Elucidating a chemical defense mechanism of Antarctic sponges: A computational study

Sai Lakshmana Vankayala^{a,1}, Fiona L. Kearns^{a,1}, Bill J. Baker^a, Joseph D. Larkin^b, H. Lee Woodcock^{a,*}

^aDepartment of Chemistry, University of South Florida, 4202 E. Fowler Ave., CHE205, Tampa, FL 33620, USA

^bDepartment of Chemistry, Eckerd College, 4200 54th Avenue South, St. Petersburg, FL 33711, USA

Abstract

In 2000, a novel secondary metabolite (erebusinone, Ereb) was isolated from the Antarctic sea sponge, *Isodictya erinacea*. The bioactivity of Ereb was investigated, and it was found to inhibit molting when fed to the arthropod species *Orchomene plebs*. Xanthurenic acid (XA) is a known endogenous molt regulator present in arthropods. Experimental studies have confirmed that XA inhibits molting by binding to either (or both) of two P450 enzymes (CYP315a1 or CYP314a1) that are responsible for the final two hydroxylations in the production of the molt-inducing hormone, 20-hydroxyecdysone (20E). The lack of crystal structures and biochemical assays for CYP315a1 or CYP314a1, has prevented further experimental exploration of XA and Ereb's molt inhibition mechanisms. Herein, a wide array of computational techniques – homology modeling, molecular dynamics simulations, binding site bioinformatics, flexible receptor - flexible ligand docking, and molecular mechanics-generalized Born surface area calculations – have been employed to elucidate the structure - function relationships between the aforementioned P450s and the two described small molecule inhibitors (Ereb and XA). Results indicate that Ereb likely targets CYP315a1 by interacting with a network of aromatic residues in the binding site, while XA may inhibit both CYP315a1 and CYP314a1 because of its aromatic, as well as charged nature.

Keywords: P450 enzyme, erebusinone, xanthurenic acid, homology modeling, molecular dynamics simulations, aromatic interactions

1. Introduction

Natural products research aims to isolate and characterize primary or secondary metabolites released by biological systems. Novel compounds with relevant bioactive properties are then used in pharmaceuticals, nutraceuticals, cosmetics, chemical engineering and more.^{1,2} One core component of successful natural products research is to identify sources of chemical diversity that are rooted in species diversity and thus also driven by the evolutionary factors of competition, predation, and defense.^{1,2} Sessile species hold

*Corresponding author

Email address: hlw@usf.edu (H. Lee Woodcock)

¹These authors contributed equally.

unique promise for producing novel metabolites as they often rely on chemical defense mechanisms due to their lack of mobility.² In 2000, Baker and coworkers reported one such unique aromatic metabolite from the Antarctic sponge, *Isodictya erinacea*,¹ which was named erebusinone (Ereb, Figure 1), after the Erebus Bay where the *I. erinacea* specimen was collected.

Interest in Ereb’s ecological role piqued when feeding studies performed with *Orchomene plebs* revealed Ereb inhibits molting in this small arthropod, resulting in high mortality. These observations supported the hypothesis that Ereb chemically disrupts molting in the crustacean predators of *I. erinacea*.¹ Other research has determined the effect of xanthurenic acid (XA), a known endogenous molt inhibitor, on crustacean molting^{3,4,5} and, due to structural and functional similarities to Ereb, it was proposed the two may inhibit this pathway by the same mechanism.^{1,4,5,6} XA is known to interact with, and inhibit, one or more P450s responsible for secretion of 20-hydroxyecdysone (20E); however, the exact details of how XA binds to either CYP315a1 or CYP314a1, and which binding site is favored by XA, remain unclear.^{3,4,5}

The circulatory molt inducing hormone (20E) is produced ultimately from extracellular cholesterol through a series of successive biotransformations in various physiological locations.^{7,8} The final stages of this pathway are a series of hydroxylations performed by a small set of closely related cytochrome P450 enzymes.^{27,8,9} Ecdysis, or ecdysteroidogenesis, is the biochemical pathway by which ecdysteroids (molting hormones) are produced and regulated for molting and growth in arthropods. This pathway has been studied thoroughly in insects, but remains somewhat clouded in the context of their crustacean cousins.¹⁰ There are some key results, however, that may allow us to apply knowledge about insect ecdysis to crustaceans. In insects, the intermediate ecdysone (E) is released from the prothoracic gland (PG) to surrounding tissues where it undergoes the final hydroxylation step to produce 20E.^{9,11} In crustaceans, E is secreted from the Y-organ complex (YOC) to surrounding tissues followed by hydroxylation to produce 20E, thusly the insect PG and crustacean YOC are often considered physiologically analogous, indicating a conservation of the pathway amongst arthropods.^{9,6,11,12,13,14,15,16,17} It has been amply supported that the original series of hydroxylation reactions that occur in crustaceans to produce 20E is also highly conserved in insects.^{16,17}

As mentioned, enzymes responsible for the final hydroxylations to produce 20E, are members of a small group of cytochrome P450 enzymes (also known as CYPs or P450s). Studies performed on insect (*Drosophila* and *Bombyx*) species have provided most of our understanding about ecdysteroidogenesis and its molecular pathway.^{9,16,18,19,20} Recent studies have sequenced various crustacean genomes and have identified orthologs to insect ecdysis P450s present in crustacean DNA, therefore the crustacean molting pathway can be appropriately modeled with insect sequences. The overall 20E biosynthesis begins with cholesterol and, after several steps, yields 2-deoxyecdysone (2-dE), the E precursor.^{16,21,20,10} 2-dE is hydroxylated by the C2-hydroxylase, CYP315a1, releasing E which binds to CYP314a1 for hydroxylation at C₂₀ finally producing 20E.^{22,23} Past studies report that the target for regulation by XA is either CYP315a1 or CYP314a1.^{2,5} Figure 2

²Throughout this text we may refer to such closely related P450s as the “molting protein(s)” to emphasize that these are responsible for the production of the molt-inducing hormone, 20E.

highlights these final two hydroxylation steps, which are herein investigated.

The purpose of this current work is to predict how Erebus and XA inhibit 20E biosynthesis and thus crustacean molting.^{2,3,4} From here on we refer collectively to Erebus and XA as “small molecule inhibitors,” or SMIs, Figure 1 compares these SMIs. A variety of computational methods including homology modeling, molecular dynamics (MD) simulations, binding mode analysis, flexible docking studies, and molecular mechanics - generalized Born surface area (MM-GBSA) calculations, have been used to investigate SMI binding in CYP315a1 and CYP314a1. Additionally, we compared the observed interactions between the SMIs and P450 binding sites to the interactions between the native substrates (2-dE, E, and 20E) and P450 binding sites to predict how SMI binding may interrupt enzymatic activity.

2. Computational Details

P450 Homology Modeling: Query amino acid sequences of *Drosophila melanogaster* P450s, CYP315a1 and CYP314a1, were obtained from the NIH GenBank (accession numbers AAS65010.1 and AAL86019.1 respectively).^{24,22} The BLAST (basic local alignment search tool) algorithm²⁵ was used to perform a local structural sequence alignment search through the non-redundant PDB²⁶ for a template homolog, and CYP24a1 was identified (PDB code: 3K9V chain A, a rat P450 that hydroxylates vitamin D), with the highest sequence similarity for both CYP315a1 (BLAST-bit score: 173.3) and CYP314a1 (BLAST-bit score: 166.0). Homology models of the *Drosophila* molting proteins CYP315a1 and CYP314a1 were constructed using the Prime 3.1 comparative modeling module.²⁷ The iron protoporphyrin IX (heme) and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CPS, the CYP24a1 substrate, Figure SI1) from the 3K9V-A crystal structure were retained in the query binding site to aid in homology model construction. For serial loop sampling, Prime uses a hierarchical refinement protocol with a loop buildup algorithm (see Supporting Information for more details).²⁸ Further, side-chain optimizations are performed on filtered loop candidates, followed by full minimizations of the loops until convergence of the gradient is reached ($< 0.001 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-1}$). Prime uses the OPLS-2005 all atom force field²⁹ in conjunction with generalized Born implicit solvation (GBSA) to score the energies of loop candidates. After loop refinement, the homology models were further improved using Truncated Newton

Figure 1: Structures of the small molecule inhibitors (SMIs) investigated in this work: erebusinone (Ereb) and xanthurenic acid (XA). XA is shown here in both its keto and enol forms.

Conjugate Gradient (TNCG) minimization³⁰ until the RMSD of heavy atoms reached a maximum cut-off of 0.3 Å from their respective initial coordinate geometries. Ramachandran plots were generated using PROCHECK.³¹ Finally, an online homology modeling Web-server, I-TASSER, was used to generate models for structural comparison with our models.^{32,33,34} Homology models of CYP315a1 and CYP314a1 were then simulated as described below.

Molecular Dynamics Simulations: Molecular dynamics (MD) simulations were performed using the CHARMM biomolecular simulation program.^{35,36,37} CHARMMing³⁸ and pdbtools (<https://github.com/harmlab/pdbtools>) were used to parse the homology models for correct atom types and residue numbers and to prepare the structures for simulation. The proteins were treated with the CHARMM (C22) force field, while substrate topology and parameters were evaluated using the CHARMM generalized force field (CGenFF) via the Paramchem Web-interface.^{39,40,41} The protein structures were solvated with the addition of 46,656 modified TIP3P waters⁴² in a cubic box of size of 90 x 90 x 90 Å, and Periodic Boundary Conditions (PBCs)⁴³ were applied with particle-mesh Ewald (PME)⁴⁴ to handle long-range electrostatic interactions ($\kappa = 0.34$, order of B-spline interpolation set to 6). Lennard-Jones interactions were treated with a heuristically updated atom-based cutoff scheme, where the non-bonding pair list was generated to include all atoms within 12 Å, and the switching function was evaluated for atoms between 8 and 12 Å away from the atom of interest. Non-bonding pair lists for atom images were also generated with a 12 Å radius. After waters were initially added, SHAKE was applied to restrain all heavy-atom to hydrogen bonds at parameter defined distances, and a loose/short (50 step) Steepest Descent (SD) minimization protocol was used to minimize bad contacts.

