NMR characterization of the ligand dependent dynamic changes in Cytochrome P450s

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Rebecca A. Shrem
August 2020

Copyright © 2020 by Rebecca A. Shrem.
All rights reserved.

ACKNOWLEDGEMENTS

I acknowledge my advisor, Dr. Nitin Jain, for all of his help and guidance on this project. I also acknowledge Ana Bernal for her work that formed the preliminary basis for this project. I'd like to thank my cats, Rosalind Franklin and Bourbon, and significant other, Josh, for becoming excellent lab mates in this unprecedented time of remote work only due to COVID-19.

ABSTRACT

Cytochrome P450s are a large superfamily of enzymes in humans with little sequence homology, but a conserved fold. The flexibility of P450s, particularly of their substrate recognition site (SRS) is well-documented and enables them to bind a multitude of ligands with diverse physico-chemical properties, however the dynamic factors controlling the flexibility are not clearly understood. In order to gain better understanding of conformational flexibility inherent in P450s, NMR studies were performed on a model P450 system, CYP101A1, to characterize the dynamic differences exhibited by this P450 upon binding to a suite of 4 different ligands. Different ligand-bound forms of the protein were prepared, isotopically labeled and used to obtain sequence-specific resonance assignments throughout the protein using multidimensional NMR experiments. Molecular docking was carried out to elucidate a model structure of CYP101A1 in complex with one of the ligands, ketoconazole, as there is no crystal structure available for this complex. The new structural data and resonance assignments allowed interpretation of amide exchange rates encoding slow timescale motions for the various ligand-bound forms in comparison to the ligand-free form of CYP101A1, which showed significant differences in slow timescale dynamics for not just the SRS, but also regions outside of the SRS. The results from the amide exchange studies clearly show that CYP101A1 exhibits differential dynamics for binding various ligands even though they all have the same affinity to the enzyme. The dynamic measurements were complemented by thermodynamic data from ITC measurements which helped connect the dynamic differences in the various ligand-bound forms of the protein to the thermodynamic characteristics via a novel enthalpy-entropy transduction mechanism, which has not been experimentally demonstrated so far for a P450. Understanding of this mechanism will aid in gaining deeper insights into how P450s are able to modulate their dynamics for recognition of diverse ligands, with implications for drug design in human P450s.

CONTENTS

Chapter One: Cytochrome P450 Dynamics in Ligand Recongition.....1
1.1 Protein Dynamics.....1
1.2 Methods for characterizing protein dynamics.....3
1.3 Cytochrome P450 Enzymes 5
1.4 CYP101A1.....11
1.5 Project goals.....17
Chapter Two: Preparation of CYP101 for experiments.....21
2.1 Cell growth and CYP101 protein expression 21
2.2 Protein Purification.....22
2.3 Selection of ligands to probe CYP101 dynamics 23
2.4 Preparation of NMR samples.....27
2.5 Preparation of Selective Isotope labeled samples of CYP101.....27
2.6 Two- and Three-dimensional NMR experiments for resonance assignments.....28
2.7 Collection of Amide exchange data 29
Chapter Three: Sequence-Specific NMR Resonace assignments from Mulitdimensional Heteronuclear experiments 31
3.1 Three-dimensional Heteronuclear NMR experiments for resonance assignments in CYP101 31
3.2 Sequence-specific Assignments of various forms of CYP101 34
3.3 NMR spectra of ^{15}N selectively labeled CYP101 samples.....39
Chapter Four: Modeling of CYP101-Ketoconazole Complex Structure Via Molecular Docking 42
4.1 Docking Protocol for the Ketoconazole-CYP101 complex.....44
Chapter Five: Hydrogen-Deuterium Exchange experiments to monitor ligand-dependent dynamic changes in cyp101.....53
5.1 Ligand-dependent dynamic changes in CYP101 from preliminary 2D NMR spectra.....53
5.2 Measurement of Amide Exchange Rates for Various Forms of CYP101 56
5.3 Ligand-dependent dynamic changes in the substrate recognition regions of the protein structure.....63
5.4 Amide Exchange results show ligand-dependent dynamic changes in the regions not associated with ligand binding 68
Chapter six: Thermodynamics of cyp101-ligand interactions.....72
6.1 Isothermal Calorimetry of CYP101-ligand complexes.....72
6.2 Thermodynamic parameters of CYP101-ligand complexes.....74
Chapter Seven:.....83
Conclusions and Future Directions 83
7.1 Summary of Project and Conclusions.....83
7.2 Future Directions 85

Appendix 85
List of References 111
Vita 117

LIST OF TABLES

Table 1: List of ligands used to study dynamics of CYP101……..…….26

Table 2: Summary of 3D NMR experiments on various forms of CYP101 and details of resonance assignments available……...…...38

Table 3: HDX exchange categorization table……60

Table 4: Overall amide exchange type percentages for assigned peaks in various CYP101 forms …….…….63

Table 5: Summary of ITC experiments ……..……...73

Table A.1: NMR Chemical shift Assignments for oxidized Cyp101 in complex with the substrates camphor and norcamphor……...……...87

Table A.2: NMR Chemical shift Assignments for oxidized Cyp101 in ligand-free form ……..92

Table A.3: NMR Chemical shift Assignments for oxidized Cyp101 in complex with the nicotine and ketoconazole ……96

Table A.4: Example plots of calculation of exchange rates in various CYP101 forms and table of amide exchange rates for camphor and norcamphor ……..100

Table A.5: Table of amide exchange rates for Nicotine, Ligand free and Ketoconazole……108

