

# Non-specific Drug Binding in Human Cytochrome P450 3A4: Molecular Docking without Ligand Bias

Luke Czapla<sup>1,2\*</sup>

1. College of Staten Island, The City University of New York (CUNY). Department of Chemistry and Department of Engineering Science and Physics. 2800 Victory Boulevard, Staten Island NY 10314 USA.
2. Rutgers – The State University of New Jersey. BioMaPS Institute for Quantitative Biology. 610 Taylor Road, Piscataway, NJ 08854 USA.

\* Corresponding author: Luke Czapla, [czapla@biomaps.rutgers.edu](mailto:czapla@biomaps.rutgers.edu)

Short Title: CYP3A4-mediated metabolism of substrates

*Molecular simulation is a valuable technique for inferring new hypotheses on the mechanisms of the biological machinery of life. These simulations are based on approximations of the native environment of biological macromolecules that are more accurate representations of the in vivo environment than the conditions under which structural information can be extracted from experimental crystallographic techniques, but conformational sampling remains a major challenge, making results highly speculative. Here, an analysis of a 300 ns Accelerated Molecular Dynamics trajectory of the human cytochrome P450 3A4 enzyme with no drug bound is presented along with two additional independent copies to illustrate the issues with free energy calculations and sampling. The FACTS solvation model is then used to account for the hydrophobic effect in drug binding using molecular docking with the hybrid evolutionary algorithm in EADock DSS with the SwissDock protocol. The results based on minimal assumptions demonstrate that productive binding modes of substrates are highly scored within the top clusters of docked drug poses with the “blind docking” method and thus the approach described here might be automated and extended to other systems. However, the premise that productive binding modes of substrates to human cytochrome P450 must be the very most thermodynamically favorable states may be flawed, as the process of drug metabolism is primarily driven by chemical reactivity and non-specific binding is observed in the results. Simple hypotheses based on observation of the results are presented with reflection on recent advances and the current lack of conclusive experimental evidence on important issues. The simulation data is freely available to the scientific community under the GNU Public License.*

## **Author Summary:**

*Predicting the sites of metabolism of substrates for oxidation by human cytochrome P450 3A4 using computational techniques is a very challenging problem with many approaches. We have combined molecular simulations with drug docking methodologies to provide a new picture of the active site of this protein, suggesting that it may be accessible to a range of substrates and that predictions of biologically relevant sites of metabolism may be obtained without any information about chemical reactivity in the model. The hydrophobic effect may drive non-specific drug binding but there is significant prevalence of productive modes to investigate the results further. Importantly, there is no ligand bias or active site bias in the presented approach. Reflection on the current experimental data in context of the results yields some new hypotheses and simulation data is publicly available for future exploration.*

## **Introduction:**

The human cytochrome P450 (CYP) enzymes are a family of membrane-associated heme-containing proteins intricately involved in the metabolism of xenobiotics<sup>1,2</sup> and related to the bacterial cytochrome P450 enzymes that are not significantly associated with cellular membranes. These enzymes are coupled to other proteins such as cytochrome P450 reductase<sup>3,4</sup> in a complex redox reaction cycle resulting in the oxidative metabolism of drugs within the CYP active site. One of the most important of these enzymes in humans is the cytochrome P450 3A4 protein, which is involved in the oxidative metabolism of approximately 60% of currently FDA-approved drugs.

Many compounds can act as both substrates of CYP3A4 with multiple sites of metabolism (SOM) positioned within their molecular structure, and also as inhibitors, reducing the effective rate of binding and active oxidation of other substrates within the active site. Inducers exist as well, but many bind to auxilliary nuclear receptors that act as transcriptional regulators, like the pregnane X receptor (PXR)<sup>5</sup> and constitutive androstane receptor (CAR)<sup>6</sup>, to effect hepatic induction at the transcriptional level<sup>7</sup>, and thus the role of direct binding of these inducers to CYP is currently an unresolved issue in absence of direct experimental evidence such as high-resolution structural data from X-ray crystallography. The existence of such *in vivo* genetic elements as these nuclear receptors suggests that a full understanding of CYP metabolism involves not only understanding the complex CYP catalytic cycle and coupling between reaction partners, but also other complex networks in the cell.

One important ligand for which detailed structural information is known is bromocriptine (bromoergocryptine or BEC, trade names Parlodel and Cycloset), an ergot alkaloid dopamine agonist used in the treatment of Parkinson's disease, as well as for pituitary tumors, hyperprolactinaemia, neuroleptic malignant syndrome and type II diabetes. It binds to a broad spectrum of neurotransmitter receptors

such as dopamine, serotonin, and alpha adrenergic receptors, and the mechanism of action is primarily believed to be due to its potent dopamine D2 receptor agonism. The published crystallographic structure with bromocriptine bound within the active site of human cytochrome P450 3A4 (PDB ID 3UA1) in an orientation consistent with the hydroxylation at the proline 8' and 9' carbons of the cyclic tripeptide moiety (producing the 8'-mono- and 8',9'-di-hydroxy derivatives reported by *in vivo* CYP3A4 metabolite analysis<sup>8,9</sup>) is the only currently resolved structure of human CYP3A4 with a Type I ligand ligand observed in a productive binding mode<sup>10</sup>. A Type I ligand is one that does not interact with the heme iron protoporphyrin catalytic center of the protein with a complex semi-covalent interaction, whereas Type II ligands do form these complex interactions which can be conceptually understood within the framework of ligand field theory.

So while CYP-mediated metabolism of bromocriptine plays only a minor role in the overall metabolic fate of this drug and it is primarily metabolized in the liver by hydrolysis and isomerization to 2-bromolysergic acid and 2-bromoiso-lysergic acid during first-pass metabolism, the active remainder of the drug is metabolized by CYP3A4 and it is also a very weak competitive inhibitor compared to its roughly 150 picomolar (pM) therapeutic plasma concentration (it has a large volume of distribution), so it makes an ideal target compound in validating docking methodologies for discovering productive poses using protein conformations from molecular simulations.

On the other hand, any attempt to gain insight into CYP3A4 drug binding for a wide range of substrates and inhibitors to draw an independent conclusion requires original approaches. Performing molecular simulations with a drug bound in order to take advantage of an existing “open” active site by removing the drug and docking other drugs to the active site may lead to biased docking results when other target compounds are considered. There may also be significant solvation effects within the active site that are absent under the crystallization conditions, and placing a drug into the active site may thus further limit the conformational flexibility of the protein. Sampling states according to their contribution to the configurational partition function for much simpler model systems proves difficult enough, and greater recognition must be given to the reasonable assumption that the conformation of proteins under standard solution conditions may look quite different than the conformations observed with X-ray crystallography. Thus, the approach is to remove both ligand bias as well as active site bias and allow physical theory to provide an answer without any assumptions.

In the present work, we have performed explicitly solvated Accelerated Molecular Dynamics simulations with a computational model of the human cytochrome P450 3A4 protein without any drug bound to the active site in order to provide a picture of drug binding without any bias towards a particular substrate. These Molecular Dynamics simulations have provided many valuable insights about cytochrome P450 protein structure and function in countless previous studies<sup>11-13 14-18</sup>. Residues 28-499 of the wild-type protein are simulated in a model environment

representing typical solution conditions, but without a model environment of the *in vivo* cellular location on the endoplasmic reticulum membrane. Focus is given to the thermodynamics of the protein and of ligand-protein interactions, in order to evaluate a simple model of drug binding that is dependent on the least assumptions about the system.

An analysis of the simulation is presented, revealing major conformational opening of the active site of the protein, and a single simulated structure is prepared to probe the binding of ligands to the protein with molecular docking. The premise of taking this approach is that conformational sampling with simulations may be deficient and an average structure is sufficient to gain insight into the protein after significant divergence from the crystal structure due to the change in the physical conditions such as temperature, pressure, and solvation environment. Experts in drug docking may want to further evaluate the results and not just look at one conformation. The Accelerated Molecular Dynamics technique is used in this regard as a sampling method to overcome some limitations of time scales and provide a more extensive ensemble.

A hybrid evolutionary algorithm for molecular docking analyzes possible ligand positions within and around the CYP3A4 protein using just a single energy-minimized simulated conformation, in order to evaluate whether the Accelerated Molecular Dynamics simulation technique provides any new insight into substrate and inhibitor positioning within the active site. First, the examples of Type I ligand substrates bromocriptine and progesterone are explored with the docking methodology, to demonstrate the prediction of productive poses from “blind docking” without any specific region of the protein defined to find optimal ligand binding positions. Medically relevant compounds without representative ligand-CYP3A4 crystal structures are then explored with the same methodology to test the predictions of this model, and the results for methadone with its context in medicine and drug toxicity are briefly presented with the results from analysis of many other important compounds that are substrates for human CYP3A4. A table (Table 2) summarizes the author’s results for every compound tested and it suggests that the FACTS continuum solvation model describes the hydrophobic effect in drug binding fairly accurately and positions ligands for metabolism.

Free energy analysis of the active site conformation provides some insight into which conformational states are most likely to be observed thermodynamically in absence of a bound drug, based on an effective radius of gyration as an approximation of accessibility of the active site. This data may form the basis for incorporating representative structures from the entire ensemble of protein conformations predicted by the simulation into the evaluation of ligand binding modes, accounting for the relative statistical probability of states. However, there are always significant errors in the computed potentials of mean force, due to the massive challenges in sampling the configurational partition function of these complex model systems.

The highly ranked productive binding modes from drug docking are among an ensemble of thermodynamically favorable states within the active site according to the estimated binding free energies here, and the simulation suggests that drug binding may be primarily driven by solvation effects due to the success of the FACTS continuum solvation model using the hybrid evolutionary algorithm for docking. Account for chemical reactivity combined with improved thermodynamic estimates from further calculations may help to understand the mechanism of CYP-mediated drug metabolism using transition state theory and equilibrium statistical mechanics, and thus the data has been made available for others to explore.

The results are promising and suggest that any of the conformational states in the simulation could be useful for identifying preferred binding modes within the active site for an entire set of biologically relevant ligands, but in absence of direct experimental evidence any idea that these modes are the very most thermodynamically favorable may be flawed upon reflecting on the mechanism of drug metabolism and the non-specific binding modes observed. A discussion of the docking methodology

## **Methodology:**

### *Computational Procedure*

The apo (with no bound ligand) human cytochrome P450 3A4 system was constructed using the crystallographic coordinates of the structure with Protein Data Bank<sup>19</sup> accession code 1TQN<sup>20</sup>. Missing loop residues 282-285 were built by hand and the simulated protein system consisted of residues 28-499 of the wild-type human microsomal CYP3A4. The rest of the protein was held entirely fixed while the looped residues 282-285 were energy minimized and a short continuum solvent simulation with the Generalized Born model<sup>21</sup> in NAMD 2.9 was performed for 1 nanosecond on the looped residues alone to equilibrate their positions. The system was then solvated in an orthorhombic periodic cell in TIP3P model water (maintaining only the crystallographic water seen coordinating with the heme iron in PDB ID 1TQN), with 10 positive ions (K<sup>+</sup>) and 16 negative ions (Cl<sup>-</sup>) added to neutralize the net system charge and represent an excess salt concentration of approximately 50 mM. The states of titratable amino acid residue side chains were assigned by using the command-line versions of *PropKa*<sup>22</sup> and *pdb2pqr*<sup>23,24</sup> to estimate the effective shift of pKa values due to the electrostatic and solvation environment based on the Poisson-Boltzmann equation and hydrophobic effects and to assign representative CHARMM27 states of the side chains at pH = 7. Relevant non-standard protonation states include the protonation of Asp182, Asp270, and Glu320. All histidine amino acid side chains were electrically neutral, and His28, His54, His267, His287, and His402 were protonated at ND1, while His30, His65, and His324 were protonated at NE2. The CHARMM27 protein parameters with the grid-based CMAP backbone dihedral correction<sup>25</sup> were used along with ion parameters for potassium and chloride from Dmitrii Beglov and Benoit Roux.

The heme parameters were taken from CHARMM19 parameters (from CHARMM c31b1<sup>26</sup>) and the cysteinate residue Cys442 was electrically neutral and bonded to the heme iron with the CHARMM heme to cysteinate patch, representing an Iron(III) oxidation state of the active site heme with a core having a net zero electrical charge and a -2 total charge due to the two carboxylic acid side chains off the porphyrin ring. The heme atomic partial charge parameters were compared to the cysteinate penta-coordinated ferric high-spin heme protoporphyrin RED-III<sup>27</sup> atomic partial charges (T-HM IC6) from work on recently derived improved AMBER heme parameters<sup>28</sup>, and the CHARMM atomic partial charges were determined to be in reasonable agreement (about a 15% difference in the Fe(III) atomic partial positive charge and also on the total negative partial charge of the neighboring penta-coordinated nitrogens and SG sulfur atom of Cys442). The total system consisted of 42,094 atoms with 11,437 water molecules and was assembled in version c35b5 of the CHARMM biomolecular simulation program<sup>26</sup>, while all the simulations were performed with NAMD 2.9<sup>29</sup>.

Throughout all simulations,  $k = 400 \text{ kcal/mol} \cdot \text{\AA}^2$  harmonic constraints ( $U = \frac{1}{2} k |\mathbf{x} - \mathbf{x}_0|^2$ ) were imposed on the SG side chain atom of Cys-442, the iron atom of the active site heme, and also the O atom of the crystallographic water observed coordinating above the heme iron in the active site of PDB ID 1TQN, in order to maintain their relative crystallographic bond lengths and positions  $\mathbf{x}_0$  and to maintain them near the center of the box coordinates throughout the simulations as a convenient reference frame. Initially, 200 kcal/mol·Å<sup>2</sup> harmonic constraints were placed on the non-hydrogen atoms of the protein backbone and 2 kcal/mol·Å<sup>2</sup> harmonic constraints were placed on non-hydrogen atoms of the protein side chains. After 2000 steps of energy minimization with the gradient and line search algorithm in NAMD 2.9<sup>29</sup>, the system was equilibrated for 6 ns in the NPT ensemble. A temperature of 310 K was maintained with the NAMD Langevin dynamics integrator with a damping coefficient of  $\gamma = 5.0 \text{ ps}^{-1}$ , and a pressure of 1 atm was maintained with a modified Nosé-Hoover algorithm that uses Langevin dynamics for barostat fluctuation control. The initial time step was 0.5 fs and was rapidly increased to the final time step of 2 fs over 30000 steps, with hydrogen bond length geometries constrained with the SHAKE algorithm throughout the simulations. After the first 4 ns of equilibration, protein backbone harmonic constraints were reduced to 2 kcal/mol·Å<sup>2</sup> and the protein side chain constraints were removed for 1 ns, and the final 1 ns of equilibration was performed without any harmonic constraints on the protein.