The solvated protein systems were then neutralized with KCl salt at a concentration of 0.15 M. Another short minimization procedure was conducted after neutralizing, to again reduce bad contacts: 10 steps of SD followed by 25 steps of Adopted Basis Newton Raphson minimization (both with default settings). The P450 model systems were then progressively heated, using Domain Decomposition (DOMDEC) compatible CHARMM,³⁷ from 110.15 K to 310.15 K over 100 ps (time step of 1 fs). Once the systems were successfully heated, they were allowed to equilibrate for 2 ns under isothermal-isobaric conditions (NpT ensemble), the Hoover thermostat⁴⁵ was used to maintain temperature fluctuations around 310.15 K.

Two types of post-equilibration simulations were conducted for each P450 system. The first type was a 20 ns simulation of each system, with no substrate present in the P450 binding site (we refer to these simulations as “apoenzyme” or “apo” simulations),

Figure 2: The final two hydroxylation steps in the production of 20E, which results in the initiation of the molting cycle in arthropods.

these were conducted to ensure each model was globally stable over longer time-scales. “Stability” was quantified by monitoring the root mean square deviation (RMSD) and radius of gyration (R_g) over the course of the simulation. Root mean square fluctuations (RMSFs) were also calculated for all residues in each P450 across each 20 ns trajectory (see Supporting Information) to identify which regions were the most rigid/flexible.

Apart from illustrating thermodynamic stability, MD simulations can also be used to refine homology models and generate good starting structures for flexible docking studies. Previous work by Weihua Li et al. has shown that refining homology models via MD simulation can be a necessary step in accurately predicting ligand binding modes in P450 proteins (e.g., CYP2J2).⁴⁶ From our own past experience with flexible docking studies,^{47,48,49,50,51} we have also seen that it is advisable to utilize protein structures with receptor conformations that are predisposed to ligand binding (i.e., “productive” conformations). Past research has shown that apo P450 binding sites in particular can be very malleable,^{52,53} therefore “holo-like” structures were desirable, i.e., protein structures similar to those with a bound ligand, to ensure we had productive receptor conformations for docking. Thus, we also set up structures of CYP315a1 and CYP314a1 homology models with CPS (the template protein’s, CYP24a1, ligand) retained in the active sites. These complexes were then prepared for MD simulation in CHARMM (ParamChem used to generate CPS parameters) according to the steps of solvation, neutralization, heating, and equilibration described above. Then, we conducted 2 ns post-equilibration simulations (referred to as a “holoenzyme” or “holo” simulations, due to the inclusion of CPS) to refine the homology models for docking studies. We chose shorter simulations for refinement than stability testing because, although we needed to maintain productive active site conformations, CPS is not the native ligand of either CYP315a1 or CYP314a1 so we did not want to induce non-native conformations by running long simulations. The resulting final snapshots from these 2 ns holo simulations were used to generate grids for flexible docking studies.

DOMDEC compatible CHARMM was used for all simulations, apo and holo, (except the 20 ns apoenzyme post-equilibration simulation of CYP314a1 which was conducted using the GPU based OpenMM as interfaced with CHARMM).^{37,54} The integration time steps were 2 fs and a snapshot was saved every 1000 steps.

Induced Fit Docking: An induced fit docking (flexible ligand-flexible receptor docking, IFD) protocol was employed, with Schrödinger Software,^{55,56} to predict initial binding modes of the SMIs (Ereb and XA) and the native substrates/products (E, 2-dE, and 20E) within CYP315a1 and CYP314a1 active sites. The following procedures were carried out as part of our IFD protocol; unless otherwise noted, these procedures represent default conditions in Schrödinger. First, all receptor active site residues within 9 Å around the CPS ligand were mutated to alanine, opening the binding site pocket. The choice of a 9 Å mutated region is not a default condition, but our group has seen that specifying a larger region here allows for more accurate binding mode predictions, especially when docking larger ligands.^{47,48} To initiate docking, Glide (Grid-base ligand docking with energetics) receptor grids were generated with CPS identified as the centroid of the protein active site.^{57,58} Schrödinger’s LigPrep 2.3 module was used to pre-process substrate structures

at pH 8 (not default)³ to generate relevant protonation states, tautomers, and energy minimized molecular structures.⁶⁰ GlideSP (Glide Standard Precision) docking was done for all ligands with 0.5 ligand and 0.7 receptor van der Waals scaling.^{57,58} One-hundred partially optimized poses per ligand were requested within the receptor to generate a collection of ligand+P450 complexes. Next, the mutated alanine residues were back-mutated to their respective original residue types in random orientations followed by side chain refinement.^{55,56} Subsequently, the van der Waals radii were scaled back to 1.0 for the ligand+P450 complex, followed by Truncated Newton Conjugate Gradient (TNCG) minimization,³⁰ using the OPLS-2005²⁹ All Atom force field with GB-SA to produce an ensemble of induced fit structures. All complexes within 50 kcal·mol⁻¹ of the lowest energy structure were retained for the final step, discarding other high energy structures. Finally, GlideXP (Glide Extra Precision) re-docking was performed with full van der Waals scaling to predict the relative binding energies represented by Glide scores (GScores).^{57,58,61}

Binding Mode Analysis: To test the stabilities of SMI binding modes that were predicted by IFD, the resulting poses were used as starting coordinates for 4 ns MD simulations (time step of 2 fs and snapshots were saved every 1000 steps, or every 2 ps). We conducted only 4 ns of SMI+P450 complex simulation because we were mostly concerned with allowing the SMI to find stable binding modes. These simulations were set-up and carried out with identical options (PBC/PME, LJ switching options and non-bonding list generation) as described in the “Molecular Dynamics” section above. Root mean square fluctuations (RMSFs) were calculated for each SMI in each binding site, over the final 3 ns of simulation, to illustrate the relative stability of the interactions formed upon binding. These time correlated properties were used to support hypotheses concerning which binding site may be preferred by which SMI.

MM-GBSA: The Schrödinger Suite’s Prime 3.1 module²⁷ was used to conduct molecular mechanics-generalized Born surface area continuum solvation (MM-GBSA) calculations to estimate the binding free energy of each SMI+P450 complex. A binding free energy can be estimated according to $\Delta G_{bind} = E_{complex} - (E_{protein,free} + E_{ligand,free})$. MM-GBSA is used to calculate the individual terms E_X (subscript X denotes protein/ligand/complex) as: $E_X = E_{vac,X} + \Delta G_{solv,X} = E_{vac,X} + \Delta G_{vdW,X} + \Delta G_{elec,X}$. In Prime, VSGB 2.0⁶² is used to model solvation effects implicitly in calculation of the $\Delta G_{vdW,X}$ and $\Delta G_{elec,X}$ components. Although MM-GBSA results are typically only accurate with regard to relative binding free energies, VSGB 2.0’s use in predicting binding free energies has recently been reported to be approaching the accuracy of absolute binding free energies.⁶³ This accuracy has been attributed, in part, to VSGB 2.0’s novel energy function which incorporates bonded and nonbonded parameters from the OPLS-AA (Optimized Parameters for Liquid Simulation - all atom) force field.⁶⁴ VSGB 2.0 represents an improvement upon earlier Prime implicit solvation models (SGB,⁶⁵ SGB-NP,⁶⁶ and VSGB 1.0⁶⁷) by including additional terms to enhance prediction of inter- and intra-molecular hydrogen-bonding geometries as well as π - π stacking geometries and energetics.⁶² Additionally, VSGB 2.0 has shown improved performance in handling solvation effects of long

³The choice of pH 8 for ligand preparation is not a default condition. Mangum and Shick⁵⁹ have previously reported the body fluid pH of marine decapods to be slightly basic pH (between 7.439 - 7.835 pH units), and noted that the pH levels should rise in seasons other than midsummer. Thus, we presumed that pH 8 was appropriate in preparing ligand protonation states.

loop regions.⁶²

We used Prime MM-GBSA calculations in two ways: (1) to estimate the binding free energy of the top pose so that we could compare relative binding free energies between SMI+P450 complexes, and (2) to estimate the binding free energy across the 4 ns SMI+P450 complex trajectories, so that we could observe how conformations affected binding affinity and compare binding energy averages to the estimated binding affinity of the top pose (we should see at least qualitative consistency between the two). Because Prime treats solvent effects implicitly with VSGB 2.0, all TIP3P water molecules were deleted from MD simulation frames prior to MM-GBSA calculations. MM-GBSA energies were then calculated for every 4th saved frame from the 4 ns trajectory. The MM-GBSA energy of the “top pose” was then found by searching the time correlated series for the corresponding frame. Binding free energies were averaged over the final 3 ns of each trajectory (allowing for 1 ns of equilibration).