LIST OF FIGURES

Figure 1: Timescales for protein motions and NMR experiments that examine them [4] 2
Figure 2: Conserved P450 reaction mechanism [1] 7
Figure 3: Comparison of bacterial (CYP101A1) and human P450 (CYP2C9) structures showing common fold and secondary structure elements. CYP101A1 PDB id: 2CPP and human CYP2C9 PDB id: 1OG2 10
Figure 4: Active site of CYP101 bound to substrate camphor (top) and inhibitor nicotine (bottom) with important residues marked in white 12
Figure 5: CYP101 structure with regions corresponding to the substrate recognition sites (SRS) shown in red. CYP101 PDB id: 2CPP 14
Figure 6: Superimposed structures of camphor-bound (blue) and the ligand-free (red) CYP101, showing the displacement (yellow double-headed arrow) of the FG Loop upon binding of camphor. CYP101 camphor-bound PDB id: 2CPP; ligand-free PDB id: 3L62 15
Figure 7: Solution NMR 1H-15N 2D HSQC-TROSY correlation spectra for oxidized CYP101 in ligand-free form and bound to various ligands 19
Figure 8: Zoomed in Portions of 2D HSQC-TROSY spectra for oxidized CYP101 in ligand-free and various ligand-bound forms showing differences in linewidth changes and peak splittings among the various spectra 20
Figure 9: : UV-Vis spectra of CYP101 bound to each ligand 25
Figure 10: 3D NMR heteronuclear resonance assignment experimental scheme. Red circles indicate the chemical shifts of nuclei that will be visible in a certain experiment. Note that the N and H chemical shift in the HN(CO)CA experiment is correlating to the self. 33
Figure 11: Example of sequential assignment of a stretch of residues 387-390 from ketoconazole-bound CYP101 using a 3D HNCA spectrum 36
Figure 12: Comparison of 2D ^{1}H-^{15}N HSQC-TROSY of camphor-bound CYP101 uniformly labeled (left) and selectively labeled with ^{15}N Lysine (right). Only the ^{15}N lysine peaks are seen, identifying some of the peaks in the uniformly labeled spectrum by amino acid type 40
Figure 13: Ketoconazole initial structure generated from CHEMDRAW 45
Figure 14: Flow chart showing the docking protocol for CYP101-ketoconazole complex 47
Figure 15: Docked structural model of ketoconazole-bound CYP101 generated with HADDOCK 49
Figure 16: Structural view of docking site of ketoconazole in the SRS of CYP101, showing the interactions of chlorinated phenyl ring and long chain with specific residues in CYP101 50
Figure 17: Comparison of NMR chemical shift perturbations between ligand-free and ketoconazole-bound CYP101 (left) and RMSD between ligand-free and docked ketoconazole-bound CYP101 structure (right). Color code for Chemical shift perturbation and RMSD difference: Small -Pink, Medium-Blue, Large-Cyan. 52

Figure 18: Zoomed in portions of 2D ^{1}H-^{15}N HSQC-TROSY of CYP101 in ligand-free and various ligand-bound forms, showing the different peak patterns in terms of chemical shift changes, line-broadening and peak splittings. The amino acid G189 in CYP101 is used as an example to illustrate these variations 54
Figure 19: Scheme for Hydrogen-Deuterium Exchange experiment [65] 57
Figure 20: 2D ^{1}H-^{15}N HSQC spectra of CYP101 bound to various ligands showing amide exchange at different timepoints starting with no D_2O added at 0 hrs to 40 hours after D_2O exposure. 59
Figure 21: Structures of showing the exchange patterns for each ligand highlighting the SRS 65
Figure 22: Comparison of the overall amide exchange patterns for the various CYP101 forms 69
Figure 23: ITC curves and fit for binding of nicotine to CYP101 79
Figure 24: ITC curves and fit for binding of ketoconazole to CYP101 80
Figure 25: ITC curves and fit for binding of camphor to CYP101 81
Figure 26: ITC curves and fit for binding of norcamphor to CYP101 82

CHAPTER ONE:
CYTOCHROME P450 DYNAMICS IN LIGAND RECONGITION

1.1 Protein Dynamics

Proteins are not static but dynamic entities, where dynamics of the protein apart from structure often play a critical role in executing said functions [2]. While traditionally the study of enzymes has emphasized their structural elements in relation to their function, more evidence points towards the dynamics of the protein playing an integral role. Protein function often requires the proteins to maintain a certain level of flexibility and dynamic movement. Protein dynamics can be defined as a reversible time dependent movement from atom coordinates [3]. The dynamics are not limited to a single time scale, and can occur on several time scales depending on the structural level (Figure 1) [4]. Slow timescale movements (s-min) include movements such as the global collective motions of the protein. The intermediate timescale (μs-ms) encompasses movements like folding and domain movements. The intermediate time scale also includes motions associated with both ligand binding and catalysis. The fast time scale (ps-ns) encompasses movements like side chain rotation and loop movement [5].

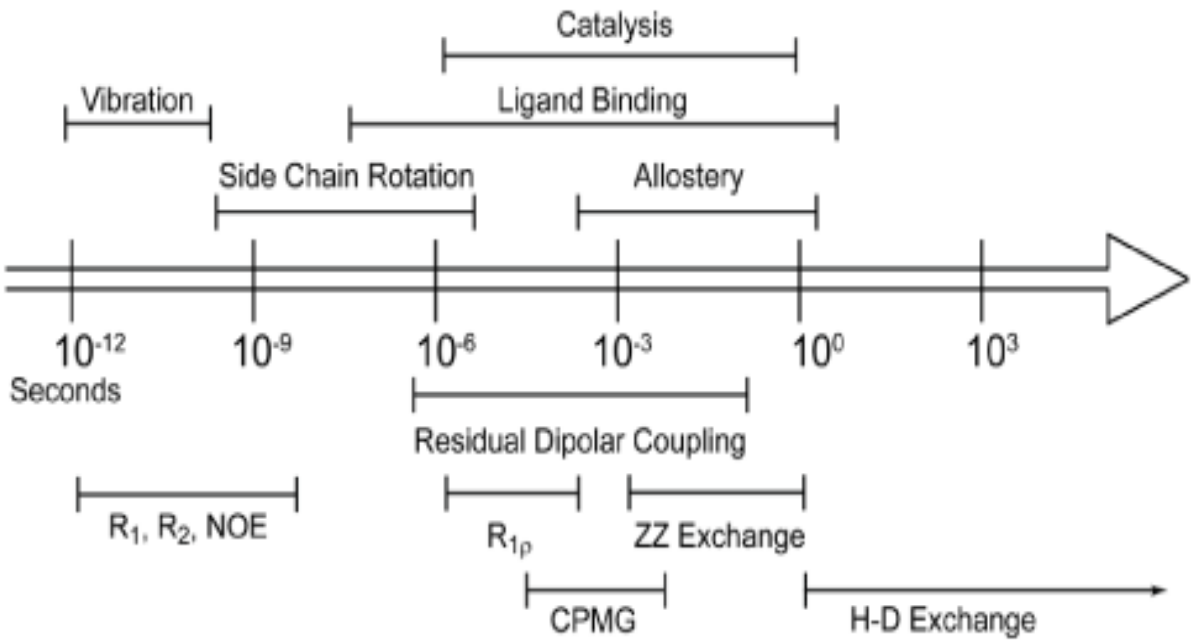


Figure 1: Timescales for protein motions and NMR experiments that examine them [4]