Following this equilibration, the system was first simulated for 50 ns with Langevin dynamics in the NVT ensemble at  $T = 310 \text{ K}$ , using the final volume after the 6 ns of NPT ensemble equilibration, with Accelerated Molecular Dynamics<sup>30</sup> used to enhance the sampling of protein conformational space, using a biased approach for approximating the configurational partition function of the system. The density of the simulated system was 1.049 g/cm<sup>3</sup> and the periodic orthorhombic cell dimensions were 87.2 Å × 78.0 Å × 60.9 Å. Electrostatics were computed with the Particle Mesh Ewald (PME) algorithm with 4<sup>th</sup> order interpolation on a grid with

spacing of approximately 1.0 Å (grid dimensions 88x80x64), a direct sum tolerance of 10<sup>-6</sup>, a direct space cutoff at 12 Å, and a switching function smoothing the potential between 10 Å to 12 Å to zero for both direct space PME electrostatics and non-bonded van der Waals 12-6 interactions.

Following the initial Accelerated Molecular Dynamics run, the entire final system configuration after this 50 ns trajectory was subjected to an additional 2000 step energy minimization to prepare it for analysis of ligand poses with a docking methodology. The structures of ligands such as bromocriptine (bromoergocryptine) and progesterone were obtained from the ZINC database<sup>31</sup> and docked to the active site of CYP3A4 in this final structure using the EADock DSS hybrid evolutionary algorithm<sup>32</sup> (using the SwissDock graphical web interface<sup>33</sup> and a Python SOAP client) without any definition of the region of interest<sup>34</sup>. Docking is performed to the protein with an active site in the reduced Fe(II) heme oxidation using CHARMM27 protein parameters and CHARMM-compatible SwissParam parameters<sup>35</sup> for the ligands that would be suitable for future Molecular Dynamics simulation studies. The relevant Fe(III) heme oxidation state from the simulation is targeted as well, and results are summarized in Table 2 and illustrated in the Figures S1 to S5 in the Supporting Material.

An additional 250 ns of simulation (for a total simulated Accelerated Molecular Dynamics run of 300 ns) was performed to sample additional protein conformations and to improve the estimates of the free energy profile with respect to the conformational accessibility of the active site of cytochrome P450 3A4, as detailed in the Methodology subsections below. The data in Figure 2 could form one basis for the selection of an expansive set of simulation frames representing the conformational landscape of the active site, and the Potential of Mean Force may provide rigorous estimates of the free energies of protein conformational states for ensemble evaluation of ligand binding free energies using the docking methodology presented here. Incorporating additional simulation copies and longer simulation runs in the evaluation of this potential of mean force (PMF) would improve the reported accuracy by more extensively sampling protein conformational space.

The relative RMSD of the energy-minimized structure after the 50-nanosecond Accelerated MD simulation run was 4.4 Å with respect to all resolved non-hydrogen crystallographic coordinates of PDB ID 1TQN. The active site in this single conformation was found to be more open than a number of ligand-bound crystal structures of CYP3A4 in the Protein Data Bank, and the active site residues had an overall radius of gyration ( $R_g$ ) of 10.81 Å as defined in the following subsection. The simulation after an additional 250 ns continued to visit open states of the active site but only explored significant states with a larger active site radius of gyration towards the end of the simulation in the final 100 ns (see Figures 3 and 8).

Finally, two additional independent 300 ns Accelerated MD simulations with the exact same starting conditions were performed to address the error in the potential of mean force. No additional analysis was performed on these trajectories but the data is freely available to the scientific community.

## *Analysis of Crystal Structures and Simulated Protein Conformations*

In order to quantify the state of the active site, amino acid residue side chains within 5 Å of any atom of the bound bromocriptine (bromoergocryptine) ligand in the CYP3A4 structure with Protein Data Bank ID code 3UA1<sup>10</sup> defined the important active site residues, and this set of residues was compared to those within 5 Å of the ritonavir ligand in the ritonavir-CYP3A4 complex with PDB ID 3NXU<sup>36</sup> and found to be 90% similar. The active site residues included Tyr53, Phe57, Asp76, Arg105, Arg106, Phe108, Ser119, Ile120, Arg212, Phe213, Phe215, Leu216, Thr224, Phe241, Ile301, Phe304, Ala305, Thr309, Ala370, Met371, Arg372, Leu373, and Glu374.

The radius of gyration  $R_g = \sqrt{\frac{1}{N} \sum |\mathbf{x}_i - \langle \mathbf{x} \rangle|^2}$  was computed for the  $N$  coordinates of non-hydrogen atoms of both the backbone and side chains of the active site residues to characterize the accessibility or effective volume for substrate and inhibitor binding within the active site. For the apo CYP3A4 crystal structure with PDB ID 1TQN, the calculated active site  $R_g$  value is 9.82 Å, while the bromocriptine-bound and ritonavir-bound crystal structures 3UA1 and 3NXU have an active site  $R_g$  value of approximately 9.81 Å and 10.71 Å. Histograms of the active site radius of gyration as reported by the Colvars module<sup>37</sup> in NAMD at every 2 fs time step during the simulation were Boltzmann weighted using the Accelerated MD boost potential to obtain the Potential of Mean Force (PMF) with respect to  $R_g$  of the active site residues, as reported in the results.

## *Accelerated Molecular Dynamics Potentials*

With the Accelerated Molecular Dynamics (aMD) enhanced sampling technique<sup>38</sup>, a boost potential can be added to various terms in a Molecular Mechanics force field to facilitate barrier crossing and to overcome the difficult problem of sampling states that are far from the initial system configuration and would only be observed at times far beyond the simulated time scales feasible for explicit solvent models using currently available computational resources. In the case of a difficult sampling problem it may not be completely possible to ascertain the accuracy of the approach, but it is an attractive alternative to running a simulation that is clearly deficient in diverging from the initial state.

Here, we applied a boost potential to the total dihedral energy of the system as performed in other simulation studies with the technique<sup>39 40</sup>, and other more sophisticated variants of the method have been used to treat cytochrome P450 enzymes<sup>13</sup>. The boost potential  $\Delta V$  determines the relative weight of each protein configuration in the simulation according to the Boltzmann factor  $\exp(-\beta \Delta V)$ , where  $\beta = \frac{1}{k_B T}$ . The boost potential that is added to the total dihedral potential  $V(\mathbf{x})$

of a configuration with coordinates  $\mathbf{x}$  when it falls below the threshold value  $E$  has the following form:

$$\Delta V(\mathbf{x}) = \begin{cases} 0 & V(\mathbf{x}) > E \\ \frac{(E - V(\mathbf{x}))^2}{\alpha + E - V(\mathbf{x})} & V(\mathbf{x}) < E \end{cases}$$

The value of  $E$  was chosen to be 3600.0 kcal/mol and the value of  $\alpha$  was chosen to be 400.0 kcal/mol based on analysis of the average total dihedral energy of a short 5 ns MD simulation run of the system in the NVT ensemble at 310 K without constraints, which yielded an average total dihedral energy of approximately 2100 kcal/mol for the 472 amino acids in this system composed of residues 28-499 of the wild-type human microsomal CYP3A4. The average boost potential  $\Delta V$  over the course of the first 50 ns Accelerated Molecular Dynamics run was 96.6 kcal/mol with a standard deviation of 13.3 kcal/mol, and in the final 250 ns of Accelerated MD the average was 94.1 kcal/mol with a standard deviation of 13.0 kcal/mol. The average boosted total dihedral potential energy of the system during the complete 300 ns Accelerated MD simulation was approximately 3350 kcal/mol.

Computed for the amino acids resolved in the 1TQN PDB structure, the relative RMSD of the non-hydrogen atoms of the protein backbone had an average of 2.1 Å and a maximum of 2.8 Å, while the relative RMSD of the non-hydrogen atoms of the entire protein had an average of 3.9 Å and a maximum of 4.4 Å over the course of the initial 50 ns Accelerated Molecular Dynamics trajectory. This indicates that the boost potential maintained very reasonable conformations throughout the simulation, and comparable Molecular Dynamics simulations without enhanced sampling of cytochrome P450 enzymes run for 250 ns using the same CHARMM27 protein force field but in a model of the membrane environment based on CHARMM36 lipid parameters (apo CYP3A4, ritonavir-bound CYP3A4, alpha-naphthoflavone bound CYP1A2, and CYP2C9) generally had an average relative RMSD of 2.3 Å and a maximum of 2.6 Å for all the non-hydrogen atoms of the protein as published results for the human aromatase protein. The CYP3A4 Molecular Dynamics simulations without enhanced sampling suggested similar protein conformational flexibility, as reported by the relative RMSD, as results for the human aromatase protein (CYP19A1) in membrane simulations with CHARMM36 lipid parameters and these same CHARMM27 CMAP protein parameters<sup>41</sup>, and so it is speculated that these simulations did not diverge as far from the starting structure over the course of the relatively short 250 ns simulation time to sample states such as those reported here from Accelerated Molecular Dynamics simulation.

The complete 300 ns Accelerated Molecular Dynamics trajectory had a protein non-hydrogen backbone relative RMSD with an average of 2.8 Å and a maximum of 3.9 Å as well as an overall non-hydrogen atom relative RMSD with an average of 4.4 Å with a maximum of 5.2 Å, computed with respect to the amino acid residues 28-281 and 286-499 resolved in PDB ID 1TQN. The protein backbone relative RMSD with respect to human CYP3A4 residues resolved in the crystal structure with Protein

Data Bank accession code 1TQN is illustrated in Figure 8, and large motions of the G' helix, coupled to motions of the F and G helix and neighboring residues, appear to be responsible for the increased backbone RMSD in the final 100 ns at the end of the 300 ns simulation. Figures 1A and 1C illustrate the major conformational rearrangements of the protein at the end of the 300 ns simulation with respect to Figure 1B illustrating the starting protein backbone structure.

#### *Docking biologically-relevant ligands using the hybrid evolutionary algorithm in EADock DSS with solvation theory in FACTS*

The CHARMM Molecular Mechanics force field and FACTS continuum solvation model<sup>42</sup> is used by the EADock DSS algorithm<sup>43</sup> to estimate the ligand pose scores as well as the binding free energies of the docked poses, by evaluating the van der Waals and electrostatic interactions along with polar and non-polar contributions to the solvation free energy of the ligand, protein, and ligand-protein complex. Importantly, the hydrophobic effect is treated with a surface tension term for the solvent accessible surface area with  $\gamma = 0.015 \text{ kcal/mol}^{\circ}\text{Å}^2$ . While this approach in general is just an approximation<sup>44</sup>, it is optimally parameterized for protein cavities and provides a simple treatment of this essential effect without the need for detailed microscopic calculations such as with water Grand Canonical Monte Carlo<sup>45</sup>.

The EADock DSS hybrid evolutionary algorithm in an Accurate setting then optimizes ligand poses using their FullFitness score and these poses are clustered, and the clusters are ranked according to the average FullFitness of the corresponding member poses<sup>46</sup> and only the highest ranked pose of a cluster is considered in the analysis here. A total of 5,000 seeded ligand poses are evaluated and reported with the scores, energies, and corresponding binding free energies during the procedure and the eight corresponding to the top ranked pose of the top eight poses.

An equilibrium dissociation constant  $K_d$  that may give insight into the competitive inhibition constant  $K_i$  for each compound to inhibit substrate metabolism by human CYP3A4 is estimated from the binding free energies for the ligand poses reported by the EADock DSS algorithm. The binding free energy is related to  $K_d$  by the following well-known relationship,  $\Delta G_{\text{binding}} = k_B T \ln K_d$ , where the value  $k_B T$  is equal to 0.616 kcal/mol at the simulated temperature of  $T = 310 \text{ K}$ . Thus, an equilibrium dissociation constant of 1  $\mu\text{M}$  is equal to a binding free energy for a docked ligand pose of approximately -8.5 kcal/mol. The values are not meant to be exact fits to experimental data but to merely demonstrate the predictions of the docking methodology and solvation model with respect to quantitative experimental data and the fit between theory and experiment. An error of 1 kcal/mol corresponds to roughly a factor of five difference in the reported values of  $K_d$  or  $K_i$ .

The highest ranked pose of each of the top eight clusters were analyzed, and the reported free energies  $\Delta G$  from the EADock DSS algorithm for these top ranked poses of the first eight clusters were all within a range of 0.6 kcal/mol or about  $1 k_B T$  for the two ligands bromocriptine and progesterone, suggesting an ensemble of possible ligand conformations within the active site and non-specific binding with prevalence of productive binding modes. The Accelerated MD simulation trajectory and ligand poses from docking to the energy-minimized final 50 ns simulated CYP3A4 conformation were analyzed with the visualization software packages UCSF Chimera<sup>47</sup> and VMD<sup>48</sup>.

## Results:

*The CYP3A4 active site is much more open in the simulated Accelerated Molecular Dynamics solution system than in the apo CYP3A4 crystallographic structures and major conformational rearrangements are observed*

The Accelerated Molecular Dynamics simulation predicts that the active site is more open under realistic solution conditions at 310 K than both the apo crystal structures and many ligand-bound crystal structures, as illustrated in Table 1 and Figure 2. The relative probability of conformations of the active site with a specified  $R_g$  value with the proper configurational Boltzmann weights from the Accelerated MD boost potential is presented as the Potential of Mean Force (PMF) or free energy profile  $A(R_g) = -k_B T \ln p(R_g)$ . The results suggest that active site conformations observed in both the apo crystal structures and in a number of the ligand-bound crystal structures (e.g., bromocriptine, metyrapone) are quite closed compared to the active site in the most probable CYP3A4 protein conformations during the apo CYP3A4 aMD simulation. Figure 2 presents the free energy profile at simulation times after the conformational state has diverged significantly from the starting structure, from analyzing both the first 200 ns and the entire 300 ns of the simulation. The oscillations in the PMF between  $R_g$  values of 10.0 Å and 11.2 Å are speculated to be mostly sampling noise due to the short length of the 300 ns Accelerated Molecular Dynamics simulation trajectory.