Top Pose Selection Protocol: In order to qualitatively and quantitatively describe the most likely binding modes of the SMIs in each binding site, we devised a protocol to select one pose from the 4 ns MD trajectory, that most likely represents the true binding mode, we will refer to these selected poses as the “top poses” for expedience. We will first briefly outline the selection protocol, and then describe the procedure with complete detail for the interested reader (see the next paragraph): 1) each 4 ns SMI+P450 simulation was clustered according to an RMSD cut-off using the CORREL module in CHARMM, 2) the representative structure of each cluster (from each simulation) was identified as the frame with the lowest interaction energy between the SMI and P450 binding site, and finally 3) MM-GBSA calculations were used to identify the representative structure with the lowest predicted binding free energy. This workflow is also described pictorially in the Supporting Information (Figure SI4).

The complete description of the top pose selection protocol is as follows: the CORREL module in CHARMM was used to perform pairwise RMSD matrix calculations between every 4th frame of each 4 ns SMI+P450 trajectory. RMSD calculations were done by comparing the frame-to-frame positions on a subset of heavy atoms: specific active site residues, the SMIs, and the heme. Residues His78, Trp84, Phe98, Phe99, Phe232, Asp285, Ala289, Asp292, Thr293, Val340, Ala341, Phe343, Cys414 and Val452 were included in RMSD calculations for CYP315a1. Residues Leu78, Tyr84, Val98, Met99, Pro232, Leu285, Asp289, Asp292, Lys293, Asn340, Ile341, Gln343, Arg414 and Glu452 were included for RMSD calculations in CYP314a1. Next, using a RMSD matrix with a radius cutoff value of 0.15 Å, unique clusters were generated for each trajectory. The structure from each cluster with the lowest interaction energy (calculated in CHARMM between the SMI and the protein) was chosen as the representative structure for that cluster. The representative structures from the five *most populated* clusters were selected to conduct binding energy calculations with MM-GBSA using Schrödinger’s Prime module. For MM-GBSA calculations, all explicit water molecules and ions were deleted as Prime treats solvent implicitly (VSGB 2.0), and the complex was scored in place (minimizing non-polar hydrogens) using Glide. The representative structure with the lowest MM-GBSA binding energy was chosen as the most likely binding mode for each SMI+P450 complex. To ensure that the “top pose” for each complex did indeed represent the average across the final 3 ns of simulation: we qualitatively compared the top-pose binding

energy to the average binding energy obtained in the “Binding Mode Analysis” section. We supplemented MM-GBSA results with “per-residue interaction energy analyses” for each top pose to quantitatively describe particular interactions that contribute to the binding energy. Per-residue interaction energies were calculated in CHARMM between the SMI and all protein residues within 10 Å. Explicit water molecules were included in these calculations to capture potential water mediated interactions.

3. Results and Discussion

Homology model construction: Initial sequence alignments using Clustal Omega,^{68,69} with template sequence CYP24a1 (see Methods), retained 23% identities for both CYP315a1 and CYP314a1. Low percentage sequence identities (i.e., falling below 20%) is a common yet major challenge in the P450 family, even among similar members of the family.⁷⁰ However, signature P450 motifs which include highly conserved familial residues assure that characteristic P450 tertiary structural folds remain conserved.¹⁸ To account for low percentage identity between target and template sequences, the Prime STA (single template alignment) module, which optimizes placement of α -helices and β -sheets, was used. Then, conserved residues were anchored to retain “signature sites”, which are known substrate receptor sites,⁷⁰ and the heme-binding region. Information regarding signature sites was obtained from the earlier sequencing of the closely related P450, CYP306a1, in Kuruma Prawn crustaceans.⁹ The sequences were further physically aligned to remove gaps between α -helices and β -sheets which resulted in slightly increased sequence identities: 26% for CYP315a1 and 27% for CYP314a1. The homology models were then built by transferring the backbone and conserved side chain atoms from template CYP24a1 to query P450 sequences, followed by protein loop refinement.²⁸ The template ligand (CPS) and heme co-factor were retained as part of the build process. Flexible loops were refined independently using Prime serial loop sampling (Supporting Information).

CYP Protein Structures: Final homology models (whose backbone dihedral distributions were verified via Ramachandran plots, Figure SI2, Table SI1) exhibit the conserved tertiary structural elements defined as the “CYP fold”, which consists of α -helices A - L and β -sheets 1 - 5 (Figure 3). Moreover, both models have a heme-binding site comprised of four α -helices (E, I, L, and K) and two α -helices near the heme binding pocket (D and J), which is characteristic of P450 heme sites.⁷¹ Final models were also validated via comparison to models generated by the I-TASSER Web-interface^{32,33,34} for protein structure prediction (Figure 3). The homology models of CYP315a1 and CYP314a1 aligned very well with the highest ranked I-TASSER generated models (ranking determined by I-TASSER), and there was strong agreement between our models and the I-TASSER models with respect to the elements of the CYP fold. CYP315a1 and CYP314a1 both have several residues to stabilize the heme prosthetic group via hydrogen bonding/electrostatic interactions: three arginines, one tryptophan, and one heme-coordinated cysteine. Although both P450s have the same tertiary folds, they still exhibit sequence and structural variability in the heme binding site. This variability is likely correlated to the differences in P450 chemical activity and substrate specificity.⁷¹ Hence, the models of CYP315a1 and CYP314a1 were superimposed to probe active site structural similarities and differences (Figure 4). Additionally, a list of residues included within the active sites is provided in

Figure 3: Overview of tertiary structures revealing the “CYP fold” for CYP315a1 and CYP314a1. The α -helix and the β -sheet labeling is consistent with the CYPED definitions of P450 general structure.⁷¹ Tertiary structural validation of 3D-homology models (purple) with respect to I-TASSER 3D-models (green) and CYP24a1 template structure (yellow). a) CYP315a1 homology model vs. CYP315a1 I-TASSER model, b) CYP315a1 homology model vs. CYP24a1 template structure, c) CYP314a1 homology model vs. CYP314a1 I-TASSER model, and d) CYP314a1 homology model vs. CYP24a1 template structure.

the supporting information (Table SI2). An aromatic cluster comprised of Phe98, Phe99, Trp84, Phe232, His78, and the heme was identified in the binding site of CYP315a1, while CYP314a1 was found to be largely devoid of similar aromatic character.

In many cases, aromatic residues have been shown to assist in structural stability, substrate recognition, secure ligand binding, protein folding, and protein-protein interactions.^{72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83} To illustrate this point, Lanzarotti et al. found that nearly half of proteins listed in the PDB contain trimeric aromatic clusters, and a third of proteins contain tetrameric and higher orders clusters, all of which contribute to tertiary structure definition and stability.⁸⁴ Furthermore, they defined an effective energy function based on distance probability (Boltzmann inversion method) and calculated in-

Figure 4: In order to probe the differences in aromatic character between the CYP315a1 and CYP314a1 active sites, we structurally superimposed our P450 enzyme models and CYP102a1, a) CYP315a1 vs. CYP102a1 b) CYP315a1 vs. CYP314a1 and c) CYP314a1 vs. CYP102a1. The carbons of CYP315a1, CYP314a1, and CYP102a1 are green, purple, and yellow respectively. The rest of the atoms have standard color notation: oxygen in red, nitrogen in blue, sulfur in yellow, and hydrogen in white. Some residues supporting the heme in the active site, including arginine and tryptophan residues as mentioned in the text above, were omitted from this image for clarity though they are present in the models.

teraction energies between pairs of aromatic residues, followed by total stabilization interaction energy calculations at the quantum mechanical MP2/6-31++G** level of theory. They concluded that 4-11 kcal·mol⁻¹ of stabilization energy can be attributed to trimeric, tetrameric, and pentameric aromatic clusters. Although aromatic interactions are often grouped together, several distinct types are observed, each with their own characteristic stabilities: sandwiched, parallel displaced, T-shaped, edge-to-face; cation- π interactions also represent stabilizing interactions found in proteins (Figure 5).^{85,86,87,88,89,90,91,92}

Figure 5: Illustration of stabilizing aromatic (π - π , cation- π) interactions with their corresponding estimated stabilization energies.^{85,89,87,86,90,91,92} It should be noted that all of the π - π interaction energies shown here were calculated in the gas phase, while the cation- π term was calculated in solution:⁸⁵ a) Sandwiched (1.8 kcal·mol⁻¹),⁸⁹ b) edge-to-face (1.6 kcal·mol⁻¹),⁸⁶ c) parallel displaced (2.8 kcal·mol⁻¹),⁸⁹ and d) T-shaped π - π interactions (2.8 kcal·mol⁻¹),⁸⁹ as well as e) cation- π interactions (2.8 kcal·mol⁻¹).⁸⁵

Commonly seen ring-ring centroid distances between aromatic residues in such clusters are between 5-8 Å.⁸⁴ Furthermore, they concluded that aromatic substrates should preferentially bind in protein active sites with significant aromatic character.⁸⁴ We have calculated aromatic ring centroid distances between all small molecules and aromatic protein residues to check if they fall within the mentioned 5-8 Å range, and as Table 4

illustrates they do.