1.2 Methods for characterizing protein dynamics

Different techniques are specialized to examine the different time scales of motion. For example, stopped flow methods can study fast picosecond-nanosecond timescale motions, whereas Forster resonance energy transfer (FRET) is a technique that can be utilized to study the intermediate time scale motions in the nanosecond to microsecond range. FRET is based on adding fluorophore tags that absorb different wavelengths of lights to the two areas or molecules that potentially interact. If the tags encounter one another, the florescent signal will change due to resonance energy transfer, which can be tracked and used to get both structural and dynamic information [6]. One large disadvantage of this technique is that the tag introduces an artificial element that can affect the protein's structure and ability to fold. It also does not have resolution at the atomic level. Single-molecule FRET is also an increasingly popular technique that is being utilized increasingly to study conformational dynamics of a protein. This technique has the advantage that it is not an averaged ensemble of structures and thus can show details at a single-molecule level that are obscured by the ensemble, like kinetic information when a system is at equilibrium [7, 8]. However, such details cannot be obtained experimentally at atomic resolution and have to be combined with MD simulations to interpret results.

X-ray crystallography is another alternative than can be used to look at fast time scale motions. Interpretation of the electron density using time-resolved crystallography can show dynamics at the atomic level in the ps- μs range [9]. One group used picosecond time-resolved Laue diffraction methods to compare wild type and a mutant myoglobin's motions in the time range of 100 ps to 3 μs [10]. However, there are some significant limitations to this technique. First, this method normally requires large quantities of protein, especially in the crystallization setup, crystal growth and screening process. Second, crystals will only form under certain conditions, which may not be representative of the conditions that the protein will face in actuality in solution. Thus, only a partial representation of conformations available to the protein can be obtained.

Molecular dynamics is an *in-silico* method that is heavily utilized to look at fast time scale motions and is often used as a complement to experimental data. Molecular dynamics is able to compute probable protein motions with great accuracy for motions

in the fs-ps time scale [11]. This technique can provide atomic resolution when paired with complementary experimental techniques such as neutron scattering or crystallography. This can act as limitation on molecular dynamics, as it does require robust experimental validation.

Spectroscopic techniques overcome some of the limitations in resolution, protein amounts, non-native conditions and modification to proteins, as dynamic motions can largely be studied experimentally using methods such as electron paramagnetic resonance (EPR) and nuclear magnetic resonance (NMR) spectroscopy under physiological conditions without any tagging or outside perturbations, giving a less biased picture of the protein motions [12]. EPR can be used to study motions in the ns-μs time scale range and monitors changes in linewidths to interpret dynamic motions [13]. Solution NMR spectroscopy in particular provides information at the atomic level on all three timescales of motion and is one of the primary techniques used to study protein dynamics [14]. Solution NMR allows for multiple conformations of the protein to be represented at once which can be detected at high resolution and even time-dependent transitions between the conformations can be characterized in great detail. Another advantage of NMR in study of dynamics is that it typically uses lower concentrations of protein compared to similar methods.

A variety of solution NMR methods are available to characterize molecular motion in solution. The most prominent among them are hydrogen-deuterium exchange (HDX) experiments that monitor ms-min slow timescale motions on the protein backbone by NMR. HDX as a technique is not limited to NMR and can also be used alternatively in mass spectrometry or in conjunction with NMR, however NMR typically provides higher atomic resolution in a site-specific manner[15]. This method involves monitoring decreasing peak intensity in a two-dimensional ^{1}H-^{15}N correlation spectrum over time after exposure to D_2O due to the so called overall “breathing” motions of the protein which opens up and closes water access channels to the interior of the protein resulting in exchange of hydrogens attached to nitrogen with deuterium as a function of this motion.

The intermediate time scales of motion on the other hand can be captured using the transverse relaxation rates for protein backbone N-H and side-chain C-H bonds

utilizing the Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence. This technique works by measuring T_2 relaxation time modulation under chemical shift refocusing conditions or lineshape analysis to get rates for chemical exchange in the μs-ms time frame [16]. Such measurements are highly useful to characterize important functional motions of proteins involved in catalytic activity, recognition events such as protein-protein and protein-ligand interactions as well as folding-unfolding events.

Finally, backbone amide relaxation experiments can be performed to monitor the fast ps-ns timescale of protein motion. These allow two relaxation time parameters to be measured which can be used indirectly to quantify protein motion: The T_1 time, which is the spin-lattice relaxation time, and T_2, which is the spin-spin transverse relaxation time[17]. These nuclear spin rates can be monitored in a residue-specific manner within the protein to get a sense of motions such as isotropic and anisotropic backbone movements by calculating order parameters along with any relevant chemical exchange rates for transitions between protein conformations as part of these movements.