Figure 3 shows the time evolution of active site opening and amino acid residue rearrangement, quantified as the active site radius of gyration  $R_g$  defined in the Methodology and also reported for the high-resolution CYP3A4 crystal structures in Table 1. It must be noted that the time trajectory of the Accelerated Molecular Dynamics simulation has no correspondence to the speculative time trajectory of events from standard Molecular Dynamics simulations without enhanced sampling, due to the Accelerated MD technique for boosting the total dihedral energy of the system, and thus the states explored with this method may only manifest in a far longer standard MD simulation than the timescale of the simulation presented here.

In addition, comparison of the protein structure at the beginning of the simulation after equilibration in Figure 1B with the structure at the end of the simulation in

Figure 1C reveals major structural rearrangements of the protein. As noted in Figures 1 and 8, the G' helix of the F'-G' junction and G helix line up end to end and the F and G helices arrange into more of an antiparallel orientation towards the end of the simulation. Significant collective protein motions occur, which suggests that under more realistic simulated conditions the protein is very dynamic and may continue to evolve beyond the short length of the current 300 ns simulation to sample many conformational transitions. These states may contribute significantly to substrate entry kinetics (which may be better evaluated with a model of the membrane present) as well as to the thermodynamics of drug binding.

*Blind docking of bromocriptine to the 50 ns simulated CYP3A4 structure suggests a productive pose consistent with the site of metabolism elucidated in the crystal structure with PDB accession code 3UA1*

As illustrated in Figure 4, we identified a binding mode from blind docking to the energy-minimized CYP3A4 structure after the initial 50 ns of Accelerated Molecular Dynamics simulation that is consistent with the site of metabolism, as observed in the high-resolution X-ray crystallographic structure of the bromocriptine-CYP3A4 complex with PDB accession code 3UA1, exposed to the catalytic center of the heme. In analyzing the top poses of the first eight clusters in the results from the EADock DSS evolutionary algorithm for bromocriptine, we found that all of these top eight binding modes showed bromocriptine within the active site of cytochrome P450 3A4, but five had the substrate positioned close to the active site heme (ligand non-hydrogen atom distances less than 6 Å).

The productive pose is the second ranked of those with close positioning of a site of the substrate to the heme center for possible oxidative metabolism with the CYP3A4 in an oxidized form during the catalytic cycle, with a distance between the 9' proline carbon of the tricyclic peptide moiety and heme iron of 3.3 Å. The top pose of the first cluster, which is the highest ranked pose consistent with a substrate positioned for oxidation, displays a binding mode consistent with N-demethylation at the quinoline 6-methyl of the lysergic acid moiety (N6) with the 6-methyl carbon at a distance of 4.2 Å from the heme iron, an oxidation reaction that is not observed experimentally in studies of CYP3A4-mediated metabolism of bromocriptine.

Our simple methodology for examining substrate and inhibitor positioning does not incorporate the reactivity of substrate sites of metabolism at all, and it is possible that an entire ensemble of ligand binding modes are explored within the active site during each *in vivo* ligand binding event. These non-productive modes observed in our blind docking studies, such as this binding mode consistent with positioning of the lysergic 6-methyl of bromocriptine for N-demethylation, may be suggestive of actual states explored by a bound ligand during an encounter with the CYP3A4 active site that simply do not have the chemical reactivity or favorably low activation energy to form significant products later in the catalytic cycle.

Whether low energy barriers connect the ensemble ligand positions from the top docked poses during a mean-field encounter with the single conformational state of the protein analyzed is not examined in this work, but the occurrence of some clustered docked states with the molecule far from the active site heme iron suggest hypothetical bromocriptine binding to the catalytic center in this frame is not particularly tight, and the active site appears to allow for much conformational rearrangement of the ligands. The concept of “conformational selection”<sup>49</sup> may be particularly useful here for understanding the effects of bromocriptine binding to the active site of human CYP3A4, but further analysis would be required to investigate whether it is important to the binding of the ligands analyzed here.

Compared to the in vitro assessment of drug interactions in the Cycloset (bromocriptine) FDA Prescribing Information, a very approximate  $K_i$  value for competitive inhibition of CYP3A4 by bromocriptine 1.5  $\mu\text{M}$  (10,000 times the therapeutic concentration), but a spectral dissociation constant  $K_s$  of 0.37  $\mu\text{M}$  has also been reported from analysis of bromocriptine binding to a truncated form of the wild-type human microsomal CYP3A4<sup>10</sup>. The productive binding mode of bromocriptine as represented by the top pose of cluster #4 has a  $\Delta G_{\text{binding}}$  of -9.22 kcal/mol from analysis with the FACTS continuum solvation model, corresponding to a  $K_d$  value of 0.32  $\mu\text{M}$ , while the top pose of cluster #1 with a non-productive hypothetical N-demethylation mode has a corresponding  $\Delta G_{\text{binding}}$  of -9.80 kcal/mol and a  $K_d$  value of 0.12  $\mu\text{M}$ . The lowest estimate from the top pose of these first eight clusters is a  $\Delta G_{\text{binding}}$  of -9.84 kcal/mol.

These binding free energies of -9.84 kcal/mol to -9.16 kcal/mol for the top 8 clustered poses from FACTS continuum solvation analysis match quite well to the high affinity of the substrate and its competitive inhibition. A lowest binding free energy for the top poses of -9.89 kcal/mol is observed from docking to this same structure in the Iron(III) heme oxidation state (Table 2 and Supporting Material Figure S1), but the thermodynamic contributions from factors such as displacement of the bound water and conversion of the heme to a high spin state during the titration of bromocriptine is not certain, nor is the specific conditions upon which this high spin conversion would manifest. Analysis of CYP3A4 nanodisc experiments suggests that the human CYP3A4 at realistic conditions exists in a state where there is 11% high spin with no drug bound<sup>50</sup>, and thus less than  $-k_B T \ln (0.11/0.89)$  or 1.29 kcal/mol may represent a reasonable cost for displacing the bound water and inducing a high spin, and this water may not be very tightly bound in the Fe(III) heme oxidation state.

In addition, experimental estimates of the increased spectral dissociation constant  $K_s$  (loss of binding affinity) in an R212A truncated human CYP3A4 mutant suggest a significant contribution to favorable bromocriptine binding from this residue<sup>10</sup>, and a direct Arg212 interaction with the ligand is not observed in the productive binding mode found in the docking result for the single simulated structure here. Protein conformational states from frames of the simulation where the Arg212 residue is better positioned to interact with the cyclic tripeptide moiety of bromocriptine

could be hypothesized to have a more favorable  $\Delta G_{\text{binding}}$  for observed productive binding modes reported by the docking algorithm, and it is possible that this interaction may assist in maintaining the ligand with the site of metabolism exposed near the catalytic center during oxidation.

Table 2 and Supporting Material Figure S1 shows the hypothetical productive pose from docking to this simulated structure with the heme moiety in the Fe(III) heme state, and the hypothetical productive pose moves up to the top pose of cluster #2, with a non-productive binding mode with the valine side chain of the cyclic tripeptide positioned for metabolism as the top pose of cluster #1. The state analyzed in this treatment of docking to the Iron(III) state may be more relevant to the metabolism of these substrates, since it is further along in the CYP catalytic cycle and may drive the process forward.

*Blind docking of progesterone to the same simulated CYP3A4 structure suggests productive binding modes consistent with progesterone sites of metabolism*

As illustrated by Figure 5, the blind docking approach using the same energy-minimized structure after 50 ns of Accelerated Molecular Dynamics simulation used for docking bromocriptine predicts that progesterone is positioned within the active site of cytochrome P450 3A4 in all of the top eight clusters, in contrast to the progesterone binding mode observed in the crystallographic structure with PDB ID 1W0F<sup>51</sup>. In the high resolution structure of the progesterone-CYP3A4 complex, the progesterone molecule was not crystallized within the active site, but rather the ligand is observed associated with the hydrophobic protein surface that positions the core of the protein within the membrane of the endoplasmic reticulum and is likely involved in the entry and exit of ligands into the CYP3A4 active site.

It is known that progesterone has complex metabolic interactions with many CYP enzymes as well as other cellular elements, and that hydroxylation at the 16 $\alpha$ , 6 $\beta$ , and 2 $\beta$  carbon positions results in the major CYP3A4 oxidation products of 16 $\alpha$ -, 6 $\beta$ -, and 2 $\beta$ -hydroxyprogesterone molecules<sup>52</sup>. In addition, CYP2C9 yields an additional 21-hydroxyprogesterone as the major product of progesterone oxidation, with the 16 $\alpha$ - and 17 $\alpha$ -hydroxyprogesterone as minor products. Cooperative binding may also play some role in the process of CYP-mediated metabolism.

The first relevant binding mode observed in docking with the EADock DSS algorithm and consistent with the sites of metabolism of progesterone is the top pose of the first cluster as illustrated in Figure 5A. The progesterone ligand here is orientated in a productive pose for the 16 $\alpha$ -hydroxylation observed experimentally as a major product of the CYP3A4-mediated metabolism of progesterone, and relevant hydrogen-bonded interactions with the Ser119 side chain and with the backbone amide of the Glu374 are illustrated in the Figure. The distance between the 16 $\alpha$  carbon and the heme iron is 3.4 Å. In addition, the top pose of the third cluster has progesterone positioned with the aliphatic C21 at a distance of 3.4 Å from the heme

iron for the 21-hydroxylation that is the major product of CYP2C9-mediated progesterone metabolism. Finally, in the top pose of the eighth cluster, the  $2\beta$  carbon hydrogen is at a distance of 2.4 Å from the heme iron (a  $2\beta$  carbon-iron distance of 3.1 Å), positioning the progesterone molecule for the  $2\beta$ -hydroxylation consistent with CYP3A4-mediated metabolism of progesterone as illustrated in Figure 5B.

In addition, a residue believed to be essential for cooperative hydroxylation of progesterone, Phe304<sup>53</sup>, is observed to be fairly distant from the region of progesterone binding in these binding modes revealed from docking to the single simulated CYP3A4 structure analyzed here. The difference in reported free energies between the top poses of Clusters #1 and Clusters #8 is 0.52 kcal/mol, ( $\sim 0.8 k_B T$ ), suggesting that these analyzed states are among an ensemble of possible binding modes and that the relative occurrence of this  $2\beta$ -hydroxylation binding mode from Cluster #8 is approximately 40% of the occurrence of the  $16\alpha$ -hydroxylation binding mode from Cluster #1, based on the approximate free energy analysis using the FACTS continuum solvation model in EADock DSS. However, these groups have different chemical reactivities, making a thorough analysis of the oxidation products unattainable with the models presented here.

The top docked pose of cluster #1 has a  $\Delta G_{\text{binding}}$  of -8.1 kcal/mol, suggesting an equilibrium dissociation constant  $K_d$  of 1.9 μM compared to an experimental competitive inhibition constant of 0.9 μM<sup>2</sup>. This suggests that there is some reasonable agreement between the two from the limited analysis performed here. In addition, progesterone is a known allosteric effector of CYP3A4<sup>51, 54</sup> and the simple model here does not appear to predict such a binding mode and allosteric effects from drug binding to human CYP3A4 are not addressed by this present study. The presence of two out of the top eight clustered poses with progesterone quite distant (ligand non-hydrogen atoms at a distance greater than 6 Å) from the heme iron catalytic center suggests that the positioning for metabolism here is not particular tight, similar to the results for bromocriptine from docking to this single simulated structure.

In addition, as also observed in the docking results for bromocriptine, there is no overwhelming thermodynamic favorability for the poses that may represent productive binding modes with respect to the non-productive modes in the top clustered poses. Table 2 shows the results from docking progesterone to this same structure when the heme is in the Iron(III) oxidation state instead of the Iron(II) state considered here.

*Blind docking of Type II ligand erythromycin to the same simulated CYP3A4 structure reveals putative peripheral binding sites on the surface of the protein*

As illustrated in Figure 6, docking to this same energy-minimized simulated CYP3A4 structure from the end of 50 ns of Accelerated Molecular Dynamics simulation leads

to top poses of the first eight clusters where erythromycin is bound to the surface of the protein instead of within the active site. Because the active site radius of gyration  $R_g$  value for this particular simulated structure is smaller than the value seen in the erythromycin-CYP3A4 crystal structure with PDB ID 2J0D (10.81 Å for the simulated apo CYP3A4 conformation vs. 10.85 Å for the active site in 2J0D), accommodating erythromycin is quite difficult, as the ligand-CYP3A4 crystal structure complexes appear to have a very compact active site to begin with. Protein Data Bank structure with ID 2V0M revealing two bound ketoconazole molecules within the active site, one having the direct coordination to the heme characteristic of a Type II ligand inhibitor, has the largest active site of the ones analyzed here with  $R_g = 11.15$  Å. In addition, cooperative binding to CYP3A4 has been observed for a number of important endogenous substrates<sup>55</sup>, and thus energetically favorable open active site conformations with  $R_g$  values around 11.28 Å and larger (observed in Figure 2) may be extremely important for understanding the binding of large ligands as well as the cooperative binding of multiple molecules within the active site.