One experimental mutation study, in particular, has demonstrated the importance of aromatic residues as substrate selectors in the CYP102a1, a P450, active site.⁹³ CYP102a1 contains a BC-loop phenylalanine (Phe87). When Phe87 is mutated to other non-polar residues, CYP102a1 has modified substrate specificity and regio-selectivity.⁹³ Thus, it was concluded that Phe87 was functionally critical to CYP102a1 activity. Subsequent studies have compared CYP102a1 to 8,164 other P450s (within the CYPED) and found that 73% of residues that align with Phe87 are non-polar.⁷¹ Of these non-polar residues, phenylalanine and leucine were found to be the most conserved (at 22% each). Additionally, an article by Oostenbrink & coworkers investigated many possible cooperative binding patterns of two molecules of Aflatoxin B₁ (an aromatic compound) in the active site of CYP3A4.⁵² They observed that π - π stacking interactions between the two molecules of AFB₁ and the heme group, or between the two molecules and Phe304, were maintained throughout the course of MD simulations, and in some cases were present in almost the entirety of the trajectory.⁵² They also found that aromatic interactions between ligands and the heme in CYP3A4 resulted in proper orientation of AFB₁ for mono-oxygenation, which is the known function of CYP3A4 (≤ 6 Å from heme Fe).⁵² Thus, we know aromatic interactions can play a major role in P450 substrate recognition and reactivity.

To examine the role of aromatic residues in CYP315a1’s active site, we used the pairwise local structural alignment tool ProBiS (Figure SI3).^{94,95,96} In order to determine the differences in aromaticity between CYP315a1 and CYP314a1 active sites, we have superimposed their binding sites as well as with CYP102a1. We found that CYP315a1’s Phe98 occupies the same structural position (relative to the heme and the overall CYP fold) as CYP102a1’s Phe87 (Figure 4a) and thus may serve a similar biochemical function. Similar to CYP102a1’s Phe87, CYP315a1’s Phe98 may play a role in the recognition and stabilization of the substrates and uncharged aromatic inhibitors over the heme active center. For example, when Phe98 forms a π stacking interaction with a substrate, it also interacts with Phe99 via aromatic interactions. Therefore, the binding of a ligand with an aromatic moiety (e.g., our SMIs) to the CYP315a1 active site, has the potential to add a “layer” to the established aromatic cluster (Phe98, Phe99, Trp84, Phe232, His78, and heme), and thus provide additional stability for said ligand upon binding. Comparing CYP102a1 and CYP314a1 shows that Val114 occupies the Phe87 position (Figure 4c). Out of the P450s examined in the CYPED, 12% had a Val at this structural position, indicating Val114 is semi-conserved amongst P450s.⁷¹

MD simulations: To monitor system stability, as described in the Methods section, a 20 ns apoenzyme simulation was conducted for each homology model. R_g and RMSD values (Figure SI8 and SI9) converged during the 20 ns simulations indicating that both systems were stable and well equilibrated. To quantify dynamics on a per-residue basis, RMSFs of all heavy atoms were calculated over the course of the simulations. The RMSF plots for CYP315a1 and CYP314a1 (Figure SI10) show that residues from both active sites are relatively rigid throughout the simulation (i.e., $\text{RMSF} < 2$ Å). Following MD simulations, Ramachandran plots were again compiled. We observed an increase in population of residues within the most favored regions for ϕ - ψ distribution: from 87% to 90% for CYP315a1 and 84% to 89% for CYP314a1 after MD Simulation (Figure SI11, Table SI3).

Native substrate/product binding, orientation, and stability: Binding sites for docking simulations were again taken from the final structures of 2 ns holoenzyme simulations. To illustrate that our constructed homology models retained their known function, the native substrates/products were docked in their respective active sites using the IFD protocol. All resulting poses were then ranked according to the Glide Score (an estimate of the binding energy). The lowest energy poses were visually inspected for “chemically intuitive” geometries. As expected, when using a simulated protein-ligand complex (i.e., a simulated holoenzyme complex rather than an apoenzyme) for flexible docking studies (see Methods), intuitive binding geometries of the native substrates/products were obtained. Visual inspection of IFD poses was supplemented with per-residue interaction energy analyses in CHARMM to provide insight on active site residues’ stabilizing/destabilizing contributions towards substrate binding (Table SI6 and SI7).

The reactant (2-dE) and product (E) retained the same overall conformation in CYP315a1 (Figure 6a, Table SI6), with Phe98 stabilizing the conjugated ketone moieties of 2-dE and E via π stacking (2-dE: $-4.05 \text{ kcal}\cdot\text{mol}^{-1}$, E: $-7.01 \text{ kcal}\cdot\text{mol}^{-1}$). Furthermore, three residues (Asp285, Asp292, Met451) stabilize the reactant/product in CYP315a1 active site via water mediated hydrogen bonds to the 14-hydroxy and 22-hydroxy groups of both 2-dE and E (2-dE to Asp285: $-15.94 \text{ kcal}\cdot\text{mol}^{-1}$, 2-dE to Asp292 is not stabilizing, 2-dE to Met451: $-5.12 \text{ kcal}\cdot\text{mol}^{-1}$, E to Asp285: $-10.05 \text{ kcal}\cdot\text{mol}^{-1}$, E to Asp292 $-9.74 \text{ kcal}\cdot\text{mol}^{-1}$, E to Met451: $-4.49 \text{ kcal}\cdot\text{mol}^{-1}$). Aspartate residues, along with π - π support from Phe98, appear to be important for holding 2-dE in a favorable orientation for hydroxylation at the C₂ position and the calculated distance between the C₂ carbon and heme iron (2-dE: $4.42 \text{ \AA}$; E: $4.71 \text{ \AA}$) reflects this known behavior. The native substrates also interact favorably via van der Waals interactions to the heme (2-dE: $-11.58 \text{ kcal}\cdot\text{mol}^{-1}$, E: $-11.00 \text{ kcal}\cdot\text{mol}^{-1}$).

In the case of CYP314a1, the flexible docking protocol using holoenzyme models again was able to predict substrate/product binding modes that clearly reflect hydroxylation at the C₂₀ position of E (Figure 6b, Table SI7). The C₂₀ atom of both substrate and product is oriented above the heme iron and at a proper distance for hydroxylation (E: $5.27 \text{ \AA}$; 20E: $5.25 \text{ \AA}$). For both E and 20E in CYP314a1, the central gonane structure is stabilized by van der Waals interactions from Leu480 (E: $-3.72 \text{ kcal}\cdot\text{mol}^{-1}$, 20E: $-4.15 \text{ kcal}\cdot\text{mol}^{-1}$), van der Waals as well as electrostatic stabilization from Ala302 (E: $-5.56 \text{ kcal}\cdot\text{mol}^{-1}$, 20E: $-9.06 \text{ kcal}\cdot\text{mol}^{-1}$) and Ala303 (E: $-2.97 \text{ kcal}\cdot\text{mol}^{-1}$, 20E: $-9.01 \text{ kcal}\cdot\text{mol}^{-1}$), and electrostatic stabilization from Glu306 (E: $-5.99 \text{ kcal}\cdot\text{mol}^{-1}$, 20E: $-9.00 \text{ kcal}\cdot\text{mol}^{-1}$). Waters in the active site mediate hydrogen bonding between E and Cys368 ($-5.34 \text{ kcal}\cdot\text{mol}^{-1}$) and Ala209 ($-2.64 \text{ kcal}\cdot\text{mol}^{-1}$). Similarly, waters mediate hydrogen bonding between 20E and Pro92 ($-4.79 \text{ kcal}\cdot\text{mol}^{-1}$), Gln213 ($-5.71 \text{ kcal}\cdot\text{mol}^{-1}$), and Thr365 ($-6.90 \text{ kcal}\cdot\text{mol}^{-1}$). Whereas 2-dE and E are stabilized by Phe98 through π - π interactions, a similar aromatic substrate recognizing residue is not present in the CYP314a1 active site.

Ecdysteroidogenesis regulation/inhibition: Naya et al. have proposed that XA modulates 20E release by regulating CYP315a1 or CYP314a1, whereas Baker and co-workers have posited that CYP314a1 may be the target.^{3,1} As discussed, Ereb (isolated from Antarctic sea sponge) is also capable of inhibiting the ecdysteroidogenic pathway.^{1,4,5} We have identified the likely protein target for each of the SMI using our computational

Figure 6: Orientations of native substrate/product in both models a)CYP315a1 and b)CYP314a1. Substrate/product carbons for 2-dE in yellow, E in green, and 20E in pale yellow.

approach.

IFD provided initial poses for Erebin and XA in CYP315a1 and CYP314a1. To gain additional insight about SMI+P450 interactions, we conducted 4 ns MD simulations from resultant IFD structures. We observed, based on ring-ring distances Figure SI16, that within the first 1 ns of simulation in CYP315a1, Erebin adjusted its positioning relative to Phe98 and the heme, thereby optimizing its distance from both active-site aromatic features (time correlated distance plots given in SI, Figure SI16). For this reason, analysis for each SMI+P450 simulation was done over the final 3 ns of the 4 ns simulations. We computed heavy-atom SMI RMSFs and average MM-GBSA energies over the final 3 ns, and isolated top poses according to the clustering scheme described in the Methods. Table 1 summarizes the RMSF and MM-GBSA results, additionally the top poses are described in detail in the following subsections. The $\Delta G_{\text{bind}}(\text{avg})$ for Erebin in CYP315a1 is $-47.54 \text{ kcal}\cdot\text{mol}^{-1}$ versus $-36.43 \text{ kcal}\cdot\text{mol}^{-1}$ in CYP314a1. This is a difference in binding free energy of over $10 \text{ kcal}\cdot\text{mol}^{-1}$, suggesting Erebin would preferentially bind to CYP315a1. In the case of XA, both binding free energies are very similar: $-22.83 \text{ kcal}\cdot\text{mol}^{-1}$ in CYP315a1 and $-24.45 \text{ kcal}\cdot\text{mol}^{-1}$, suggesting perhaps no preferential target.

