.37</td>
<td>0.59</td>
<td>0.30</td>
<td>0.36</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Abbreviations: ELL: Ellipticine; MXL: Mexilitine; PNC: Phenacetin; PPD: Propranolol; RZL: Riluzole; TCN: Tacrine; LDC: Lidocaine; PPF: Propafenone.

For ELL and MXL the distance increased during the equilibration period of the simulation.
Table 5.

Occurrence of hydrogen bonds between ligands and CYP1A2 residues, Heme and Water.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>ANF</th>
<th>ELL</th>
<th>FYN</th>
<th>MXL</th>
<th>PNC</th>
<th>PPD</th>
<th>RZL</th>
<th>TCN</th>
<th>APN</th>
<th>BPL</th>
<th>LDC</th>
<th>PPF</th>
<th>QDN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr-118</td>
<td>2.0</td>
<td>-</td>
<td>23.5</td>
<td>-</td>
<td>3.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>16.7</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr-124</td>
<td>0.5</td>
<td>11.5</td>
<td>0.5</td>
<td>0.5</td>
<td>-</td>
<td>2.0</td>
<td>4.3</td>
<td>95.0</td>
<td>-</td>
<td>-</td>
<td>1.7</td>
<td>-</td>
<td>0.5</td>
</tr>
<tr>
<td>Thr-321</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>35.8</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thr-385</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thr-498</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>90.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Phe-125</td>
<td>10.0</td>
<td>-</td>
<td>20.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10.0</td>
<td>16.7</td>
<td>-</td>
<td>10.0</td>
</tr>
<tr>
<td>Phe-226</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.7</td>
<td>6.7</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Phe-256</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>8.5</td>
<td>-</td>
<td>5.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Phe-260</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp-313</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>12.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp-320</td>
<td>33.5</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>3.5</td>
<td>-</td>
<td>-</td>
<td>13.0</td>
<td>-</td>
<td>78.3</td>
<td>33.5</td>
<td></td>
</tr>
<tr>
<td>Asn-257</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.0</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asn-312</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.5</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>Ala-317</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>23.3</td>
<td>-</td>
<td>3.5</td>
<td>-</td>
<td>7.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gly-316</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>19.0</td>
<td>-</td>
<td>1.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Heme</td>
<td>4.5</td>
<td>-</td>
<td>-</td>
<td>36.3</td>
<td>-</td>
<td>8.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Water</td>
<td>129.5</td>
<td>161.5</td>
<td>136.5</td>
<td>172.0</td>
<td>232.0</td>
<td>201.5</td>
<td>18.8</td>
<td>159.0</td>
<td>0.5</td>
<td>1.0</td>
<td>-</td>
<td>4.5</td>
<td>129.5</td>
</tr>
<tr>
<td>Total</td>
<td>181.0</td>
<td>173.5</td>
<td>189.5</td>
<td>174.0</td>
<td>238.0</td>
<td>222.5</td>
<td>118.7</td>
<td>254.5</td>
<td>28.0</td>
<td>21.5</td>
<td>27.5</td>
<td>215.0</td>
<td>181.0</td>
</tr>
</tbody>
</table>


*a* the occurrence of a hydrogen bond was monitored using a geometric criterion. A hydrogen bond is defined by a minimum donor-hydrogen-acceptor angle of 135° and a maximum hydrogen-acceptor distance of 0.25 nm.
Figure 1
Figure 2

A

- Alpha-Naphthoflavon (ANF)
- Ellipticin (ELL)
- Furafylline (FYH)
- Mexiletine (MXL)
- Phenacetin (PNC)
- Propranolol (PPD)
- Riluzole (RZL)
- Tacrine (TCN)

B

- Aprindine (APN)
- Bepridil (BPL)
- Lidocaine (LDC)
- Propafenone (PPF)
- Quinidine (QDN)
Figure 4

- Training set
- Test set
Figure 5