From analysis of the data in Figure 2, simulated CYP3A4 conformations with an active site that is comparably open to the largest active site observed in the crystal structures of the ligand-CYP3A4 complexes (according to the metric used here) are quite thermodynamically favorable according to the calculated potential of mean force or free energy profile with respect to the active site radius of gyration  $R_g$ . The blind docking approach predicts this simulated CYP3A4 structure where the active site is too closed has little affinity for erythromycin, and docking to the simulated structure at the end of the 300 ns Accelerated MD simulation where the active site appears much more open (according to the measured radius of gyration) reveals a binding mode consistent with the site of metabolism for N-demethylation of erythromycin (Supporting Material Figure S6). Thus, the quantitative estimation of competitive inhibition for this ligand with the FACTS continuum solvation model using docking results would need to incorporate simulation frames with larger values of the active site  $R_g$  and also account for the relative free energies of the conformational states using data like Figure 2, with relatively low contributed weights of docked states from simulated protein conformations with  $R_g \lesssim 11$  Å, and this could be understood as a form of conformational selection. In addition, as a Type II ligand, erythromycin also displays a quasi-irreversible inhibitory binding mode in PDB 2J0D, which cannot be treated within the framework of the models presented here.

*Application of the blind docking method to medicinal compounds to the same structure reveals a possible productive N-demethylation binding mode of R-(*-*)-methadone as the top pose of cluster #3*

An effective treatment for opioid dependence with almost 50 years of research into the treatment of patients with this condition is methadone or (RS)-6-

dimethylamino-4,4-diphenyl-3-heptanone, originally developed in 1937 by German scientists and used as an opioid painkiller and later explored to find beneficial properties in opioid dependency in 1964 at Columbia University Medical Center during a heroin epidemic in New York City<sup>56</sup>. The levorotary R-(*-*)-methadone has strong properties as an opioid mu receptor agonist and the dextrorotatory S-(+)-methadone has additional properties as an NMDA receptor antagonist, and the combination of the two may account for its properties in addiction treatment. The compound is metabolized primarily by cytochrome P450 3A4, 2D6, and 2B6 (as well as a number of other CYPs) and has some reported mixed inhibition properties for CYP3A4. Biological variability in CYP3A4 activity among individuals, as well as variability in other cytochrome P450 enzymes involved in methadone clearance, accounts for the large differences in methadone bioavailability amongst patients in Methadone Maintenance Treatment (MMT) programs as well as for drug toxicity, particularly in patients prescribed the medication as an analgesic. Recent molecular modeling studies have probed the effect of different human cytochrome P450 alleles on the metabolism of this drug<sup>57</sup>, and here we provide a simple analysis of a relevant enantiomer to extend the methodology to compounds not seen in crystallographic structures of human CYP3A4.

We were able to identify a putative productive N-demethylation binding mode of R-(*-*)-methadone to human cytochrome P450 3A4 as the top pose of the third cluster from blind docking with EADock DSS using the energy-minimized frame after 50 ns of simulation, as identified in Figure 7. The first two clusters reveal non-productive modes where each phenyl rings of the compound have an aromatic carbon approximately 3.5 Å from the central iron of the heme moiety in each of the two top clustered poses. The distance between one of the carbons of the 6-amino group and the iron at the center of the heme in the productive binding mode is 5.0 Å and a hydrogen-iron distance of 4.0 Å is identified in the Figure illustrating this top pose of cluster #3.

The corresponding binding free energy of this binding mode is  $\Delta G_{\text{binding}} = -7.26$  kcal/mol, corresponding to a  $K_d$  value of 7.6 μM. However, due to possible Type II ligand interactions resulting in quasi-irreversible inhibition, it may not be possible to conclude anything definite from such quantitative results, and one study reports a  $K_i$  value of 100 μM for racemic methadone inhibition of CYP3A4-mediated nifedipine oxidation<sup>58</sup> but there appears to be a lack of extensive experimental data from other independent measurements. In addition, the N-demethylated metabolite may then undergo spontaneous cyclization to the product EDDP (2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine), complicating the analysis of the oxidation products.

*Application of the docking methodology using the same simulated structure in the Iron(III) heme oxidation state for twelve medicinal compounds reveals many binding modes consistent with sites of metabolism, with summary of results*

Our results suggest that the docking methodology used here with the single simulated structure is very successful in identifying sites of metabolism for medicinal compounds by considering the ligand positioning alone, and the application of this method to predicting possible sites of metabolism for new compounds may be very promising in combination with estimates for the chemical reactivity of the sites of metabolism of these compounds. We have looked at a number of other important drugs and were able to identify productive binding modes consistent with sites of metabolism among the top poses of the first eight clusters. Supporting Material Figures S1 through S5 illustrate putative productive binding modes of bromocriptine, midazolam, oxycodone, carbamazepine, and escitalopram from top poses revealed by docking to the Iron(III) heme oxidation state instead of the Fe(II) state considered in the docking results presented above, and Table 2 summarizes the finding for many drugs from docking to this Iron(III) state that is critical to the mechanism of substrate metabolism. Erythromycin again does not fit into the active site of this same structure, but Supporting Material Figure S6 shows the putative productive binding mode of erythromycin from docking to the simulated structure at the end of the complete 300 ns of Accelerated MD.

In summary, the human CYP3A4 active site appears to easily accommodate many drugs in the quite open state examined here in the single simulated structure. Productive binding modes of substrates may be thermodynamically favorable orientations among a complex landscape of binding modes that are ultimately driven by chemical reactivity in the process of oxidation, since there is no significant separation between the productive binding modes and the non-productive modes in these top clustered poses. It is possible that these non-productive modes observed may also induce a high-spin state and drive forward the CYP catalytic cycle, and the process of oxidation may be quite a complex dynamical process for different substrates. Further work is necessary to clarify the process for different substrates and the effect of ligand binding on the active site. The effect of decreased ligand entropy upon binding, which is neglected in the  $\Delta G_{\text{binding}}$  calculations using the FACTS continuum solvation model, appears to be minimal in binding to this quite open active site conformation for many of the ligands analyzed. Finally, significant conformational rearrangements of the entire protein are observed over the course of the 300 ns Accelerated Molecular Dynamics simulation, and extending the simulation to continue conformational sampling may reveal even greater insights into possible conformational transitions of human CYP3A4, as predicted by simulations using classical Molecular Mechanics force fields.

## Discussion:

The positioning of substrates and inhibitors in the active site of CYP3A4 is a challenging problem that is coupled to a number of other extremely difficult problems across many disciplines in predicting the metabolic patterns of cytochrome P450 enzymes. The hydrophobic effect may be the primary driving force for substrate positioning, as accounted for by the FACTS continuum solvation

model and the EADock DSS FullFitness scoring procedure. By evaluating only the top pose of a cluster as scored by the algorithm, it is possible to find significant correlation between ligand poses and productive binding modes. While the author of this work is not a drug docking expert, a physics-based approach as presented here may be a future direction for others in the field. In the proposed methodology, ligand bias would be removed and the laws of thermodynamics and statistical mechanics can address certain issues involved in protein-ligand binding.

By adding a boost potential to the total dihedral potential energy of the system, a bias which necessitates Boltzmann reweighting, but not one that makes any assumption that the active site looks any more open or closed under the simulated solution conditions than in the original crystallographic coordinates resolved under the crystallization conditions, we were able to definitely conclude that the active site in Accelerated Molecular Dynamics simulations at typical solution conditions looks quite a bit more open than in the apo crystal structures and also in the ligand-bound CYP3A4 crystal structures of many compounds such as bromocriptine, progesterone, and metyrapone.

In addition, significant evolution in the protein structure is observed over the course of the simulation that is not observed in the simulations without enhanced sampling, and these changes may give insight into dynamical states that are relevant to the biology of these enzymes, but like all simulation work, it is entirely speculative. It is possible the boost potential may induce partial unfolding, but this may in fact help the system explore the phase space of the system to converge to the correct distribution. Nobody knows what these proteins really look like outside of the laboratory of crystallization experiments and from indirect structural information deduced from other approaches such as solution NMR, and protein bicelles may make an ideal model system for such studies. There is a balance of entropic and enthalpic forces and it is possible that some proteins are much more condensed in the crystal form under the experimental conditions.

The drug binding mode analyses from the docking methodology using the hybrid evolutionary algorithm with FACTS continuum solvation model in EADock DSS suggests that there is an entire landscape of thermodynamically favorable ligand positions within the active site of the protein. The results from the top scored poses of the first eight clusters gives insight into possible orientations of the ligands that could represent productive binding modes during the process of metabolism. While these putative productive binding modes are thermodynamically favorable, they do not appear to separate at all from non-productive binding modes in the analysis reported here. Thus the mechanism of positioning and signaling for metabolism appears to be very complex and the consideration of additional factors such as chemical reactivity should improve any model of substrate metabolism. The poses obtained for drug docking could also be integrated in quantum mechanical models based on transition state theory.

Bromocriptine, an important drug for which detailed experimental structural data is available for the productive binding mode for 8'- and 9'-hydroxylation, appears to

have highly scored docked poses consistent with this site of metabolism from docking to different oxidation states of the heme in Iron(II) and Iron(III). In addition, as the active site appears quite compact in the bromocriptine-CYP3A4 Protein Data Bank structure with accession code 3UA1, it may fit into many active site states in the simulation.

The thermodynamic estimates from docking to just a single state visited in the simulation may be quite reasonable, and future work could investigate more of the factors stabilizing this productive binding mode for oxidation during the complex catalytic cycle. Continuum solvation approaches techniques such as water and molecule Grand Canonical Monte Carlo may be used to investigate

There is reasonable quantitative agreement between theoretical estimates for the binding free energies of bromocriptine and progesterone and their experimental competitive inhibition equilibrium constants, and whether this is just by chance may be studied in future analyses including other Type I ligands. The theoretical binding free energy estimates from FACTS continuum solvation analysis in EADock DSS compared to the experimental values for bromocriptine and progesterone from their competitive inhibition equilibrium constants (and spectral dissociation constants, when available) are quite similar, and further study is required to analyze additional contributions from other states in the simulation. The ensemble of ligand binding modes may also give insight into states along a transition pathway during the complex process of positioning during CYP-mediated drug metabolism, and it is possible that the docking analysis using the single conformational state here is a reasonable first approximation of the thermodynamics of drug binding.

Molecular simulations combined with ligand analysis may provide detailed insights into the contribution of this positioning to metabolism, by incorporating not only ensemble binding modes of the ligand but possibly also examining CYP3A4-ligand interactions with explicit solvent simulations and other techniques for detailed analysis of the observed binding mode. The combination of these approaches with models for substrate chemical reactivity may give greater insight into the mechanism of drug metabolism, since there appears to not be a significant thermodynamic separation in estimated  $\Delta G_{\text{binding}}$  values between observed binding modes that may be productive and those which appear to be nonproductive, and the positioning alone cannot account for the entire process.

Representative modeling of the *in vivo* membrane environment is expected to play an important role in simulated active site dynamics as well, but as the validation of ligand binding modes depends on experimental structural evidence that is currently most detailed in high-resolution X-ray crystallographic structures observed outside of the biological lipid bilayer environment of cellular organelles such as the endoplasmic reticulum, the focus of the present study on ligands observed in these crystal structures with a simple solution model is essential for developments leading to incorporating lipid bilayers in future work and deciphering any subtle changes to binding induced by the change in environment. The influence of the membrane may be most pronounced in the kinetics of substrate entry and in the

accessibility of substrate entry and exit channels that are exposed to the membrane or near the membrane, and also on hydrophobic membrane-associated amino acid residues (there is a particular abundance of phenylalanine residues near the F and G helices at the putative membrane interface of CYP3A4) that reorient within the membrane environment<sup>14</sup>.

Future work will consider more of the biologically relevant factors for CYP-mediated metabolism to improve estimates of ligand-positioning thermodynamics, and more precise models accounting for atomic polarizability<sup>59</sup> as well as complex semi-covalent interactions of the active site heme with water and type II ligands during various stages of the catalytic cycle (based partly on atomic partial charges derived from Quantum Mechanical calculations<sup>28</sup> and even the use of QM/MM methodologies<sup>60</sup>) will also improve the quantitative accuracy of this approach and other approaches<sup>61</sup>. As the Type I ligands bromocriptine and progesterone do not form these complex semi-covalent interactions with reduced ferrous Fe(II) and ferric Fe(III) site oxidation heme states in the CYP catalytic cycle, these compounds were chosen as ideal candidates for the initial docking study with the models presented here.

The use of Potential of Mean Force calculations within Accelerated Molecular Dynamics simulations demonstrates one interesting solution to the problem of understanding the conformation of the human CYP3A4 active site for substrate and inhibitor binding, and the method developed here is general and may be extended to other cytochrome P450 enzymes such as CYP2C9, CYP2D6<sup>62</sup> and CYP1A2<sup>63</sup> and to the study of allosteric effects due to drug binding in these systems. It may also be used to investigate entirely different protein systems<sup>64</sup> to better understand conformational changes due to the binding of ligands to a receptor and the resulting influence on complex biological networks<sup>65 66</sup>. The Accelerated Molecular Dynamics enhanced sampling technique is able to sample significant conformational transitions of the human CYP3A4 enzyme over the course of a fairly short simulated time scale, and the structure of the enzyme and its active site as predicted by the model is significantly different than the more compact state observed in the initial crystal structure used for the modeling here.

Ultimately, the development of mathematical models utilizing results from both experiments and computational modeling on multiple length scales will resolve some of the massive challenges surrounding important scientific issues such as those presented by the *in vivo* metabolism of drugs by cytochrome P450 enzymes. While the results presented here in no way solve the major challenges ahead, they suggest that these molecular simulations may have predictive value in estimating the contribution of ligand positioning to the mechanisms of CYP-mediated metabolism and to the inhibition of this metabolism. This positioning may be a major factor in the oxidative process, based on evaluating the substrate docking results and observing the prevalence of thermodynamically favorable binding modes consistent with known sites of metabolism. Additional ensemble-based docking evaluation of drug binding modes with simulation results such as those

obtained here, in combination with methods for understanding substrate chemical reactivity, may add new insights into ligand positioning and oxidation during the complex process of substrate metabolism within the active site of cytochrome P450 3A4.