It is interesting to speculate about how XA and Erebin's different binding free energies in either active site reflects their different roles in the crustacean. Erebin is a proposed inhibitor of ecdysis, while XA is a regulator. Thus, it might be expected that Erebin binds with higher binding affinity to ecdysis proteins than XA because it should ideally cause nearly-irreversible inhibition. However, XA as a regulator should bind to these active sites without permanently disrupting their function. The predicted binding free energies in Table 1 support these similar yet distinct roles.

In the next two subsections we will describe the top poses for each SMI+P450 complex. Additionally, we will provide a quantified means for rationalizing these poses, using "per-

residue interaction energies”. Past articles, by Florián & coworkers⁹⁷ and Bren & coworkers,⁹⁸ have carefully demonstrated that a free energy, such as a binding free energy, can often be decomposed into distinct contributing terms (e.g., group-to-group electrostatic or van der Waals terms). They also illustrate that, through free energy decomposition, one can more clearly elucidate dominant interactions when examining solute/environment affinities.⁹⁷ Therefore, in addition to visual inspection of all poses, we have also calculated (in CHARMM) electrostatic and van der Waals interaction energies between the docked SMI and each residue within 10Å, we refer to these results as “per-residue interaction energies”, Tables 2 and 3.

Figure 7: The top scoring SMI poses in the CYP315a1 active site: a) Erebinone, b) XA.

Erebinone: The following describes the top pose of Erebinone in CYP315a1 (Figure 7a) with the aromatic ring-ring distances and interaction energies given (Table 2a) for this particular pose. Table 4 lists the average aromatic ring-ring distances taken over the final 3 ns of the 4 ns SMI+P450 simulation. Erebinone’s aromatic moiety forms a parallel displaced interaction to the heme plane below (ring centroid distances to heme 5.57 Å). The orientation of the aromatic moiety over the heme also allows for the pyrrole nitrogens to electrostatically interact with Erebinone’s amine substituent, resulting in -25.58 kcal·mol⁻¹ of stabilization from the heme. Interaction energy analysis also reveals a water mediated hydrogen bond between Erebinone’s hydroxyl and Thr293’s side chain (-10.79 kcal·mol⁻¹). The tail carbonyl of Erebinone forms another water mediated hydrogen bond to Asp285 (-4.44 kcal·mol⁻¹). Finally, Erebinone participates in an edge-to-face aromatic interaction with Phe98 (5.06 Å, -3.35 kcal·mol⁻¹) as well as an edge-to-face aromatic interaction with His78 (7.91 Å, -0.18 kcal·mol⁻¹). Erebinone’s interactions with Phe98 and His78 represents not only a stabilizing interaction from two residues, but also an additional layer to the already established aromatic network: Phe98 is itself stabilized by pseudo edge-to-face aromatic

Figure 8: The top scoring SMI poses in the CYP314a1 active site: a) Erebia, b) XA.

interactions with Phe99 (6.47 Å), Phe99 forms a pseudo T-shaped interaction to Trp84 (6.63 Å), and Trp84 establishes a pseudo edge-to-face interaction with Phe232 (5.04 Å) forming an aromatic cluster of hexameric order upon binding of Erebia (Table 4, Figure 7).

The top pose of Erebia in CYP314a1 (Figure 8a, Table 3a) is defined largely by hydrogen bonding interactions along with a single parallel displaced aromatic interaction with the heme. Erebia's aromatic ring is oriented such that its hydroxyl and amine substituents can interact electrostatically with the heme pyrrole nitrogens below, similar to the binding of Erebia in CYP315a1. This aromatic interaction coupled with electrostatics results in -21.99 kcal·mol⁻¹ of stabilization energy from the heme. Furthermore, Erebia's hydroxyl aromatic substituent donates a hydrogen bond to Thr307's side chain (-7.33 kcal·mol⁻¹). The tail carbonyl of Erebia also accepts dual hydrogen bonds from backbone -NH groups of Phe367 (-4.73 kcal·mol⁻¹) and Cys368 (-5.44 kcal·mol⁻¹). Finally, the tail amine moiety forms a hydrogen bond network with two sequential bridging waters to the backbone carbonyl of Asp220 (-7.71 kcal·mol⁻¹).

In summary, comparing average binding free energies (Table 1) for the two Erebia complexes, we predict that Erebia will preferentially target CYP315a1. Careful consideration of the interaction energies provides further explanation: Erebia interacts with the aromatic cluster present in CYP315a1 and additionally accesses hydrogen bonds, thus being dually stabilized in the CYP315a1 binding site. In CYP314a1, Erebia is stabilized by interactions with the heme below and hydrogen bonds, but there is no large aromatic cluster to provide further stabilization.

XA: Similar average binding free energies for XA bound in CYP315a1 and CYP314a1 indicate that XA may bind both active sites (Table 1). The following describes the top

pose of XA in CYP315a1 (Figure 7b, Table 2b) again with instantaneous aromatic ring-ring distances listed for the top pose (Figure SI16), however average ring-ring distance can be found in Table 4. XA interacts with heme in a parallel displaced fashion (3.49 Å) while XA’s carboxylate group interacts with one of the heme’s propionate groups via two bridging water molecules, resulting in an overall stabilizing contribution from heme of -17.92 kcal·mol⁻¹. XA’s carboxylate also interacts with Pro406 via two successive hydrogen bonds from bridging water molecules, for a total interaction energy of -18.14 kcal·mol⁻¹. XA’s ketone moiety is oriented to accept a hydrogen bond from Thr293’s hydroxyl side chain (-10.98 kcal·mol⁻¹). Electrostatic interactions are seen between XA’s carboxylate tail and Arg346 (-5.21 kcal·mol⁻¹). Though XA is oriented in a T-shaped fashion with respect to Phe98, the interaction energy from this residue is not as stabilizing as the Ere_b - Phe98 interaction (-2.50 kcal·mol⁻¹ vs. -3.35 kcal·mol⁻¹ for Ere_b, 5.97 Å XA to Phe98 distance) likely as a result of XA’s hydroxyl and amine moieties that are between the aromatic rings of Phe98 and XA. XA is oriented with an edge-to-face interaction to His78, (10.91 Å, -0.07 kcal·mol⁻¹). The aromatic network was again identifiable in the CYP315a1 and XA complex: Phe98 edge to face to Phe99 (5.60 Å), Phe99 parallel displaced to Trp84 (6.16 Å), and Phe232 edge to face to Trp84 (5.30 Å) (Table 4).

The top pose for XA in CYP314a1 (Figure 8b, Table 3b), reveals binding in this active site is primarily driven by hydrogen bonding and electrostatic interactions. XA’s carboxylate forms hydrogen bonds with the backbone amine groups of Phe367 (-19.96 kcal·mol⁻¹) and Cys368 (-6.20 kcal·mol⁻¹). The same XA carboxylate also participates in consecutive water mediated hydrogen bonds with the backbone carbonyl of Leu480 (-7.43 kcal·mol⁻¹). Further, XA’s ring hydroxyl substituent, participates in a water mediated hydrogen bond (through two water molecules) to Pro92’s backbone carbonyl (-8.42 kcal·mol⁻¹). Finally, the parallel displaced interaction between XA and the heme accounts for -6.12 kcal·mol⁻¹. Thus, the charged groups of XA encourage stabilization via hydrogen bonding and electrostatic interactions. However, XA’s two fused aromatic rings and may still benefit from interaction with the aromatic network as does Ere_b in CYP315a1.

4. Conclusions:

Homology models of CYP315a1 and CYP314a1 (*Drosophila*) were constructed using CYP24a1 as a template. Short (2 ns) MD simulations of holoenzyme homology models (using the template’s ligand as a “stand in”) were performed to generate refined structures with “productive” binding sites for later docking studies; apoenzyme homology models were simulated for 20 ns to monitor system stability via RMSD and R_g calculations. Additionally, RMSF analyses from the apoenzyme simulations have shown that both active sites are fairly rigid (with RMSF’s < 2 Å) which can be important for substrate recognition and binding. Structure and sequence comparisons to the CYPED database and CYP102a1 crystal structure using I-TASSER models also show that our homology modeling plus MD simulation methodology predicted the characteristic P450 structural elements. ProBiS binding site analysis and structure superimposition revealed that CYP315a1’s Phe98 should play a similar role to CYP102a1’s Phe87 in mediating substrate selectivities and substrate binding orientation.

Substrates/products/SMIs were docked in CYP315a1 and CYP314a1 using a flexible receptor-flexible ligand docking protocol. The conjugated ketone moieties of 2-dE

and E were seen to interact favorably via π -stacking with Phe98 in CYP315a1. This interaction secures 2-dE over the heme iron at a distance and orientation appropriate for hydroxylation at the C-2 carbon, which supports the hypothesis that Phe98 acts as a substrate recognizing residue. The top pose of EreB in CYP315a1 also features an aromatic interaction between the SMI and Phe98, which represents an addition to the already established aromatic network. We therefore propose that this EreB-Phe98 interaction is key to the inhibition of CYP315a1 by EreB. Additionally, the estimated binding free energies for EreB in CYP315a1 and CYP314a1 suggest that EreB would bind preferentially to CYP315a1. We hypothesize that this preferential binding is largely due to aromatic interactions between EreB and CYP315a1, a hypothesis which we support with interaction energy analysis. Due to XA’s aromatic as well as charged functional groups, XA can optimize aromatic and electrostatic interactions in both active sites. Thus we hypothesize that XA may not have a strong preference for either of the P450s, which is supported by the similar low binding energies for XA in either active site. In short, we have concluded that XA may bind to both active sites, while aromatic interactions drive binding of EreB in CYP315a1. Thus, inhibition of the production of 20E by exogenous EreB may be a result of EreB’s aromatic interactions with Phe98 and the heme group of CYP315a1.