1.3 Cytochrome P450 Enzymes

Enzymes are proteins that perform a specialized function, namely catalysis of a specific reaction, converting substrates into products. This function typically involves movement of structural elements, mainly in the catalytic site, as part of dynamic motions of the protein. Cytochrome P450s are a super family of enzymes that are found in every kingdom of life, with over a million estimated unique P450s in existence[18]. P450s are monooxygenases, which add an oxygen to a carbon hydrogen bond. While the substrates of P450s may vary, all P450s share a conserved reaction mechanism [19]. The typical general reaction catalyzed by P450s is represented in equation (1) and is performed to make a more soluble product compared to the substrate. The solubility of the products helps with reducing the toxicity of the substrates by safely discarding them through aqueous means and can act as a defense mechanism in metabolism of xenobiotic compounds.

$$\mathrm{RH} + \mathrm{O_2} + 2\,\mathrm{e^-} + 2\,\mathrm{H^+} \rightarrow \mathrm{ROH} + \mathrm{H_2O} \qquad \textbf{\textit{Eqn 1.}}$$

The catalytic mechanism starts when a substrate binds to the enzyme (Figure 2) [1]. Once the substrate is bound, the iron in the heme is reduced from its +3 oxidation state to +2 oxidation state by an electron donor, like NADPH or ferredoxin, to move forward in the catalytic cycle [20]. Oxygen binds in the next step in the mechanism and gets converted to radical form by accepting an electron from the iron in the heme. Only after the completion of this step will a second electron be accepted, making an iron-oxene. Two protons then enter the system and form one molecule of water with one of the oxygen atoms, leaving behind an oxygen-iron complex. This causes rearrangement at the heme center, which allows for the hydroxylation of the substrate to product by using the remaining oxygen atom and will end the cycle when the product exits the active site.

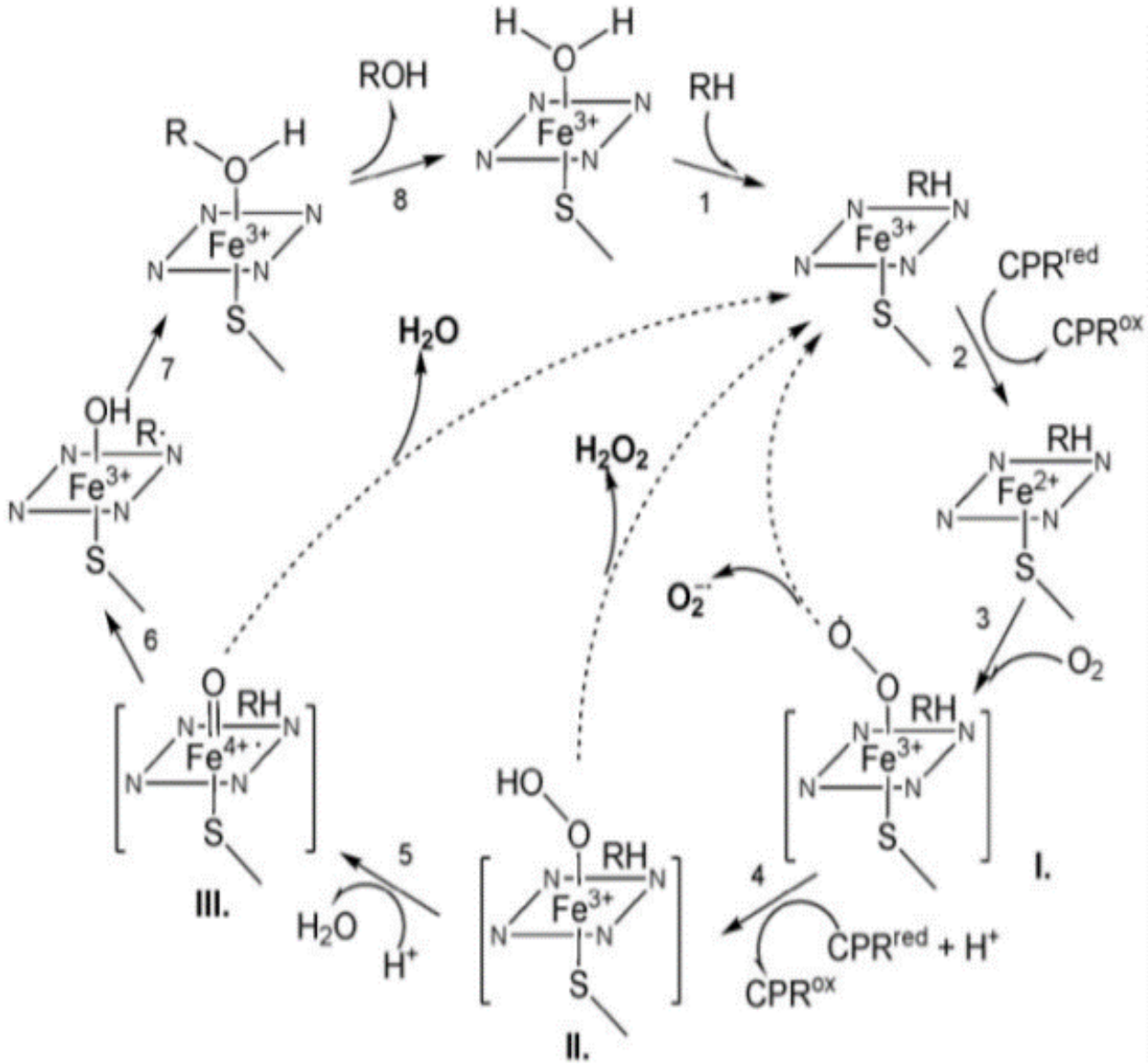


Figure 2: Conserved P450 reaction mechanism [1]

While cytochrome P450s are present in most organisms, the primary use of them can vary greatly. In plants, a typical function is to use P450s to generate secondary metabolites to defend against stressors, such as pesticides or bacterial infections [21]. Plant P450s have also been associated with the ability of microbes to colonize plant roots [22]. Bacteria utilize P450s as part of their metabolic processes or in sterol biosynthesis[23]. In humans, P450s are used primarily for xenobiotic processing, fatty acid metabolism, and hormone synthesis [24]. Another area where P450s have been utilized is biocatalysis, as the chemical reactions they help carry out is very costly without them and requires very high heat to facilitate when the enzymes are not used [25].