#### **ACKNOWLEDGEMENTS:**

The computer simulations and analyses reported here were supported by time on the TACC Stampede supercomputer through XSEDE Startup TG-MCB140034 to Luke Czapla, as well as by time on the CUNY HPC and Rutgers BioMaPS computational clusters. Luke Czapla thanks Maria Kontogianni at SIUE and Tom Poulos at UC Irvine for helpful advice with the simulations, Aurélien Grosdider at the Swiss Institute of Bioinformatics for generous support with EADock DSS and SwissDock, and Arben Jusufi at CSI CUNY for support and postdoctoral funding.

#### **REFERENCES:**

1. de Montellano, P. R. O., *Cytochrome P450: Structure, Mechanism, and Biochemistry*. Springer: 2005.
2. Sevrioukova, I. F.; Poulos, T. L., Understanding the mechanism of cytochrome P450 3A4: recent advances and remaining problems. *Dalton Transactions* **2013**, 42, (9), 3116-3126.
3. Wang, M.; Roberts, D. L.; Paschke, R.; Shea, T. M.; Masters, B. S.; Kim, J. J., Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc Natl Acad Sci U S A* **1997**, 94, (16), 8411-6.
4. Sevrioukova, I. F.; Li, H.; Zhang, H.; Peterson, J. A.; Poulos, T. L., Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci U S A* **1999**, 96, (5), 1863-8.
5. Cheng, Y.; Redinbo, M. R., Activation of the human nuclear xenobiotic receptor PXR by the reverse transcriptase-targeted anti-HIV drug PNU-142721. *Protein Science* **2011**, 20, (10), 1713-1719.
6. Xu, R. X.; Lambert, M. H.; Wisely, B. B.; Warren, E. N.; Weinert, E. E.; Waitt, G. M.; Williams, J. D.; Collins, J. L.; Moore, L. B.; Willson, T. M.; Moore, J. T., A Structural Basis for Constitutive Activity in the Human CAR/RXR $\pm$  Heterodimer. *Molecular cell* **2004**, 16, (6), 919-928.
7. Dickins, M., Induction of Cytochromes P450. *Current Topics in Medicinal Chemistry* **2004**, 4, (16), 1745-1766.
8. Maurer, G.; Schreier, E.; Delaborde, S.; Loosli, H. R.; Nufer, R.; Shukla, A. P., Fate and disposition of bromocriptine in animals and man. I: structure elucidation of the metabolites. *Eur J Drug Metab Pharmacokinet* **1982**, 7, (4), 281-92.
9. Maurer, G.; Schreier, E.; Delaborde, S.; Nufer, R.; Shukla, A. P., Fate and disposition of bromocriptine in animals and man. II: Absorption, elimination and metabolism. *Eur J Drug Metab Pharmacokinet* **1983**, 8, (1), 51-62.

10. Sevrioukova, I. F.; Poulos, T. L., Structural and mechanistic insights into the interaction of cytochrome P4503A4 with bromoergocryptine, a type I ligand. *J Biol Chem* **2012**, 287, (5), 3510-7.
11. Hendrychova, T.; Anzenbacherova, E.; Hudecek, J.; Skopalik, J.; Lange, R.; Hildebrandt, P.; Otyepka, M.; Anzenbacher, P., Flexibility of human cytochrome P450 enzymes: molecular dynamics and spectroscopy reveal important function-related variations. *Biochim Biophys Acta* **2011**, 1814, (1), 58-68.
12. Shahrokh, K.; Cheatham, T. E., 3rd; Yost, G. S., Conformational dynamics of CYP3A4 demonstrate the important role of Arg212 coupled with the opening of ingress, egress and solvent channels to dehydrogenation of 4-hydroxy-tamoxifen. *Biochim Biophys Acta* **2012**, 1820, (10), 1605-17.
13. Markwick, P. R.; Pierce, L. C.; Goodin, D. B.; McCammon, J. A., Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam. *J Phys Chem Lett* **2011**, 2, (3), 158-164.
14. Denisov, I. G.; Shih, A. Y.; Sligar, S. G., Structural differences between soluble and membrane bound cytochrome P450s. *J Inorg Biochem* **2012**, 108, 150-8.
15. Baudry, J.; Li, W.; Pan, L.; Berenbaum, M. R.; Schuler, M. A., Molecular docking of substrates and inhibitors in the catalytic site of CYP6B1, an insect cytochrome p450 monooxygenase. *Protein Eng* **2003**, 16, (8), 577-87.
16. Yuki, H.; Honma, T.; Hata, M.; Hoshino, T., Prediction of sites of metabolism in a substrate molecule, instanced by carbamazepine oxidation by CYP3A4. *Bioorg Med Chem* **2012**, 20, (2), 775-83.
17. Kuhn, B.; Jacobsen, W.; Christians, U.; Benet, L. Z.; Kollman, P. A., Metabolism of sirolimus and its derivative everolimus by cytochrome P450 3A4: insights from docking, molecular dynamics, and quantum chemical calculations. *J Med Chem* **2001**, 44, (12), 2027-34.
18. Li, W.; Liu, H.; Luo, X.; Zhu, W.; Tang, Y.; Halpert, J. R.; Jiang, H., Possible pathway(s) of metyrapone egress from the active site of cytochrome P450 3A4: a molecular dynamics simulation. *Drug Metab Dispos* **2007**, 35, (4), 689-96.
19. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000**, 28, (1), 235-42.
20. Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F., The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. *J Biol Chem* **2004**, 279, (37), 38091-4.
21. Tanner, D. E.; Chan, K. Y.; Phillips, J. C.; Schulten, K., Parallel Generalized Born Implicit Solvent Calculations with NAMD. *J Chem Theory Comput* **2011**, 7, (11), 3635-3642.
22. Rostkowski, M.; Olsson, M. H.; Sondergaard, C. R.; Jensen, J. H., Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct Biol* **2011**, 11, 6.
23. Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A., PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* **2007**, 35, (Web Server issue), W522-5.

24. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **2004**, 32, (Web Server issue), W665-7.
25. Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* **2004**, 25, (11), 1400-15.
26. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J Comput Chem* **2009**, 30, (10), 1545-614.
27. Dupradeau, F. Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys* **2010**, 12, (28), 7821-39.
28. Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 3rd, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J Comput Chem* **2012**, 33, (2), 119-33.
29. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005**, 26, (16), 1781-802.
30. Wang, Y.; Harrison, C. B.; Schulten, K.; McCammon, J. A., Implementation of Accelerated Molecular Dynamics in NAMD. *Comput Sci Discov* **2011**, 4, (1).
31. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G., ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling* **2012**, 52, (7), 1757-1768.
32. Grosdidier, A.; Zoete, V.; Michielin, O., Fast docking using the CHARMM force field with EADock DSS. *J Comput Chem* **2011**.
33. Grosdidier, A.; Zoete, V.; Michielin, O., SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res* **2011**, 39, (Web Server issue), W270-7.
34. Grosdidier, A.; Zoete, V.; Michielin, O., Blind docking of 260 protein-ligand complexes with EADock 2.0. *J Comput Chem* **2009**, 30, (13), 2021-30.
35. Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O., SwissParam: a fast force field generation tool for small organic molecules. *J Comput Chem* **2011**, 32, (11), 2359-68.
36. Sevrioukova, I. F.; Poulos, T. L., Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. *Proc Natl Acad Sci U S A* **2010**, 107, (43), 18422-7.
37. Fiorin, G.; Klein, M. L.; Hénin, J., Using collective variables to drive molecular dynamics simulations. *Molecular Physics* **2013**, 1-18.

38. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys* **2004**, 120, (24), 11919-29.
39. Hamelberg, D.; Shen, T.; McCammon, J. A., Phosphorylation Effects on cis/trans Isomerization and the Backbone Conformation of Serine–Proline Motifs: Accelerated Molecular Dynamics Analysis. *Journal of the American Chemical Society* **2005**, 127, (6), 1969-1974.
40. Wereszczynski, J.; McCammon, J. A., Using Selectively Applied Accelerated Molecular Dynamics to Enhance Free Energy Calculations. *J Chem Theory Comput* **2010**, 6, (11), 3285-3292.
41. Park, J.; Czapla, L.; Amaro, R. E., Molecular Simulations of Aromatase Reveal New Insights Into the Mechanism of Ligand Binding. *Journal of Chemical Information and Modeling* **2013**, 53, (8), 2047-2056.
42. Haberthur, U.; Caflisch, A., FACTS: Fast analytical continuum treatment of solvation. *J Comput Chem* **2008**, 29, (5), 701-15.
43. Zoete, V.; Grosdidier, A.; Cuendet, M.; Michelin, O., Use of the FACTS solvation model for protein–ligand docking calculations. Application to EADock. *Journal of Molecular Recognition* **2010**, 23, (5), 457-461.
44. Chandler, D., Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, 437, (7059), 640-7.
45. Woo, H. J.; Dinner, A. R.; Roux, B., Grand canonical Monte Carlo simulations of water in protein environments. *J Chem Phys* **2004**, 121, (13), 6392-400.
46. Grosdidier, A.; Zoete, V.; Michelin, O., EADock: Docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins: Structure, Function, and Bioinformatics* **2007**, 67, (4), 1010-1025.
47. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, 25, (13), 1605-12.
48. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996**, 14, (1), 33-8, 27-8.
49. Munz, M.; Hein, J.; Biggin, P. C., The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. *PLoS Comput Biol* **2012**, 8, (11), e1002749.
50. Das, A.; Grinkova, Y. V.; Sligar, S. G., Redox potential control by drug binding to cytochrome P450 3A4. *J Am Chem Soc* **2007**, 129, (45), 13778-9.
51. Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H., Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, 305, (5684), 683-6.
52. Yamazaki, H.; Shimada, T., Progesterone and testosterone hydroxylation by cytochromes P450 2C19, 2C9, and 3A4 in human liver microsomes. *Arch Biochem Biophys* **1997**, 346, (1), 161-9.
53. Domanski, T. L.; He, Y. A.; Harlow, G. R.; Halpert, J. R., Dual role of human cytochrome P450 3A4 residue Phe-304 in substrate specificity and cooperativity. *J Pharmacol Exp Ther* **2000**, 293, (2), 585-91.

54. Jiang, Y.; Ortiz de Montellano, P. R., Cooperative effects on radical recombination in CYP3A4-catalyzed oxidation of the radical clock beta-thujone. *Chembiochem* **2009**, 10, (4), 650-3.
55. Roberts, A. G.; Campbell, A. P.; Atkins, W. M., The thermodynamic landscape of testosterone binding to cytochrome P450 3A4: ligand binding and spin state equilibria. *Biochemistry* **2005**, 44, (4), 1353-66.
56. Joseph, H.; Stancliff, S.; Langrod, J., Methadone maintenance treatment (MMT): a review of historical and clinical issues. *Mt Sinai J Med* **2000**, 67, (5-6), 347-64.
57. Kamal, N. N. S. B. N. M.; Lim, T. S.; Tye, G. J.; Ismail, R.; Choong, Y. S., The Effect of CYP2B6, CYP2D6, and CYP3A4 Alleles on Methadone Binding: A Molecular Docking Study. *Journal of Chemistry* **2013**, 2013, 7.
58. Iribarne, C.; Dreano, Y.; Bardou, L. G.; Menez, J. F.; Berthou, F., Interaction of methadone with substrates of human hepatic cytochrome P450 3A4. *Toxicology* **1997**, 117, (1), 13-23.
59. Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Polarizable Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *J Chem Theory Comput* **2013**.
60. Lonsdale, R.; Houghton, K. T.; Zurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J., Quantum mechanics/molecular mechanics modeling of regioselectivity of drug metabolism in cytochrome P450 2C9. *J Am Chem Soc* **2013**, 135, (21), 8001-15.
61. Kirchmair, J.; Williamson, M. J.; Tyzack, J. D.; Tan, L.; Bond, P. J.; Bender, A.; Glen, R. C., Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. *J Chem Inf Model* **2012**, 52, (3), 617-48.
62. Tyzack, J. D.; Williamson, M. J.; Torella, R.; Glen, R. C., Prediction of cytochrome P450 xenobiotic metabolism: tethered docking and reactivity derived from ligand molecular orbital analysis. *J Chem Inf Model* **2013**, 53, (6), 1294-305.
63. Tu, Y.; Deshmukh, R.; Sivaneri, M.; Szklarz, G. D., Application of molecular modeling for prediction of substrate specificity in cytochrome P450 1A2 mutants. *Drug Metab Dispos* **2008**, 36, (11), 2371-80.
64. Daber, R.; Stayrook, S.; Rosenberg, A.; Lewis, M., Structural analysis of lac repressor bound to allosteric effectors. *J Mol Biol* **2007**, 370, (4), 609-19.
65. Bond, L. M.; Peters, J. P.; Becker, N. A.; Kahn, J. D.; Maher, L. J., 3rd, Gene repression by minimal lac loops in vivo. *Nucleic Acids Res* **2010**, 38, (22), 8072-82.
66. Czapla, L.; Grosner, M. A.; Swigon, D.; Olson, W. K., Interplay of Protein and DNA Structure Revealed in Simulations of the lac Operon. *PLoS One* **2013**, 8, (2), e56548.
67. Ekroos, M.; Sjogren, T., Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc Natl Acad Sci U S A* **2006**, 103, (37), 13682-7.
68. Sevrioukova, I. F.; Poulos, T. L., Pyridine-substituted desoxyritonavir is a more potent inhibitor of cytochrome P450 3A4 than ritonavir. *J Med Chem* **2013**, 56, (9), 3733-41.