5. Supporting Information:

The following can be found in SI: Loop buildup algorithm details, CPS ligand structure, Ramachandran plots for homology models and final structures from MD simulations (holoenzyme and apoenzyme structures), details of structural superimposition results from ProBiS, trajectory analyses of 20 ns apo state simulations, detailed description of top pose selection, MM-GBSA results across 4 ns trajectory of simulated complexes, loop refinement details, methods for water simulation for electrostatic interaction calculation, and per residue interaction energy analyses for the SMIs/substrates/products. This material is available online free of charge.

6. Acknowledgments:

The authors would like to acknowledge NIH (1K22HL088341-01A1), NIH (5K22HL113045), NSF (CHE-1156853, CHE-1531590, CHE-1464946), and the University of South Florida (start-up) for funding. Computations were performed on local USF computers at the USF Research Computing Center. Also, this work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575 (XSEDE grant MCB120133).

References

- [1] B. Moon, Y. C. Park, J. B. McClintock, B. J. Baker, Structure and Bioactivity of Erebusinone, a Pigment from the Antarctic Sponge *Isodictya erinacea*., *Tetrahedron*. *56* (2000) 9057–9062.
- [2] C. D. Amsler, J. B. McClintock, B. J. Baker, Secondary Metabolites as Mediators of Trophic Interactions Among Antarctic Marine Organisms., *Amer. Zool.* *41* (2001) 17–26.
- [3] Y. Naya, W. Miki, M. Ohnishi, M. Ikeda, K. Nakanishi, Endogenous Xanthurenic Acid as a Regulator of the Crustacean Molt Cycle., *Pure Appl. Chem.* *61* (1989) 465–465.
- [4] Y. Naya, M. Ohnishi, M. Ikeda, W. Miki, K. Nakanishi, What is Molt-inhibiting hormone? The Role of an Ecdysteroidogenesis Inhibitor in the Crustacean Molting Cycle., *Proc. Natl. Acad. Sci. USA* *86* (1989) 6826–6829.
- [5] Y. Ohnishi M. N., Effect of Xanthurenic Acid on P450-dependent Biotransformation by Molting Glands In Vitro., *Experientia*. *50* (1994) 654–657.
- [6] Y. Naya, K. Kishida, M. Sugiyama, M. Murata, W. Miki, M. Ohnishi, K. Nakanishi, Endogenous Inhibitor of Ecdysone Synthesis in Crabs., *Experientia*. *44* (1988) 50–52.
- [7] E. S. Chang, B. A. Sage, J. D. O'Connor, The Qualitative and Quantitative Determinations of Ecdysones in Tissues of the Crab, *Pachygrapsus Crassipes*, Following Molt Induction., *Gen Comp. Endocrinol.* *30* (1976) 21–33.
- [8] L. I. Gilbert, J. T. Warren, A Molecular Genetic Approach to the Biosynthesis of the Insect Steroid Molting Hormone., *Vitam Horm.* *73* (2005) 31–57.
- [9] H. Asazuma, S. Nagata, H. Nagasawa, Inhibitory Effect of Molt-inhibiting Hormone on Phantom Expression in the Y-Organ of the Kuruma Prawn, *Marsupenaeus Japonicus*., *Arch. Insect Biochem. Physiol.* *72* (2009) 220–233.
- [10] R. Niwa, T. Sakudoh, T. Namiki, K. Saida, Y. Fujimoto, H. Kataoka, The Ecdysteroidogenic P450 CYP302a1/Disembodied from the Silkworm, *Bombyx Mori*, is Transcriptionally Regulated by Prothoracicotropic Hormone., *Insect Mol. Biol.* *14* (2005) 563–571.
- [11] E. Spaziani, H. H. Rees, W. L. Wang, R. D. Watson, Evidence that Y-organs of the Crab Cancer *Antennarius* Secrete 3-Dehydroecdysone., *Mol. Cell. Endocrinol.* *66* (1989) 17–25.
- [12] C. Dauphin-Villemant, D. Bocking, C. Blais, J. Y. Toullec, R. Lafont, Involvement of a 3β -hydroxysteroid Dehydrogenase Activity in Ecdysteroid Biosynthesis., *Molec. Cell. Endocrinol.* *128* (1997) 139–149.
- [13] F. Lachaise, M. F. Meister, C. Hétru, R. Lafont, Studies on the Biosynthesis of Ecdysone by the Y-organs of *Carcinus maenas*., *Molec. Cell. Endocrinol.* *45* (1986) 253–261.

- [14] L. Wang, T. Jiang, D. Gusfield, A More Efficient Approximation Scheme for Tree Alignment., *SIAM J. Comput.* *30* (2000) 283–299.
- [15] E. Spaziani, T. C. Jegla, W. L. Wang, J. A. Booth, S. M. Connolly, C. C. Conrad, M. J. Dewall, C. M. Sarno, D. K. Stone, R. Montgomery, Further Studies on Signaling Pathways for Ecdysteroidogenesis in Crustacean Y-Organs., *Amer. Zool.* *41* (2001) 418–429.
- [16] K. F. Rewitz, L. I. Gilbert, Daphnia Halloween Genes that Encode Cytochrome P450s Mediating the Synthesis of the Arthropod Molting Hormone: Evolutionary Implications., *BMC Evol. Biol.* *8* (2008) 1–8.
- [17] K. F. Rewitz, R. Rybczynski, J. T. Warren, L. I. Gilbert, The Halloween Genes Code for Cytochrome P450 Enzymes Mediating Synthesis of the Insect Molting Hormone., *Biochem. Soc. Trans.* *34* (2006) 1256–1260.
- [18] K. F. Rewitz, B. Styrishave, A. Lobner Olsen, O. Andersen, Marine Invertebrate Cytochrome P450: Emerging Insights from Vertebrate and Insects Analogies., *Comp. Biochem. Physiol. C* *143* (2006) 363–381.
- [19] P. Hlavica, Insect Cytochromes P450: Topology of Structural Elements Predicted to Govern Catalytic Versatility., *J. Inorg. Biochem.* *105* (2011) 1354–1364.
- [20] J. T. Warren, A. Petryk, G. Marqués, J. P. Parvy, T. Shinoda, K. Itoyama, J. Kobayashi, M. Jarcho, Y. Li, M. B. O’Connor, C. Dauphin-Villemant, L. I. Gilbert, Phantom Encodes the 25-hydroxylase of *Drosophila Melanogaster* and *Bombyx Mori*: A P450 Enzyme Critical in Ecdysone Biosynthesis., *Insect Biochem. Mol. Biol.* *34* (2004) 991–1010.
- [21] T. Yoshiyama, T. Namiki, K. Mita, H. Kataoka, R. Niwa, Neverland is an Evolutionally Conserved Rieske-domain Protein that is Essential for Ecdysone Synthesis and Insect Growth., *Development.* *133* (2006) 2565–2574.
- [22] J. T. Warren, A. Petryk, G. Marqués, M. Jarcho, J. P. Parvy, C. Dauphin Villemant, M. B. O’Connor, L. I. Gilbert, Molecular and Biochemical Characterization of Two P450 Enzymes in the Ecdysteroidogenic Pathway of *Drosophila Melanogaster*., *Proc. Natl. Acad. Sci. USA* *99* (2002) 11043–11048.
- [23] A. Petryk, J. T. Warren, G. Marqués, M. P. Jarcho, L. I. Gilbert, J. Kahler, J. P. Parvy, Y. Li, C. Dauphin-Villemant, M. B. O’Connor, Shade is the *Drosophila* P450 Enzyme that Mediates the Hydroxylation of Ecdysone to the Steroid Insect Molting Hormone 20-hydroxyecdysone., *Proc. Natl. Acad. Sci. USA* *100* (2003) 13773–13778.
- [24] S. Misra, M. A. Crosby, C. J. Mungall, B. B. Matthews, K. S. Campbell, P. Hradecky, Y. Huang, J. S. Kaminker, G. H. Millburn, S. E. Prochnik, C. D. Smith, J. L. Tupy, E. J. Whitfield, L. Bayraktaroglu, B. P. Berman, B. R. Bettencourt, S. E. Celniker, A. D. N. J. de Grey, R. A. Drysdale, N. L. Harris, J. Richter, S. Russo, A. J. Schroeder, S. Q. Shu, M. Stapleton, C. Yamada, M. Ashburner, W. M. Gelbart, G. M. Rubin, S. E. Lewis, Annotation of the *Drosophila Melanogaster* Euchromatic Genome: A Systematic Review., *Genome Biol.* *3* (2002) 1–22.