The ability to activate carbon-hydrogen bonds gives these enzymes an important role in the metabolism of xenobiotics as well as in endogenous synthesis of ligands. There are 57 human P450s that are primarily expressed in the liver and constitute 80% of drug metabolism activity in humans [26], [27]. CYP3A4 is the most abundant of human liver P450s and processes more than 50% of the oxidatively metabolized drugs [28]. However, the metabolic activity of CYP3A4 can vary greatly in individuals due to presence of genetic polymorphisms. There are more than 78 known single nucleotide polymorphisms (SNPs) for this protein, and certain isoforms have been associated with an increased cancer risk [29]. This SNP problem is not just restricted to CYP3A4. Another human P450 that exhibits a large number of SNPs is CYP2C9. CYP2C9 comprises approximately 20% of the total liver P450 content, making it one of the major human metabolic P450s [30]. There are 30 different SNPs currently identified for the CYP2C9 gene, but there are only 5 isoforms expressed in significant amounts [31]. These isoforms vary in their ability to process drugs, leading to differential drug efficacy and toxicity. Individuals can be categorized into three phenotypic categories; extensive drug metabolizers, intermediate drug metabolizers, and poor drug metabolizers, depending on which copies of CYP2C9 are present in their genome as well as which drug is being metabolized [32]. Notably, all of the mutations for the five major polymorphisms are outside of the active site or do not directly participate in the catalytic activity for CYP2C9. This indicates an allosteric effect on either substrate binding or catalysis, via regions previously not associated with substrate binding and catalysis.

Because of their significant role in drug metabolism, P450s are a major drug target and are of much interest to the pharmaceutical industry [24]. However, the presence of SNPs complicates the drug development process due to individual metabolic variations. Between these various roles and research into effects of SNPs, P450s make up a multibillion-dollar industry.

Despite little sequence homology and only a few conserved catalytic residues, all P450s still share a common fold and heme prosthetic group. For example, despite there only being approximately 15% sequence homology between CYP2C9 and the bacterial P450 CYP101A1, the fold is fairly similar and the same general structure is maintained (Figure 3). This conserved fold can be seen in the major secondary structure elements in and around the active site, such as the helix that anchors the heme group or the number of key structural elements involved in substrate binding. The family loses it similarities however when looking at the type of interactions in the regions associated with substrate binding, or the specific substrate recognition site (SRS). The secondary structure elements that make up this region can be variable between how they adapt structurally and dynamically to different substrates in different P450s, which lends to the ability of P450s to bind to a wide array of diverse substrates. For the P450s to be able to bind substrates with different physical and chemical properties, they must be able to readjust the secondary structures in the SRS to match the features of that particular substrate. These rearrangements must be dynamic to facilitate binding, making the dynamics and substrate binding inextricable. The proteins must also maintain a certain level of flexibility not just in the SRS, but throughout the protein to make this possible. Studies have shown that the loop that closes the substrate channel, known as the F-G loop, which is common to all P450s, is very flexible [33] and works in tandem with other flexible elements of the SRS to accomplish ligand binding. In this research project, we have examined this flexibility by studying differential substrate binding by CYP101A1, a model system for the P450 family in terms of structure and dynamics.

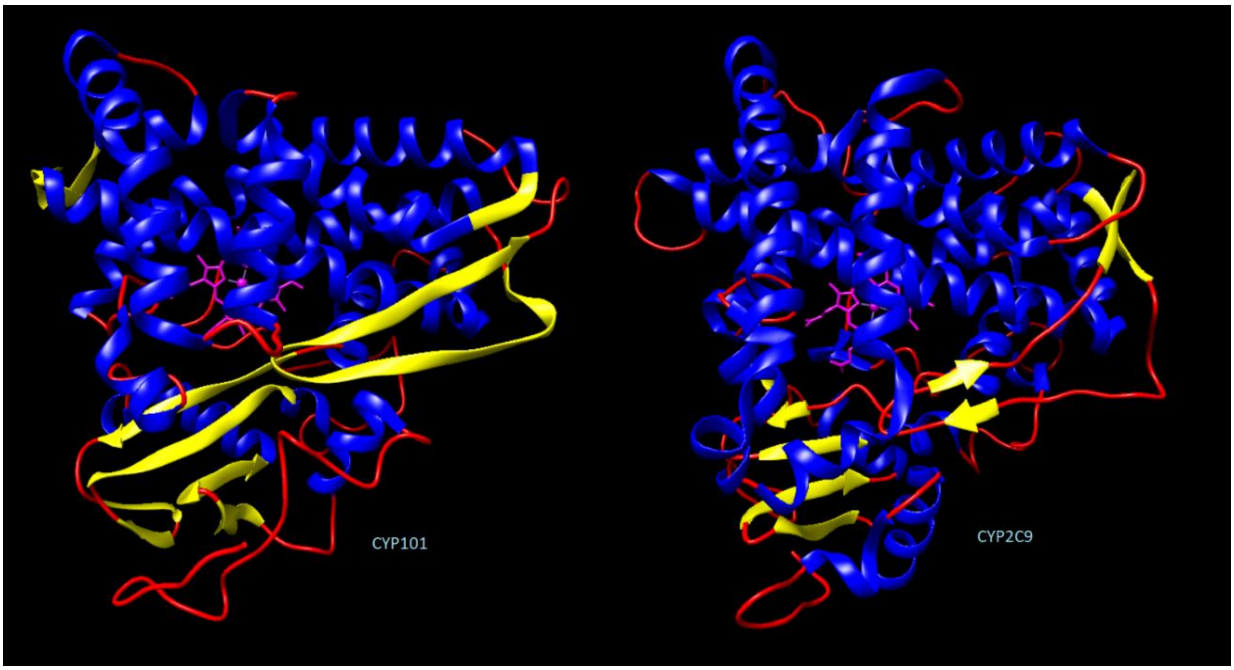


Figure 3: Comparison of bacterial (CYP101A1) and human P450 (CYP2C9) structures showing common fold and secondary structure elements. CYP101A1 PDB id: 2CPP and human CYP2C9 PDB id: 1OG2

1.4 CYP101A1

CYP101A1 ($P450_{cam}$ or CYP101) isolated from the bacteria *Pseudomonas putida* was the first P450 in the entire cytochrome P450 superfamily to be purified, crystallized and have its structure solved [34]. CYP101 inserts an oxygen atom into a C-H bond of its natural substrate, camphor, to synthesize 5-exo-hydroxycamphor which can then be metabolized further by the bacteria to generate energy. CYP101 follows the same catalytic reaction scheme as other P450s shown in Figure 1. CYP101 is reduced by its redox partner, the ferredoxin protein putidaredoxin (Pdx), in two separate steps to complete the reaction. Pdx is the only electron donor that CYP101 can use to complete the mechanism, and mechanistic studies have found that, besides the contribution of electrons, there is also a specific effector role for Pdx in the mechanism [35].