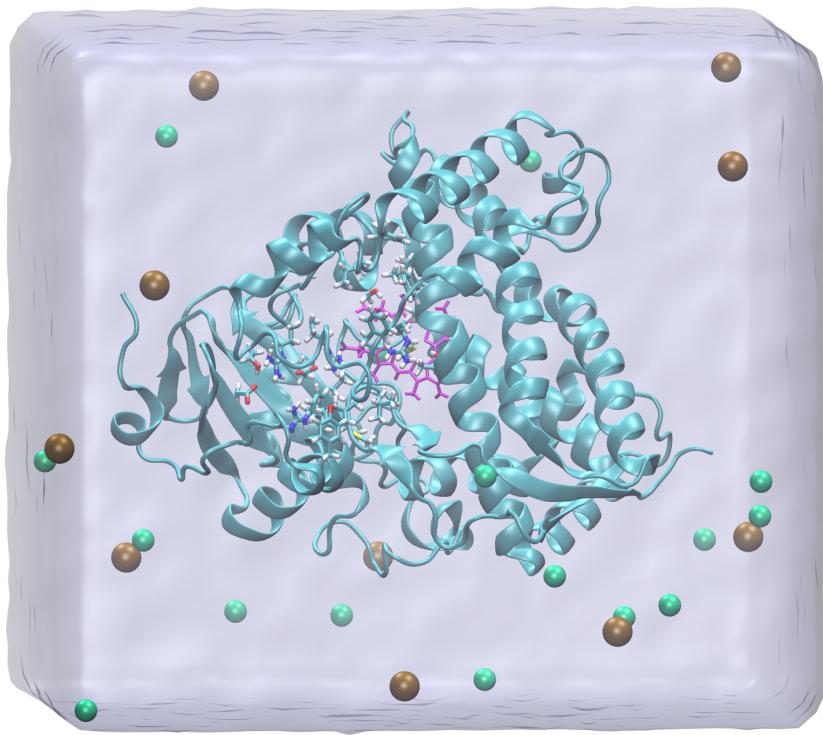
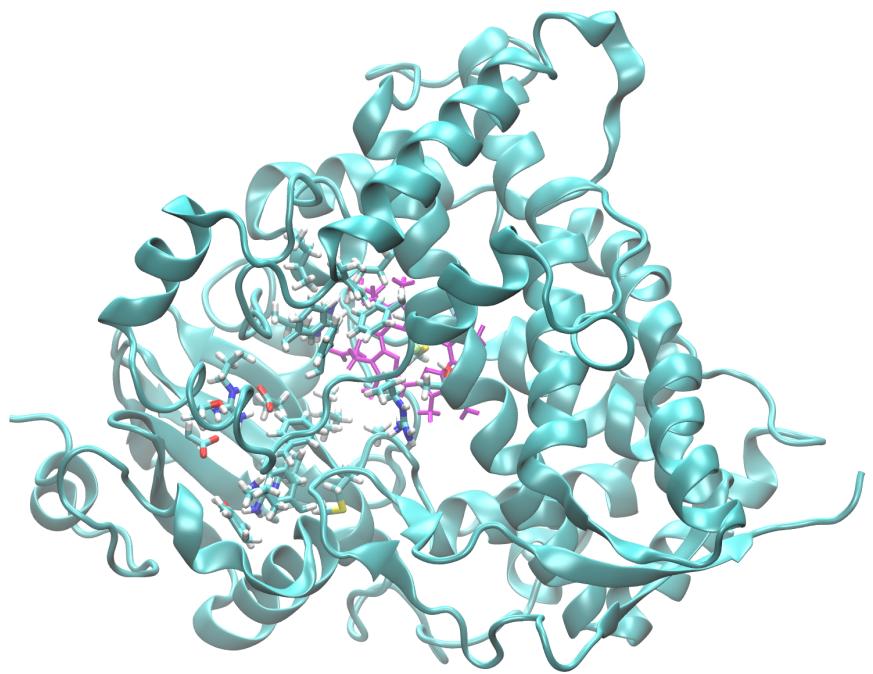


Figure 1A



*Figure 1B*

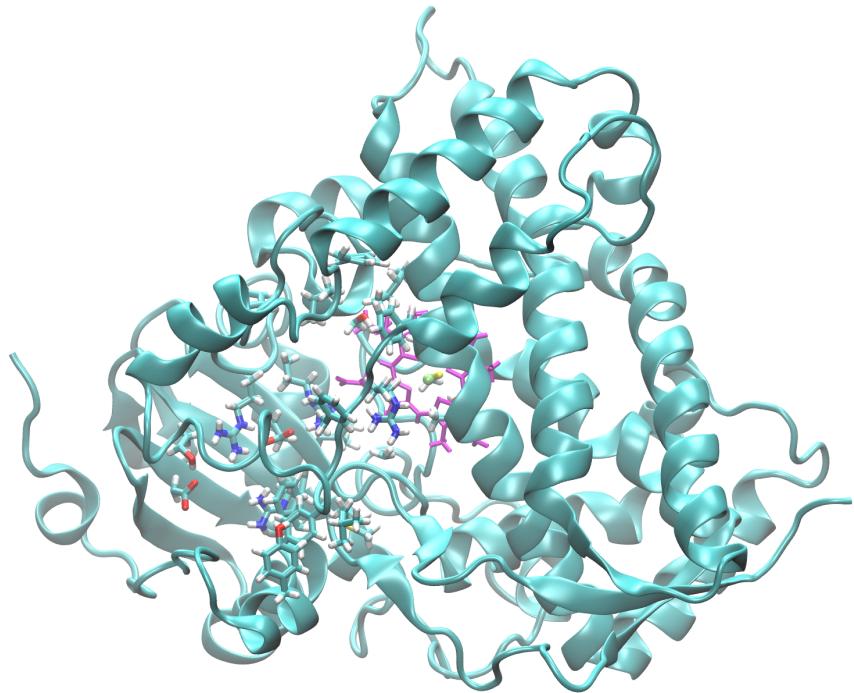


Figure 1C

**Figure 1 – Depiction of the simulation after 300 ns of Accelerated Molecular Dynamics compared to the initial structure.** The entire system of 42094 atoms with 7757 protein atoms (CYP3A4 and the heme moiety), 11,437 water molecules (one coordinated to the heme Fe(III) center with constraints), and 26 ions (10 K<sup>+</sup> and 16 Cl<sup>-</sup>) is depicted in Figure 1A with the final conformation after 300 ns illustrated from a view into the active site from the distal side of the protein, observing the G and G' helices have aligned nearly end-to-end and are much closer to parallel, and the looped region after the G helix has rearranged into an alpha helix, with water molecules reduced to a density surface using the VolMap tool in VMD. The simulated system at the beginning of the simulation after 6 ns of NPT ensemble equilibration starting from the apo human CYP3A4 structure with PDB ID 1TQN is rendered in Figure 1B. The active site amino acid residues Tyr53, Phe57, Asp76, Arg105, Arg106, Phe108, Ser119, Ile120, Arg212, Phe213, Phe215, Leu216, Thr224, Phe241, Ile301, Phe304, Ala305, Thr309, Ala370, Met371, Arg372, Leu373, and Glu374 revealed in the bromocriptine-CYP3A4 structure with Protein Databank accession code 3UA1 (as detailed in the methodology), along with the Cys442 side chain coordinated with the heme iron, are depicted with the licorice representation in VMD. The active site heme porphyrin atoms are depicted in

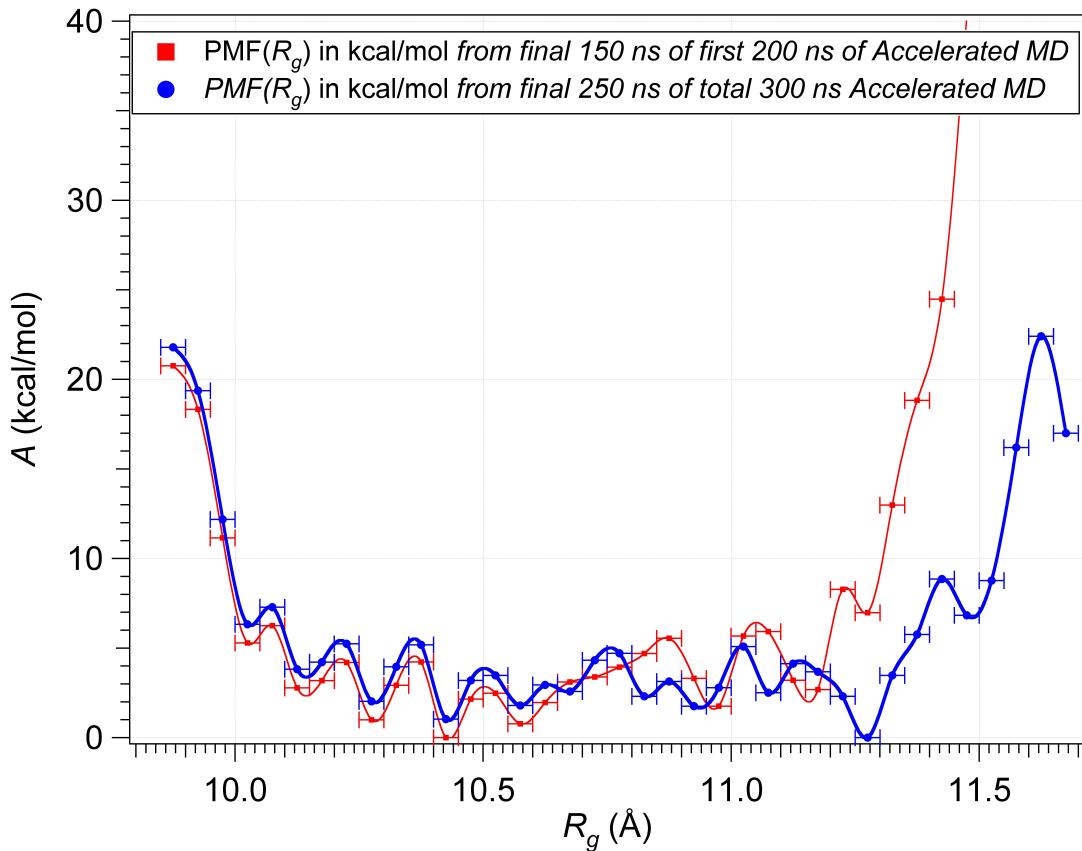
magenta, with the modeled central Iron in the Fe(III) oxidation state depicted as a van der Waals radius sphere in lime. In Figure 1C, the system after 300 ns of Accelerated Molecular Dynamics (as in Figure 1A) is rendered in the same format and viewing angle as Figure 1B. A comparison of the structures in Figures 1B with Figure 1C shows significant structural rearrangements and greater accessibility of the active site under the solution conditions of the simulation, as quantified by comparison of Table 1 for the crystal structures with the relative probability of states of the simulation in Figure 2.

PDB ID	Resolution (Å)	Active site $R_g$ (Å)	Ligand
3UA1	2.15	9.81	Bromoergocryptine
3NXU	2.00	10.71	Ritonavir
2V0M <sup>67</sup>	2.80	11.15	Ketoconazole × 2*
2J0D <sup>67</sup>	2.75	10.85	Erythromycin
1W0F <sup>51</sup>	2.65	9.89	Progesterone**
1W0G <sup>51</sup>	2.73	9.98	Metyrapone
1TQN/1W0E <sup>51</sup> /4I3Q <sup>68</sup>	2.05/2.80/2.60	9.82/9.86/9.79	<b>NO LIGAND</b>

**Table 1 – Crystallographic structures deposited in the Protein Data Bank and their corresponding active site accessibility as a radius of gyration.** The PDB accession code, X-ray diffraction resolution, active site accessibility as the value of the active site residue radius of gyration  $R_g$  defined in the methodology, and the name of the bound ligand observed in the active site for each crystal structure is listed. The active site radius of gyration estimates the accessibility of the CYP3A4 active site for ligand binding to each structure.

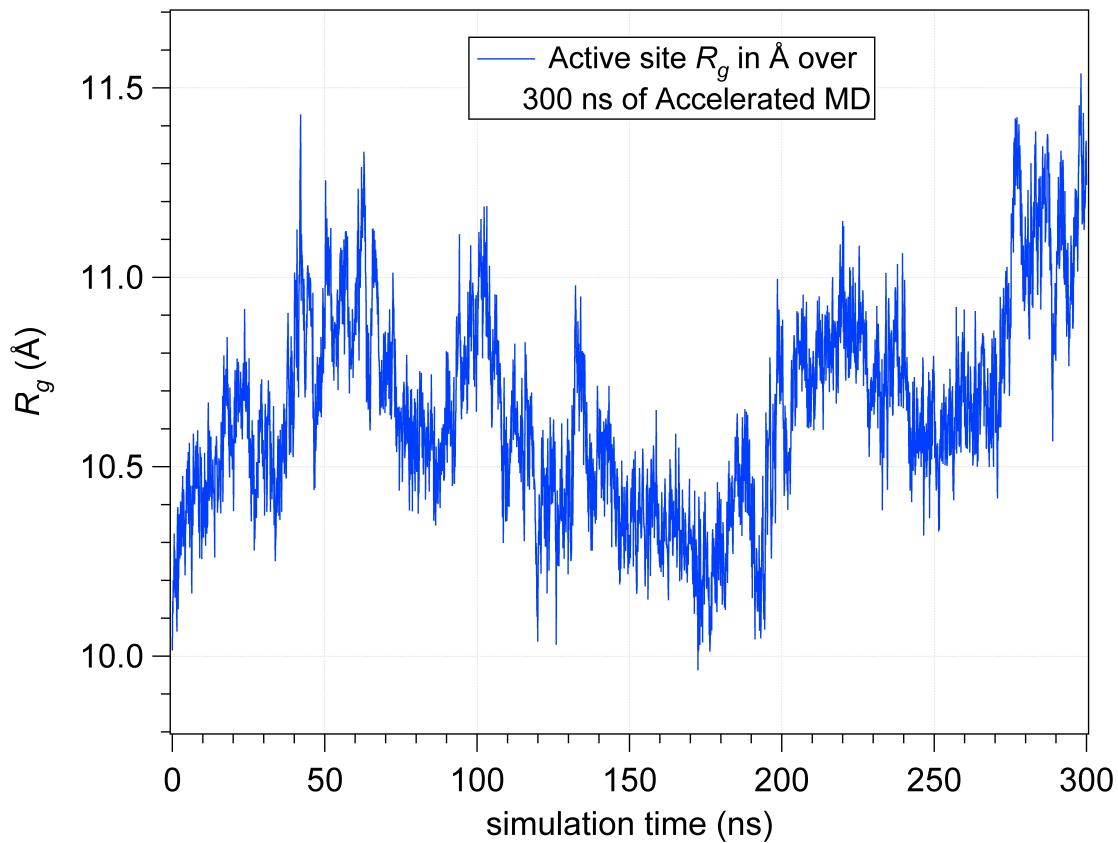
\* There are two ketoconazole molecules observed in the active site of crystal structure PDB ID 2V0M.