- [25] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs., *Nucleic Acids Res.* *25* (1997) 3389–3402.
- [26] J. L. Sussman, D. Lin, J. Jiang, N. O. Manning, J. Prilusky, O. Ritter, E. E. Abola, Protein Data Bank (PDB): Database of Three-dimensional Structural Information of Biological Macromolecules., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* *54* (1998) 1078–1084.
- [27] Schrödingers, LLC, New York, NY, Prime., version 3. Edition (2012).
- [28] M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. F. Day, B. Honig, D. E. Shaw, R. A. Friesner, A Hierarchical Approach to All-Atom Protein Loop Prediction., *Proteins.* *55* (2004) 351–367.
- [29] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.* *118* (1996) 11225–11236.
- [30] L. Grippo, F. Lampariello, S. Lucidi, A truncated newton method with nonmonotone line search for unconstrained optimization, *Journal of Optimization Theory and Applications* *60* (1989) 401–419.
- [31] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures., *J. Appl. Crystallogr.* *26* (1993) 283–291.
- [32] Y. Zhang, I-TASSER server for protein 3D structure prediction., *BMC bioinformatics* *9* (2008) 40–48.
- [33] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction., *Nat. Protoc.* *5* (2010) 725–738.
- [34] A. Roy, J. Yang, Y. Zhang, COFACTOR: an accurate comparative algorithm for structure-based protein function annotation., *Nucleic Acids Res.* *40* (2012) W471–W477.
- [35] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, CHARMM: The Biomolecular Simulation Program., *J. Comp. Chem.* *30* (2009) 1545–1614.
- [36] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations., *J. Comp. Chem.* *4* (1983) 187–217.

- [37] A.-P. Hynninen, M. F. Crowley, New faster CHARMM molecular dynamics engine., *J. Comp. Chem.* *35* (2014) 406–413.
- [38] B. T. Miller, R. P. Singh, J. B. Klauda, M. Hodoscek, B. R. Brooks, I. Woodcock H. Lee, CHARMMing: A New, Flexible Web Portal for CHARMM., *J. Chem. Inf. Model.* *48* (2008) 1920–1929.
- [39] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell, CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields., *J. Comput. Chem.* *31* (2010) 671–690.
- [40] K. Vanommeslaeghe, A. D. MacKerell Jr, Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing., *J. Chem. Inf. Model.* *52* (2012) 3144–3154.
- [41] K. Vanommeslaeghe, E. P. Raman, A. D. MacKerell Jr, Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges, *J. Chem. Inf. Model.* *52* (2012) 3155–3168.
- [42] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* *102* (1998) 3586–3616.
- [43] A. Brünger, C. L. Brooks, M. Karplus, Stochastic boundary conditions for molecular dynamics simulations of st2 water, *Chem. Phys. Lett.* *105* (1984) 495–500.
- [44] T. Darden, D. York, L. Pedersen, Particle mesh ewald: An $n\log(n)$ method for ewald sums in large systems, *J. Chem. Phys.* *98* (1993) 10089–10092.
- [45] W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A* *31* (1985) 1695–1697.
- [46] W. Li, Y. Tang, H. Liu, J. Cheng, W. Zhu, H. Jiang, Probing Ligand Binding Modes of Human Cytochrome P450 2J2 by Homology Modeling, Molecular Dynamics Simulation, and Flexible Molecular Docking., *Proteins.* *71* (2008) 938–949.
- [47] S. L. Vankayala, J. C. Hargis, H. L. Woodcock, Unlocking the Binding and Reaction Mechanism of Hydroxyurea Substrates as Biological Nitric Oxide Donors., *J. Chem. Inf. Model.* *52* (2012) 1288–1297.
- [48] S. L. Vankayala, J. C. Hargis, H. L. Woodcock, How Does Catalase Release Nitric Oxide? A Computational Structure Activity Relationship Study., *J. Chem. Inf. Model.* *53* (2013) 2951–2961.

- [49] G. T. Pauly, N. A. Loktionova, Q. Fang, S. L. Vankayala, W. C. Guida, A. E. Pegg, Substitution of Aminomethyl at the Meta-position Enhances the Inactivation of O6-Alkylguanine-DNA Alkyltransferase by O6-Benzylguanine., *J. Med. Chem.* *51* (2008) 7144–7153.
- [50] J. C. Hargis, S. L. Vankayala, J. K. White, H. L. Woodcock, Identification and Characterization of Non-covalent Interactions that Drive Binding and Specificity in DD-Peptidases and β -Lactamases., *J. Chem. Theory Comput.* *10* (2014) 855–864.
- [51] S. L. Vankayala, Computational Approaches for Structure Based Drug Design and Protein Structure-Function Prediction, Graduate thesis and dissertation, University of South Florida (2013).
- [52] U. Bren, J. E. Fuchs, C. Oostenbrink, Cooperative binding of aflatoxin b1 by cytochrome p450 3a4: A computational study, *Chem. Res. in Toxicol.* *27* (2014) 2136–2147.
- [53] U. Bren, C. Oostenbrink, Cytochrome p450 3a4 inhibition by ketoconazole: Tackling the problem of ligand cooperativity using molecular dynamics simulations and free-energy calculations, *J. Chem. Inf. Model.* *52* (2012) 1573–1582.
- [54] P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, et al., OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation, *J. Chem. Theory Comput.* *9* (2013) 461–469.
- [55] M. P. Jacobson, R. A. Friesner, Z. Xiang, B. Honig, On the Role of the Crystal Environment in Determining Protein Side-chain Conformations., *J. Mol. Biol.* *320* (2002) 597–608.
- [56] M. P. Jacobson, G. A. Kaminski, R. A. Friesner, C. S. Rapp, Force Field Validation Using Protein Side Chain Prediction., *J. Phys. Chem. B* *106* (2002) 11673–11680.
- [57] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy, *J. Med. Chem.* *47* (2004) 1739–1749.
- [58] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, J. L. Banks, Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening, *J. Med. Chem.* *47* (2004) 1750–1759.
- [59] C. Mangum, J. Shick, The pH of body fluids of marine invertebrates, *Comp. Biochem. Physiol., A: Comp. Physiol.* *42* (1972) 693–697.
- [60] L. Schrödinger, LigPrep, version 2.3 (2009).
- [61] R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz, Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for proteinligand complexes, *J. Med. Chem.* *49* (2006) 6177–6196.

- [62] J. Li, R. Abel, K. Zhu, Y. Cao, S. Zhao, R. A. Friesner, The vsqb 2.0 model: A next generation energy model for high resolution protein structure modeling, *Proteins: Struct., Funct., and Bioinf.* **79** (2011) 2794–2812.
- [63] C. Mulakala, V. N. Viswanadhan, Could mm-gbsa be accurate enough for calculation of absolute protein/ligand binding free energies?, *J. Mol. Graph. Model.* **46** (2013) 41–51.
- [64] M. J. Robertson, J. Tirado-Rives, W. L. Jorgensen, Improved peptide and protein torsional energetics with the opl-aa force field, *J Chem. Theory Comput.* **11** (2015) 3499–3509.
- [65] A. Ghosh, C. S. Rapp, R. A. Friesner, Generalized born model based on a surface integral formulation, *J. Phys. Chem. B* **102** (1998) 10983–10990.
- [66] K. Zhu, D. L. Pincus, S. Zhao, R. A. Friesner, Long loop prediction using the protein local optimization program, *Proteins: Struct., Funct., and Bioinf.* **65** (2006) 438–452.
- [67] K. Zhu, M. R. Shirts, R. A. Friesner, Improved methods for side chain and loop predictions via the protein local optimization program: variable dielectric model for implicitly improving the treatment of polarization effects, *J. Chem. Theory Comput.* **3** (2007) 2108–2119.
- [68] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, D. G. Higgins, Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments using Clustal Omega., *Mol. Syst. Biol.* **7** (2011) 539–545.
- [69] M. Goujon, H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, R. Lopez, A New Bioinformatics Analysis Tools Framework at EMBL-EBI., *Nucleic Acids Res.* **38** (2010) W695–W699.
- [70] S. E. Graham, J. A. Peterson, How similar are P450s and what can their differences teach us?, *Arch. Biochem. Biophys.* **369** (1999) 24–29.
- [71] D. Sirim, M. Widmann, F. Wagner, J. Pleiss, Prediction and Analysis of the Modular Structure of Cytochrome P450 Monooxygenases., *BMC Struct. Biol.* **10** (2010) 34–45.
- [72] L. Serrano, M. Bycroft, A. R. Fersht, Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles., *J. Mol. Biol.* **218** (1991) 465–475.
- [73] S. Burley, G. Petsko, Aromatic-aromatic interaction: a mechanism of protein structure stabilization, *Science* **229** (1985) 23–28.
- [74] R. Bhattacharyya, U. Samanta, P. Chakrabarti, Aromatic-aromatic interactions in and around alpha-helices., *Protein Eng.* **15** (2002) 91–100.
- [75] S. Aravinda, N. Shamala, C. Das, A. Sriranjini, I. L. Karle, P. Balaram, Aromatic-aromatic interactions in crystal structures of helical peptide scaffolds containing projecting phenylalanine residues., *J. Am. Chem. Soc.* **125** (2003) 5308–5315.