The active site of the P450 family is well conserved, and CYP101 is no exception (Figure 4). It has multiple conserved residues found on the I helix and β-3 loop of the protein, which is buried inside and away from the bulk solvent. These residues include a cystine (C357) coordinates with the heme through its sulfur atom, anchoring it to the L helix, and a threonine-asparagine pair (D251-T252) that act as part of the proton/electron transfer pathway for protonation of oxygen to water. Mutations of T252 results in the uncoupling of proton transfer and the production of peroxide instead, and mutations to D251 result in reduced electron transfer. D251 has also been implicated in substrate access to the active site [36]. Tyrosine 96 on the B' helix plays a role in camphor's orientation in the active site by hydrogen bonding to the oxygen atom to keep it in place. All substrates, not just camphor, upon binding displace a water molecule that coordinates with the 6th coordination site on the heme, which frees that site up for oxygen binding. In the case of inhibitors, the inhibitor will bind directly to the heme and occupy the same 6th coordination site, preventing oxygen's attachment. Once the reaction has been completed, the product then exits the active site by a hydrophilic channel near the BC loop and β-1 Sheet.

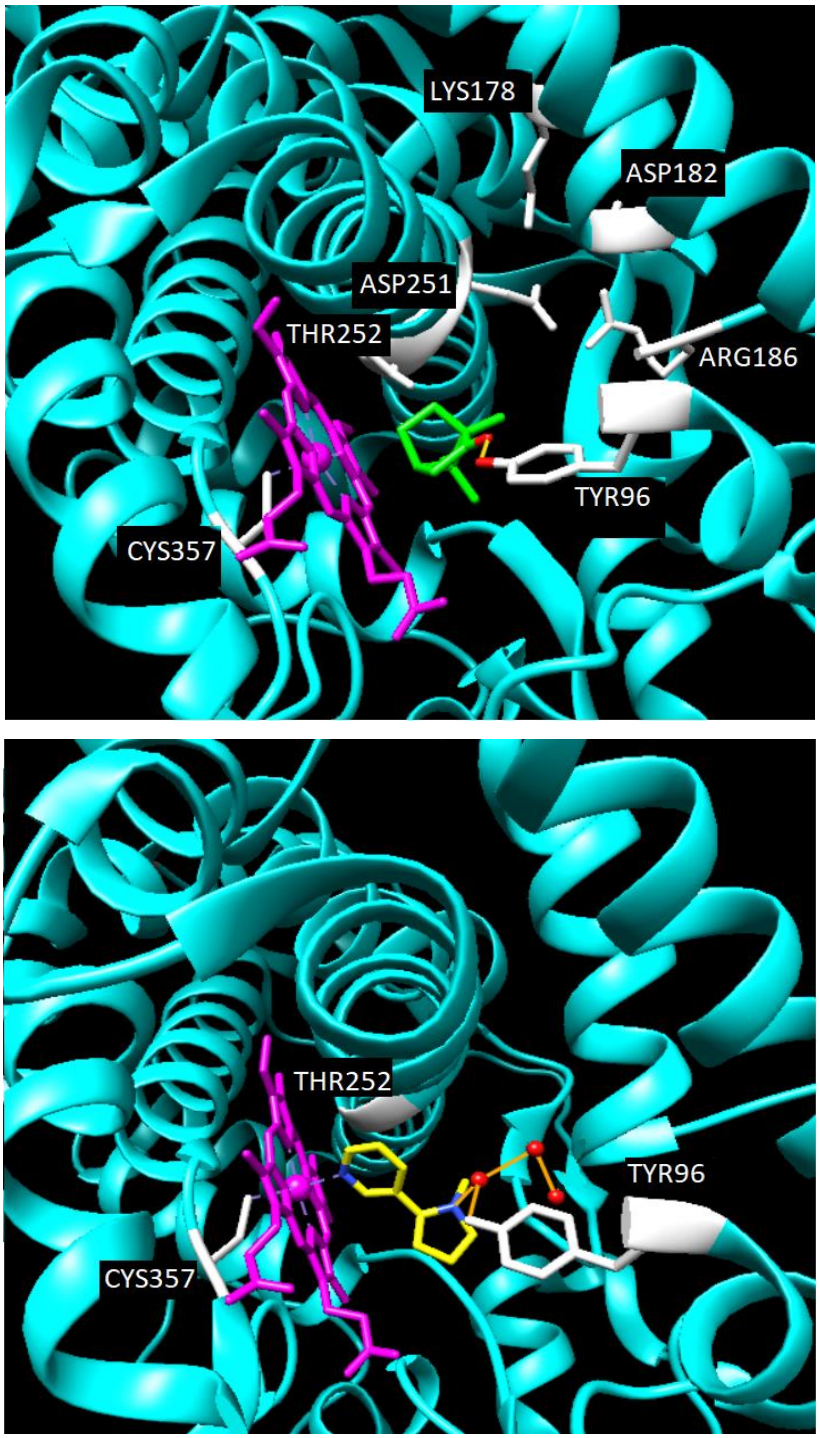


Figure 4: Active site of CYP101 bound to substrate camphor (top) and inhibitor nicotine (bottom) with important residues marked in white

The substrate binding and the catalytic conversion of substrate to product are two energetically independent events for P450s, with the substrate entering the enzyme through a channel that is distinct from the active site [37]. This channel whose opening is formed by the SRS for CYP101, consists of the F and G helices, connected by the FG loop, the β-5 sheet, a portion of the N terminal region and the B' helix (Figure 5). These regions are very flexible and show the least conservation among all P450s [38]. Important residues for these regions include Y29, F87, Y96 and F193, which show displaced side chains when bound to small or large ligands [39]. The channel can exist in an open state when no ligand is present and will convert to a closed state by moving the F and G helixes towards the heme via the FG loop after binding most ligands [37], making the distance the FG loop travels approximately 10 Angstroms between states (Figure 6).