\*\* Progesterone ligand may not be considered positioned within the active site in PDB ID 1W0F and the authors report this as a putative peripheral binding site<sup>51</sup>.



**Figure 2 – The Potential of Mean Force as a function of the radius of gyration of active site residues estimates the increased active site accessibility observed in the simulation.** In Figure 2A, the value  $A = -k_B T \ln p(R_g)$  is plotted for the probability histogram  $p(R_g)$  of the radius of gyration, using the final 250 ns of the 300 ns Accelerated Molecular Dynamics simulation (125,000,000 total protein configurations with the time step  $\Delta t = 2$  fs) collected for bins of width  $\Delta R_g = 0.05\text{\AA}$ , where each protein configurational state has a weight of  $\exp(-\beta\Delta V)$  (blue data). The same analysis of the final 150 ns of the first 200 ns of the simulation is presented (using these 75,000,000 protein configurations, red data) to demonstrate the results from analysis prior to significant structural changes observed in the final 100 ns, as quantified in Figure 8. The points of the histogram are plotted with horizontal error bars representing the bin width of 0.05  $\text{\AA}$ , the minimum value of  $A$  is subtracted from each data point to

define a reference point with  $A = 0$  kcal/mol, and the curve (dashed line) is fit to cubic spline functions using the plotted histogram points (markers with error bars). Estimating the error is a challenging problem, but from inspection of the figure it appears to be beyond quantification because protein conformational sampling is a massive challenge for such a short simulation and the system continues to evolve. The data shows that the apo CYP3A4 active site has a broad distribution of conformations and is more open under the simulated solution conditions at  $T = 310$  K than in the crystallographic forms, and the analysis of the final 250 ns of the total 300 ns simulation time suggests the behavior of the simulation at longer Accelerated MD time scales.



**Figure 3 – Time evolution of the active site accessibility as measured by the radius of gyration of active site residues.** The value of the active site radius of gyration  $R_g$  (as defined in the Methodology) for the simulated protein conformations at 100 ps intervals is plotted over the course of the entire 300 ns Accelerated Molecular Dynamics simulation (the initial 50 ns followed by an additional 250 ns of simulation). The data shows that the system evolves from the initially closed crystallographic conformation (with radius of gyration  $R_g = 9.82$  Å prior to equilibration) of the active site elucidated in the crystallographic coordinates of the wild-type human CYP3A4 in the structure with Protein Data Bank accession code 1TQN, and the active site remains in significantly more open conformations for the remainder of the simulation, comparable to and sometimes greater than the active site accessibility seen in the crystal structures of human CYP3A4 with bound ligands ( $R_g = 9.81$  Å to 11.15 Å). The data shown here is not Boltzmann weighted and Figure 2 illustrates the resulting histograms considering the boost potential  $\Delta V$  in the Accelerated Molecular Dynamics simulation.

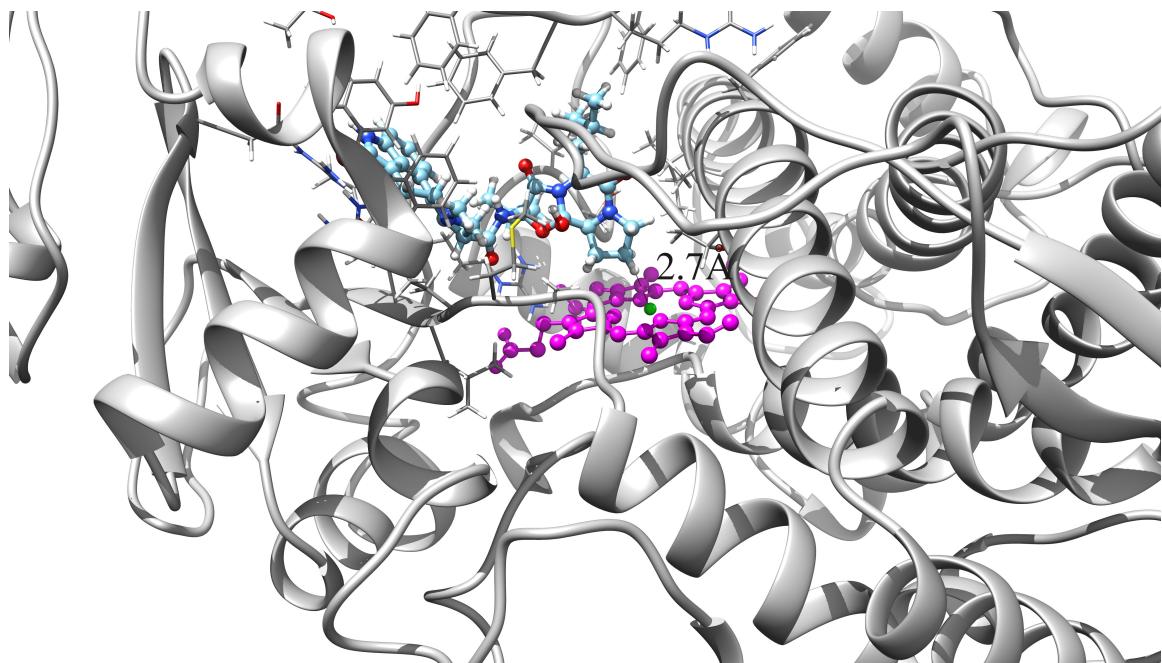


Figure 4A

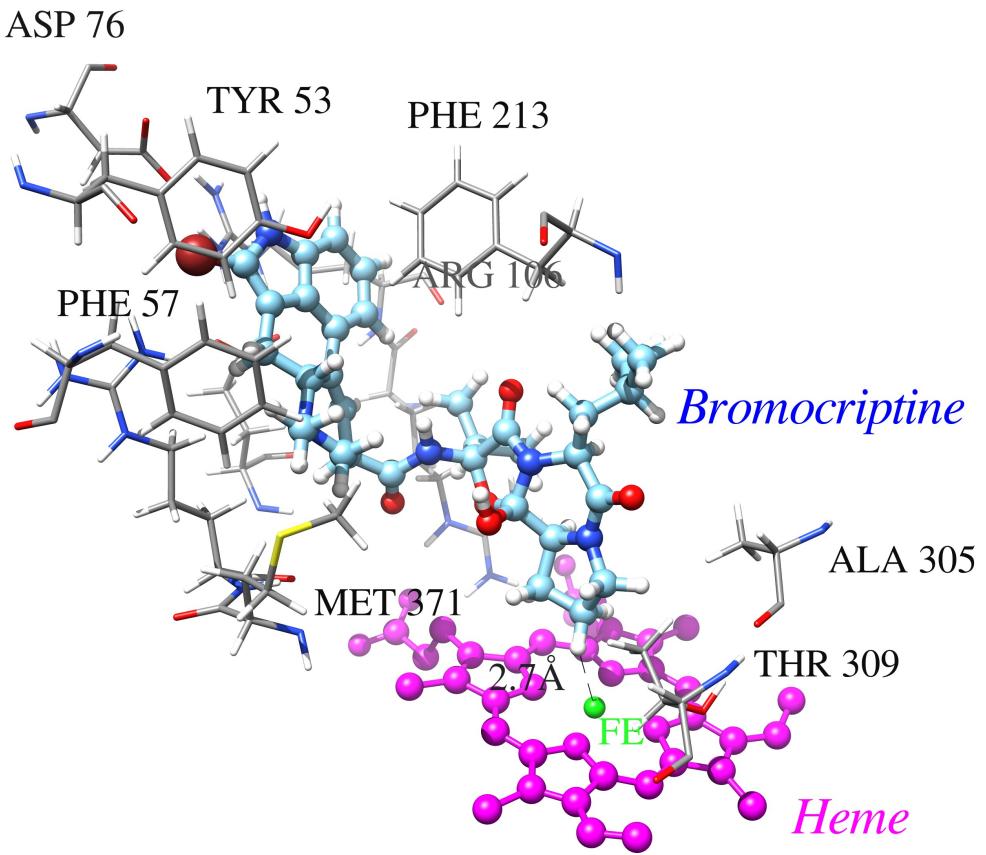


Figure 4B

**Figure 4 – A productive binding mode of bromocriptine is suggested among the top poses of the first eight clusters from blind docking with EADock DSS.** Upon examining the highest ranked docked poses from the first eight clusters using the EADock DSS evolutionary algorithm, we identify as the second top docked pose with the ligand close enough to the catalytic heme center for hypothetical productive oxidation as the top docked pose of cluster #4, and it appears to represent a productive binding mode for the 8' and 9' hydroxylation of the proline ring of the bromocriptine cyclic tripeptide moiety. Figure 4A shows a side view into the active site with the ligand positioned in this pose and identifies a distance of 2.7 Å between one of the hydrogens of the aliphatic 9' proline carbon and the heme iron, and the distance between the 9' carbon and iron is 3.4 Å. Figure 4B shows a more detailed view of the active site residues observed to form close contact interactions with bromocriptine in the bromocriptine-CYP3A4 crystal structure with PDB accession code 3UA1 that are also observed in this docked binding mode. The active

site residues near the ligand in the docked pose are illustrated using a stick representation, the bromocriptine ligand is rendered with a ball and stick representation with carbon atoms in cyan, and the heme is rendered with a ball and stick representation with the reduced iron in green and the porphyrin ring in magenta. The illustrations were rendered with UCSF Chimera.