- [76] L. A. Eidenschink, B. L. Kier, N. H. Andersen, Determinants of fold stabilizing aromatic-aromatic interactions in short peptides., *Adv. Exp. Med. Biol.* *611* (2009) 73–74.
- [77] L. M. Espinoza-Fonseca, J. García-Machorro, Aromatic-aromatic interactions in the formation of the MDM2-p53 complex., *Biochem. Biophys. Res. Commun.* *370* (2008) 547–551.
- [78] A. F. Pereira de Araujo, T. C. Pochapsky, B. Joughin, Thermodynamics of interactions between amino acid side chains: experimental differentiation of aromatic-aromatic, aromatic-aliphatic, and aliphatic-aliphatic side-chain interactions in water., *Biophys. J.* *76* (1999) 2319–2328.
- [79] R. Chelli, F. L. Gervasio, P. Procacci, V. Schettino, Stacking and T-shape Competition in Aromatic–Aromatic Amino Acid Interactions, *J. Am. Chem. Soc.* *124* (2002) 6133–6143.
- [80] J. B. Mitchell, C. L. Nandi, I. K. McDonald, J. M. Thornton, S. L. Price, Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding?, *J. Mol. Biol.* *239* (1994) 315–331.
- [81] N. Kannan, S. Vishveshwara, Aromatic clusters: a determinant of thermal stability of thermophilic proteins., *Protein Eng.* *13* (2000) 753–761.
- [82] U. Samanta, D. Pal, P. Chakrabarti, Packing of aromatic rings against tryptophan residues in proteins., *Acta Crystallogr. Sect. D-Biol. Crystallogr.* *55* (1999) 1421–1427.
- [83] R. M. Johnson, K. Hecht, C. M. Deber, Aromatic and cation- π interactions enhance helix-helix association in a membrane environment., *Biochem.* *46* (2007) 9208–9214.
- [84] E. Lanzarotti, R. R. Biekofsky, D. A. Estrin, M. A. Marti, A. G. Turjanski, Aromatic–Aromatic Interactions in Proteins: Beyond the Dimer., *J. Chem. Inf. Model.* *51* (2011) 1623–1633.
- [85] B. P. Dimitrijević, S. Z. Borozan, S. b. Stojanović, π – π and Cation– π Interactions in Protein–Porphyrin Complex Crystal Structures., *R. Soc. Chem. Adv.* *2* (2012) 12963–12972.
- [86] E. C. Lee, B. H. Hong, J. Y. Lee, J. C. Kim, D. Kim, Y. Kim, P. Tarakeshwar, K. S. Kim, Substituent Effects on the Edge-to-Face Aromatic Interactions., *J. Am. Chem. Soc.* *127* (2005) 4530–4537.
- [87] C. D. Sherrill, T. Takatani, E. G. Hohenstein, An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene- Methane, and Benzene- H₂S., *J. Phys. Chem. A* *113* (2009) 10146–10159.
- [88] S. E. Wheeler, Local Nature of Substituent Effects in Stacking Interactions., *J. Am. Chem. Soc.* *133* (2011) 10262–10274.

- [89] M. O. Sinnokrot, E. F. Valeev, C. D. Sherrill, Estimates of the Ab Initio Limit for π - π Interactions: The Benzene Dimer, *J. Am. Chem. Soc.* *124* (2002) 10887–10893.
- [90] L. M. Salonen, M. Ellermann, F. Diederich, Aromatic rings in chemical and biological recognition: energetics and structures., *Angew. Chem. Int. Ed. (English)* *50* (2011) 4808–4842.
- [91] J. C. Ma, D. A. Dougherty, The Cation- π Interaction, *Chem. Rev.* *97* (1997) 1303–1324.
- [92] J. P. Gallivan, D. A. Dougherty, Cation- π interactions in structural biology, *Proc. Natl. Acad. Sci.* *96* (1999) 9459–9464.
- [93] A. Seifert, S. Vomund, K. Grohmann, S. Kriening, V. B. Urlacher, S. Laschat, J. Pleiss, Rational Design of a Minimal and Highly Enriched CYP102A1 Mutant Library with Improved Regio-, Stereo- and Chemoselectivity., *ChemBioChem.* *10* (2009) 853–861.
- [94] J. Konc, T. Cesnik, J. T. Konc, M. Penca, D. Janezic, ProBiS-Database: Precalculated Binding Site Similarities and Local Pairwise Alignments of PDB Structures., *J. Chem. Inf. Model.* *52* (2012) 604–612.
- [95] J. Konc, D. Janezic, ProBiS: A Web Server for Detection of Structurally Similar Protein Binding Sites., *Nucleic Acids Res.* *38* (2010) W436–W440.
- [96] J. Konc, B. T. Miller, T. Stular, S. Lesnik, H. L. Woodcock, B. R. Brooks, D. Janezic, Probis-charmming: Web interface for prediction and optimization of ligands in protein binding sites, *J. Chem. Inf. Model.* *55* (2015) 2308–2314.
- [97] U. Bren, V. Martnek, J. Florián, Decomposition of the solvation free energies of deoxyribonucleoside triphosphates using the free energy perturbation method, *J. Phys. Chem. B* *110* (2006) 12782–12788.
- [98] M. Bren, J. Florián, J. Mavri, U. Bren, Do all pieces make a whole? Thiele cumulants and the free energy decomposition, *Theor. Chem. Acc.* *117* (2007) 535–540.

Table 1: $\Delta G_{\text{bind}}(\text{avg})$ in $\text{kcal}\cdot\text{mol}^{-1}$ and RMSFs in $\text{\AA}$ of SMIs in each SMI+P450 complex. The values reported are averages that have been calculated from the last 3 ns of the 4 ns SMI+P450 simulations.

	Ereb		XA	
	$\Delta G_{\text{bind}}(\text{avg})$	RMSF	$\Delta G_{\text{bind}}(\text{avg})$	RMSF
CYP315a1	-47.53	1.86	-22.83	1.63
CYP314a1	-36.43	1.65	-24.45	1.58

Table 2: List of per residue interaction energies between CYP315a1 residues and SMIs (SMIs = Ereb, XA). Interaction energies were calculated in CHARMM for the top poses. E_{Tot} represents the total interaction energy, while E_{vdW} and E_{elec} represent the van der Waals and electrostatic components. All energies are listed in $\text{kcal}\cdot\text{mol}^{-1}$. A “-” denotes pairs in which no appreciable interaction energy was calculated.

	Ereb			XA		
	E_{Tot}	E_{vdW}	E_{elec}	E_{Tot}	E_{vdW}	E_{elec}
Heme	-25.58	-17.09	-8.49	8.27	-15.66	23.94
Gly96	-1.59	-0.14	-1.45	—	—	—
Leu97	-2.17	-0.95	-1.23	—	—	—
Phe98	-3.35	-2.77	-0.59	-2.50	-1.53	-0.97
His78	-0.18	-0.08	-0.10	-0.07	-0.02	-0.05
Ile282	-1.18	-1.15	-0.03	—	—	—
Asp285	1.83	-2.58	4.41	-1.60	-0.80	-0.80
Leu286	-1.19	-1.63	0.44	—	—	—
Ala290	-1.64	-1.59	-0.05	—	—	—
Thr293	-3.22	-0.49	-2.73	-10.98	-1.04	-9.95
Ala341	—	—	—	-3.27	-0.33	-2.94
Phe343	—	—	—	2.47	-0.02	2.49
Arg346	—	—	—	-5.21	-0.03	-5.18
Cys414	-1.14	-0.57	-0.57	—	—	—
Met451	—	—	—	3.24	-0.04	3.27
Val452	—	—	—	-1.78	-0.08	-1.70

Table 3: List of per residue interaction energies between CYP314a1 residues and SMIs (SMIs = Ereb, XA). Interaction energies were calculated in CHARMM for the top poses. E_{Tot} represents the total interaction energy, while E_{vdW} and E_{elec} represent the van der Waals and electrostatic components. All energies are listed in kcal·mol⁻¹. A “-” denotes pairs in which no appreciable interaction energy was calculated.

	Ereb			XA		
	E_{Tot}	E_{vdW}	E_{elec}	E_{Tot}	E_{vdW}	E_{elec}
Heme	-21.99	-11.98	-10.02	-6.12	-10.85	4.74
Arg91	6.76	-1.59	8.34	-10.14	-1.81	-8.33
Pro92	-1.84	-0.07	-1.76	-	-	-
Val114	-0.09	-0.28	0.18	-	-	-
Asp220	-1.99	-0.12	-1.87	-	-	-
Ser221	-1.02	-0.04	-0.98	-	-	-
Phe300	-1.27	-0.06	-1.21	-	-	-
Thr307	-7.33	-1.44	-5.89	-	-	-
Thr365	-1.45	-0.69	-0.77	1.05	-0.44	1.49
Ala366	-6.92	-2.20	-4.72	-18.24	-18.24	-18.24
Phe367	-4.73	0.01	-4.75	-19.57	-1.33	-18.24
Cys368	-5.44	-1.78	-3.66	-6.20	-1.36	-4.84
Leu369	-	-	-	1.02	-0.29	1.31
Arg371	2.40	-0.06	2.46	-2.21	-0.02	-2.19
Cys392	-	-	-	2.24	-0.01	2.26
Gln393	-	-	-	-2.65	-0.02	-2.64
Asn394	-	-	-	-3.24	-0.04	-3.20
Cys443	-	-	-	-2.72	-0.31	-2.41
Phe478	-	-	-	-1.65	-0.11	-1.53
Leu480	-	-	-	4.12	-0.08	4.20
Ala481	-	-	-	-2.88	-0.18	-2.70

Table 4: Summary of average ring-ring distances measured between each pair of interacting aromatic species in the final 3ns of the 4ns simulations of Ereb and XA within the CYP315a1 active site. The two leftmost columns list the particular pair of interacting species, including members of the aromatic cluster and the SMIs (SMI = Ereb, XA). Columns 3 and 4 give the distances (in Å) between ring centroids for each aromatic pair. Column 3 gives those values as seen in the Ereb pose in CYP315a1 and column 4 for XA in CYP315a1. “SMI” should be substituted for either Ereb or XA, when reviewing data for columns 3 and 4 respectively.

		Ereb	XA
Heme	SMI	5.28	5.52
SMI	Phe98	5.66	6.82
SMI	His78	7.91	10.16
Phe98	Phe99	6.46	6.41
Phe99	Trp84	6.36	6.42
Trp84	Phe232	5.42	5.41