While the flexibility of the SRS is well-known, the dynamic motions that allow this flexibility is still not completely understood. Ligand-free CYP101 is believed to sample dynamically between open, partially open and closed states [40].There have only been a few experimental studies where the dynamics of P450s have been investigated in atomic detail. Most of these experimental studies have focused on CYP101 and other bacterial P450s. One study used 2D infrared vibrational echo spectroscopy to examine the dynamics of ligand-free, camphor-, and norcamphor-bound CYP101. It was observed that the dynamics of residues in CYP101 active site was correlated with the affinity and hydroxylation of the substrates on the ps timescale. Camphor-bound CYP101 was found to be less dynamic than the norcamphor-bound form, which is in turn less dynamic than the ligand-free protein.[41] Dynamic studies on human P450s have been carried out mostly by MD simulations on short to long timescales [42, 43]. Focusing specifically at the SRS of CYP101, MD simulations verified with NMR data have also been utilized to understand the cis-trans isomerization of the ILE88-PRO89 at the start of the B' Helix. This study surmised that the switch between isomers happens on the fast timescale and is induced by the binding of Pdx, and leads to reorientation of camphor within the active site [44]. A different study incorporating amide exchange on

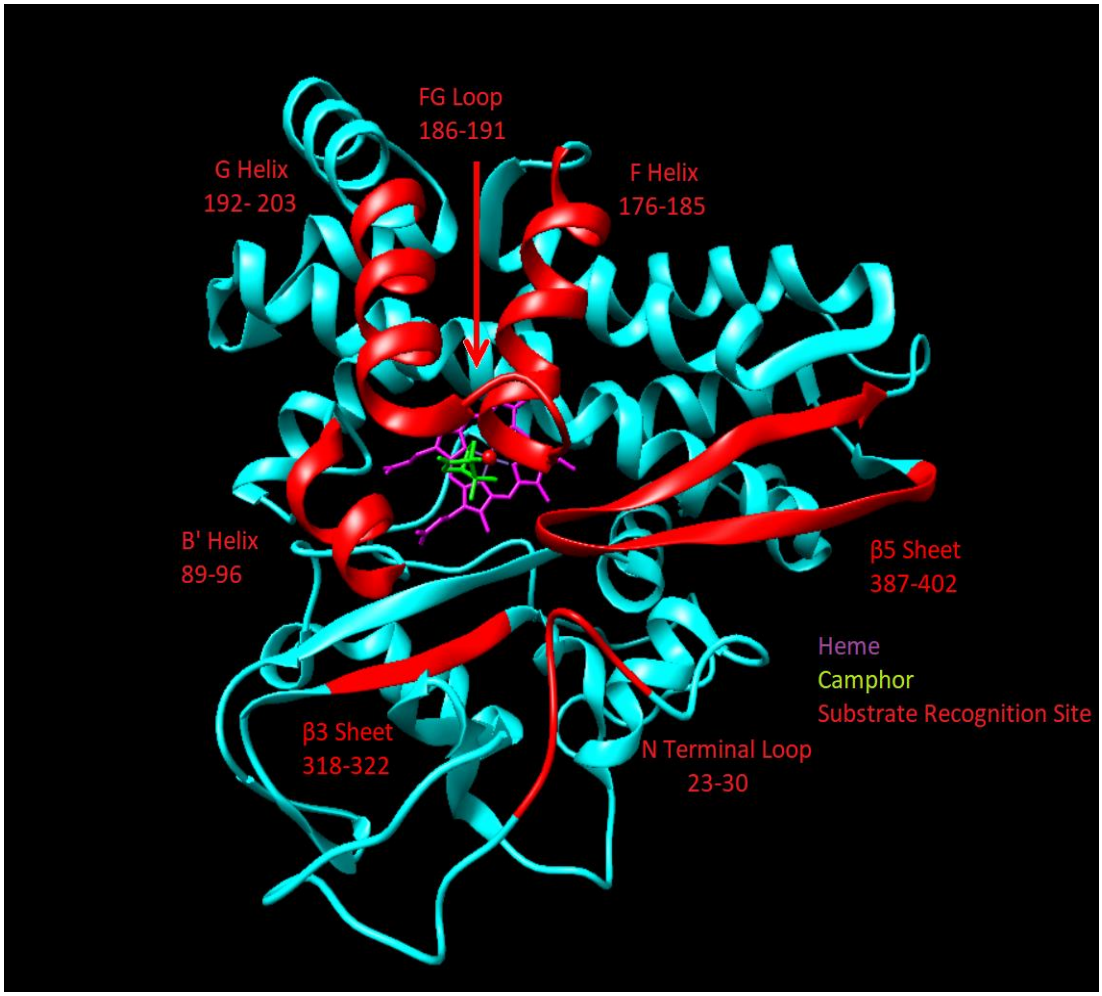


Figure 5: CYP101 structure with regions corresponding to the substrate recognition sites (SRS) shown in red. CYP101 PDB id: 2CPP