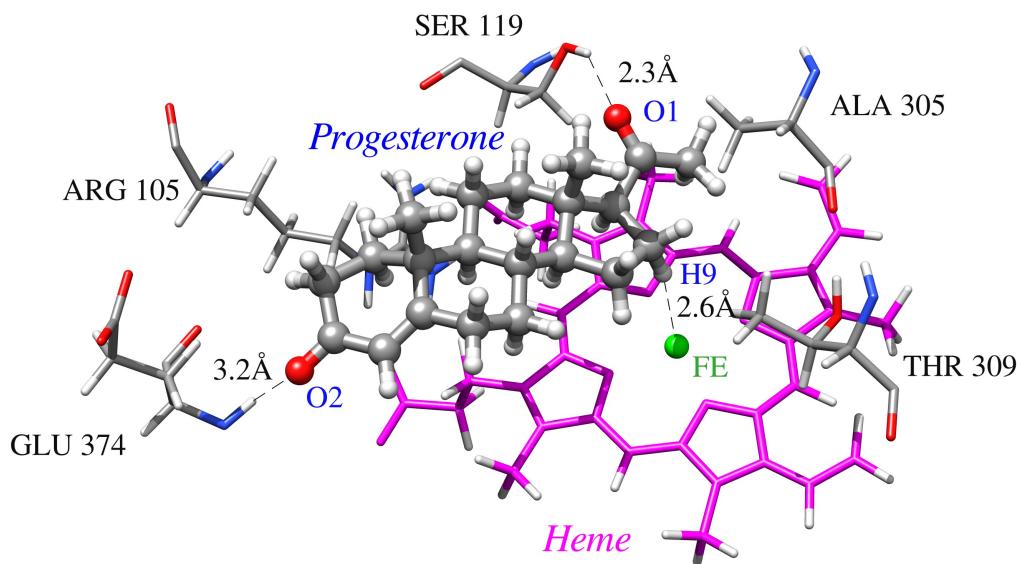


Figure 5A

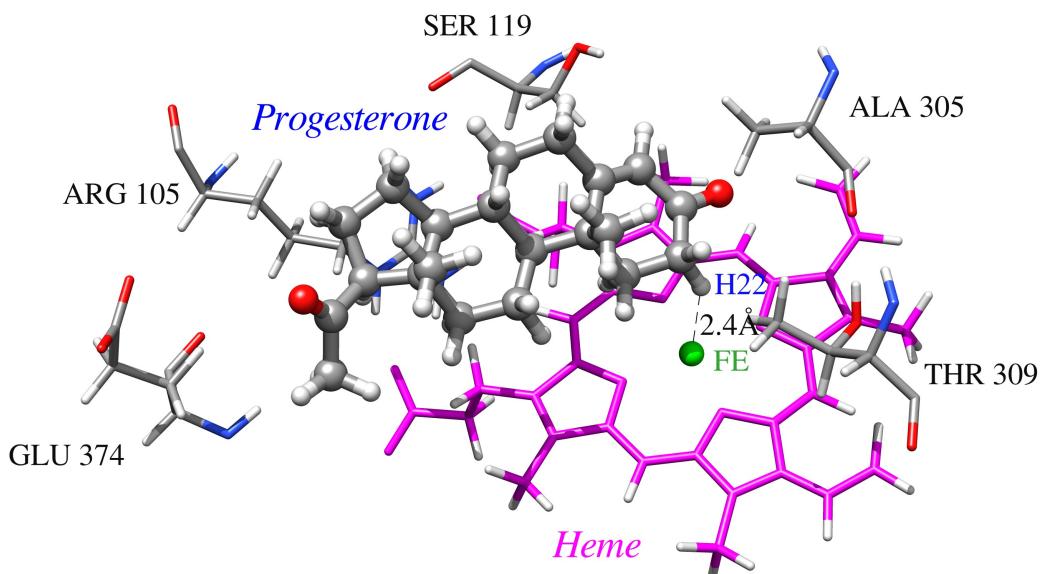
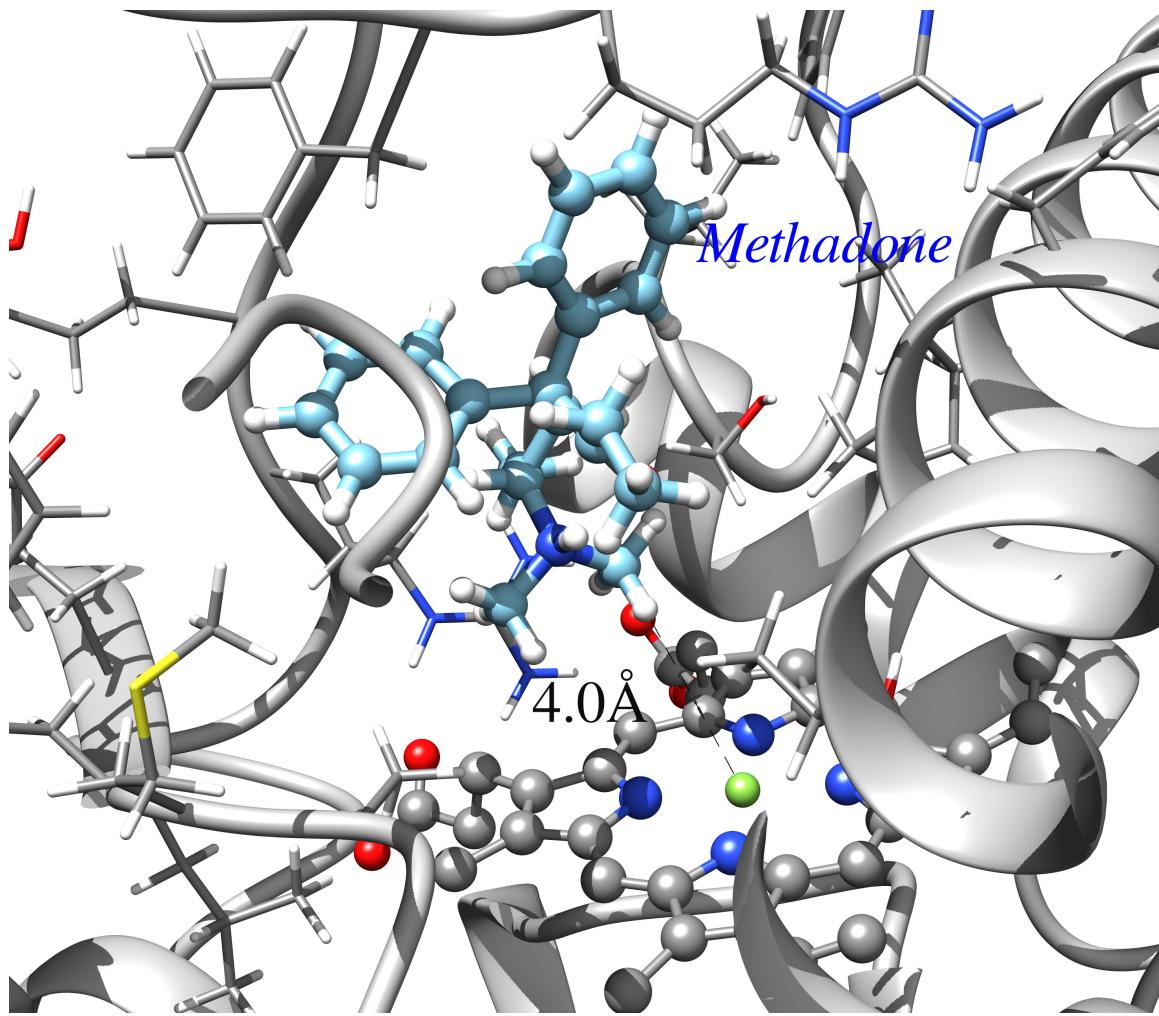


Figure 5B

**Figure 5 – Prediction of productive poses for progesterone hydroxylation in the top poses of the first eight clusters from blind docking.** Figure 4A shows the top pose of Cluster #1, consistent with the  $16\alpha$ -hydroxylation site of metabolism of progesterone by CYP3A4<sup>52</sup>. There is strong hydrogen bonding between the hydroxyl side chain of Ser-119 and the O20 oxygen (labeled O1) of progesterone, and also weak hydrogen bonding between the amide nitrogen of the backbone of Glu-374 and the O3 oxygen (labeled O2) of progesterone. The distance between the hydrogen of the aliphatic  $16\alpha$  carbon (labeled H9) and the heme iron is 2.6 Å in the docked binding mode, with a  $16\alpha$  carbon-iron distance of 3.4 Å. Figure 4B shows the top pose of cluster #8, consistent with the  $2\beta$ -hydroxylation site of metabolism of progesterone<sup>52</sup>. The hydrogen (labeled H22) from the  $2\beta$  carbon of progesterone is positioned 2.4 Å from the heme iron, with a  $2\beta$  carbon-iron distance of 3.1 Å. In addition, the top pose from cluster #3 (not pictured) has a hydrogen of the C21 aliphatic carbon at a distance of 2.6 Å from the heme iron, positioning the site of metabolism for the 21-hydroxylation of progesterone, observed as a significant product of progesterone oxidation by another cytochrome P450 enzyme, CYP2C9. The images were rendered with UCSF Chimera with a similar representation as in Figure 4B.



**Figure 6 – A putative peripheral binding site for erythromycin as revealed by the top pose of cluster #2 from blind docking.** The figure shows a top down view looking into the active site of cytochrome P450 3A4 from the distal side near the putative membrane-associated region near the F and G helices. The erythromycin position at the sharp turn junction of the F'-G' helices at the top left of the image, from the top docked pose of cluster #2 from blind docking with EADock DSS using the single simulated structure is rendered with atomic coloring with a ball-and-stick model (gray – carbon, red – oxygen, and blue – nitrogen), similar to the positioning of progesterone as revealed in Protein Data Bank crystal structure 1W0F. The active site residue side chains (as defined in the methodology and legend to Figure 1) are rendered with a wireframe representation and the heme porphyrin is represented with a ball-and-stick model with the Iron at the catalytic center of the active site in green. The top docked pose of cluster #1 has erythromycin on the proximal side of the protein away from the active site and the image of this cluster #2 binding mode that may represent either a putative peripheral binding site or simply a relevant state in the process of substrate entry is rendered with UCSF Chimera.

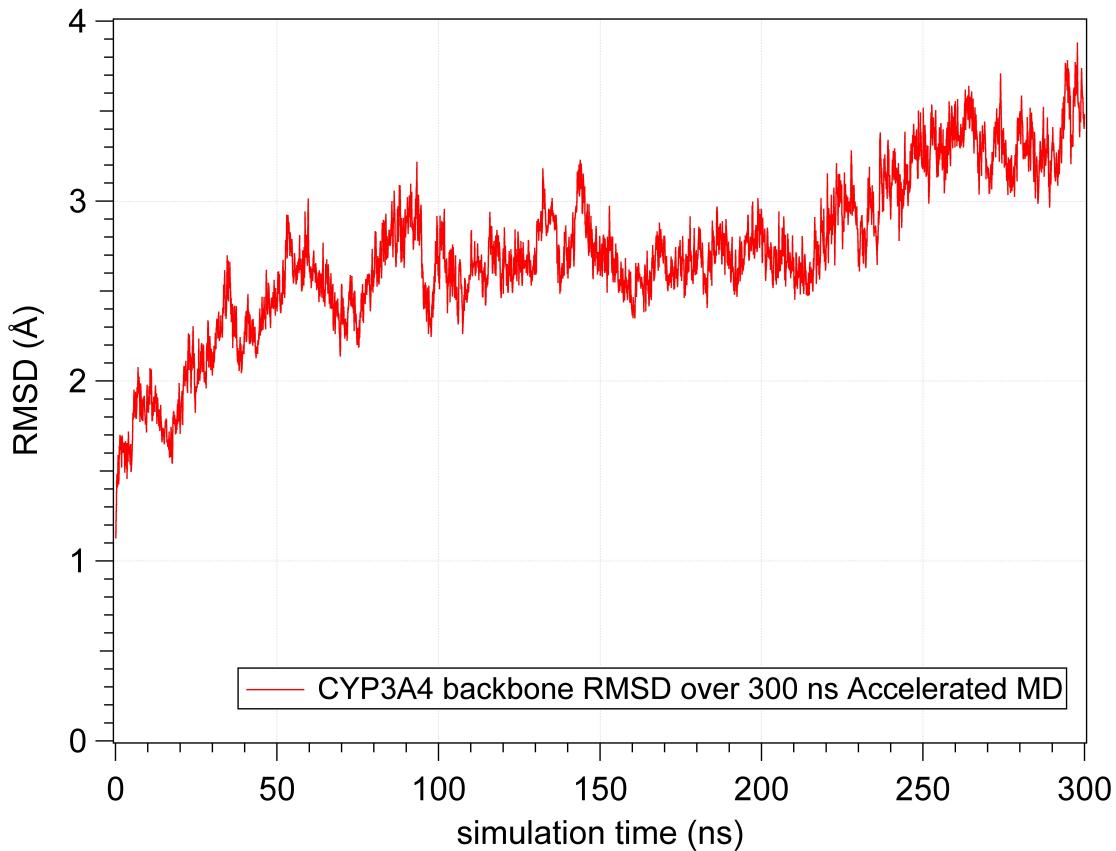


**Figure 7 – A putative N-demethylation binding mode of the medically-relevant compound R-(-)-methadone as revealed by the top docked pose of cluster #3 from blind docking.** Using the same simulated structure after 50 ns as the results for bromocriptine, progesterone, and erythromycin, a putative N-demethylation binding mode of R-(-)-methadone is revealed by blind docking with EADock DSS, as the top pose of cluster #3 with a distance of 4.0 Å between one of the hydrogen atoms of a 6-amino methyl carbon and the heme iron (the carbon-iron distance is 5.0 Å). The result suggests that productive binding modes of many substrates of interest may be located within the top eight clustered poses from EADock DSS, and additional results for other medicinal compounds supports this finding (Table 2). The image is rendered with UCSF Chimera, with active site residues and ligand depicted with a similar representation as the previous images.

<b>Drug</b>	<b>Pose cluster</b>	<b>Carbon-Fe distance (Å)</b>	<b>Carbon Position</b>	<b>ΔG<sub>binding</sub> (kcal/mol)</b>	<b>Product</b>
Bromocriptine	#2	4.9/4.2	8'/9'	-9.89	8',9'-hydroxy-
Progesterone*	#1,2,6	1) 4.7, 4.0,3.2	1) 16α	-7.93	1) 16α-hydroxy-
	#10	2) 10	2) 2β		2) 2β-hydroxy-
Amitriptyline	#6	4.3	N15-methyl	-7.20	nortriptyline
Buprenorphine	#1	3.9	17-methyl	-8.92	norprenorphine
Carbamazepine	#1	3.3/3.8	10/11	-6.49	CBZ-10,11 epoxide
Erythromycin	-	-	-	-	-
Escitalopram	#5	3.7	N-methyl	-7.70	desmethyl-
Eszopiclone	#2	1) 3.2 (N-Fe)	<b>Pyrazine N</b>	-8.48	1) -N-oxide
	#6	2) 3.2	N4-methyl		1) N-desmethyl-
(R-)Methadone	#4	4.8	N-methyl	-	N-desmethyl-
Midazolam	#4	1) 4.5	1) 4	-7.55	1) 4-hydroxy-
	#11	2) 3.6	2) 1'		2) 1'-hydroxy-
Oxycodone	#1	4.5	N9-methyl	-7.32	noroxycodone
Sildenafil	#1	4.8	N4-methyl	-9.19	N-desmethyl-
Testosterone	#2	1) 3.9	2β	-8.20	2β-hydroxy-
	#7	2) 4.0	6β		6β-hydroxy-
(R-)Venlafaxine	#2	1) 3.0	O-methyl	-7.19	O-desmethyl-
	#7	2) 4.8	N-methyl		N-desmethyl-

**Table 2 – Summary of results from a blind docking survey of drugs using the same simulated structure in the Iron(III) heme oxidation state.** The results from docking to the Fe(III) heme oxidation state from the simulation reveals that productive binding modes for many drugs are among the top poses of the first eight clusters from blind docking, suggesting that this single simulated structure used for the docking results gives insight into every docked compound besides erythromycin in Figure 6 (see Supporting Material S6 for a successful identification of a putative N-demethylation mode using a different simulated structure from the end of the 300 ns simulation). The name of the drug, the cluster number in which the top pose represents a hypothetical productive mode, the corresponding carbon-iron distance (or nitrogen-iron distance in the case of the N-oxidation of eszopiclone) for the relevant carbon (or nitrogen) group of this mode, the best  $\Delta G_{\text{binding}}$  among all the top eight clustered poses, and the product of the oxidation reaction involving the site of metabolism considered is listed in the Table. Additional binding modes consistent with other sites of the metabolism for the substrate, from top poses below the cutoff of the first eight clusters, are listed when one of these top poses of the first eight clusters represents a productive binding mode. The results demonstrate the significant role of ligand positioning in the process of CYP-mediated metabolism, but there are no apparent strong thermodynamic factors driving drug binding specifically towards productive poses because the average  $\Delta \Delta G_{\text{binding}}$  between the highest and lowest free energy top poses of the first eight clusters is typically only 0.5-1.0 kcal/mol in the structure used here for docking. A carbon-iron distance cutoff of less than or equal to 5.0 Å for a biologically relevant site of metabolism is used to define productive binding modes.

\* Poses 1,2, and 6 are all consistent with  $16\alpha$ -hydroxylation of progesterone, with significantly different orientations of the ligand within the active site.



**Figure 8** – The time evolution of the relative RMS deviation of non-hydrogen backbone atoms of the protein, with respect to resolved amino acid residues in apo CYP3A4 crystal structure with Protein Data Bank accession code 1TQN, during the course of the entire 300 ns Accelerated Molecular Dynamics simulation. The simulated protein conformations are analyzed at 100 ps intervals and the results for these 3000 conformational states are plotted in the Figure. The final 100 ns of Accelerated Molecular Dynamics reveals that the system continues to evolve in the simulation, and longer simulations may reveal additional insights, but must be carefully interpreted due to the lack of experimental data about hypothetical states as the system evolves to be much different than the crystallographic structures. The large motion responsible for this increased relative RMSD towards the end of the simulation appears to be a collective rotation of helix G followed by reorienting motions of the F and G helices to become more antiparallel with corresponding motions of the looped residues between the F helix and the F'-G' helical junction, followed by a final aligning the of G' and G helices in more of an end-to-end orientation (compare Figure 1B with Figure 1C for an illustration).