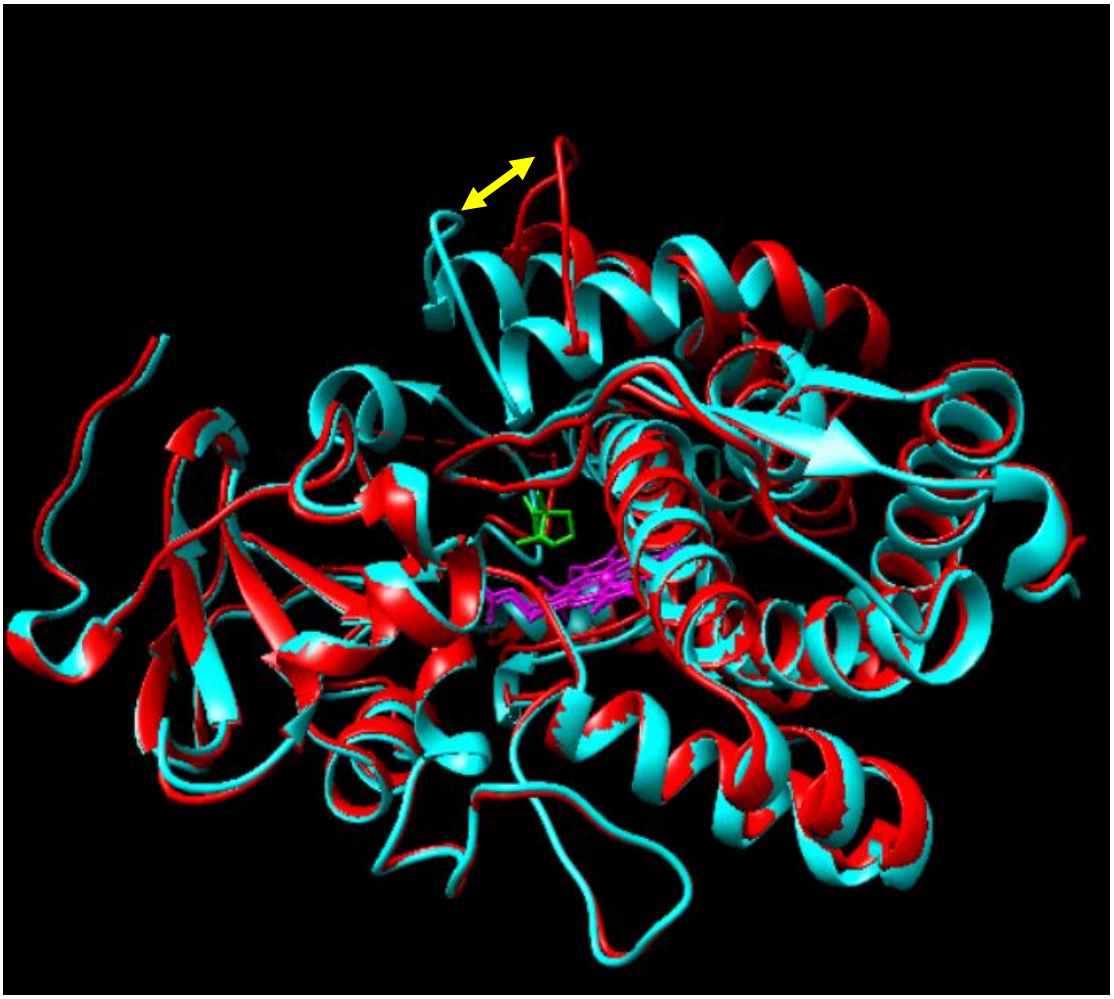


Figure 6: Superimposed structures of camphor-bound (blue) and the ligand-free (red) CYP101, showing the displacement (yellow double-headed arrow) of the FG Loop upon binding of camphor. CYP101 camphor-bound PDB id: 2CPP; ligand-free PDB id: 3L62

CYP101 was carried out using NMR and mass spectrometry which found that the B' helix exchanged out much quicker in the oxidized form of the protein compared to the reduced form [45]. A prior study utilizing NMR spectroscopy, neutron scattering and MD simulations compared the ligand-free form with the camphor-bound form in terms of backbone relaxation dynamics and found that the flexibility between the SRS and regions outside of the SRS in the protein is coupled via a network of interactions which helps modulate ligand binding [46]. The solution NMR ^{15}N relaxation measurements on CYP101 in this study characterized fast backbone N-H motions with concomitant calculation of order parameters for camphor-bound and ligand-free CYP101, both experimentally and from hydrated powder MD simulations, which concluded that the SRS exhibited the largest dynamic changes in the ps-ns time range [46]. Regions outside of the SRS, specifically in the corner regions of the triangular structure of CYP101, also had small but significant changes in dynamics between the two forms. A more recent study employing inelastic neutron scattering of hydrated powders showed presence of fast motions on the ns timescales that are collective in nature. Normal mode analysis of the network of these motions shows that the protein uses these collective motions to move along a perpendicular axis from the ligand binding channel to open and close the channel [47]. Both the NMR and neutron scattering studies were only conducted with the ligand free or camphor-bound protein, and do not necessarily represent the whole range of dynamics and motions that CYP101 can undergo with ligands with differing physico-chemical properties. Such a dynamic characterization with multiple ligands is important to explain the wide range of ligand specificity observed for P450s in various species and organisms. This dynamic understanding can also be exploited in design of new drugs that are readily recognized by the flexible SRS and appropriately metabolized by human P450s in the active site alleviating some of the clinical issues apparent in case of hyper or poor drug-metabolizing individuals due to various polymorphisms.

1.5 Project goals

The overall goal of this project is detailed characterization of P450 protein dynamics in context of multiple ligand recognition in a model P450 system providing valuable information that can be applied to other members of the P450 family to understand their diversity in ligand binding. CYP101 is a good prototypical system for such characterization for multiple reasons. For one, CYP101 can be expressed and purified as a soluble protein with soluble cofactors in large quantities. This makes it advantageous to use CYP101 rather than membrane-bound mammalian P450s which are notoriously difficult to express and purify in both soluble and membrane-bound forms at high concentrations. Second, the soluble form of CYP101 shows structural, dynamic and mechanistic aspects that are comparable to a large extent to soluble mammalian P450s with a tendency to bind similar ligands and exhibit large conformational changes similar to those observed in mammalian P450s [48]. Finally, native CYP101 was the first P450 crystallized in the late 1970s and hence is one of the most studied P450s, meaning there is a large foundation of biochemical and biophysical information available on it [34]. Since then, over 70 different structures have been solved for CYP101, in various mutant forms and bound to several ligands. Despite camphor being the natural substrate, CYP101 shows a high affinity for many other ligands, both substrates and inhibitors similar to mammalian P450s. This allows for the structural and dynamic examination of binding of the same or similar ligands that bind to human P450s, with the assumption that the dynamic mechanisms to effect binding of these ligands across multiple P450s are generally similar given the conservation of the SRS and the specific elements in these regions, although the catalytic specificity may vary. This has given impetus to use of CYP101 as a model system to understand the dynamic mechanisms that might operate in other P450s, especially human P450s. While X-ray crystallography can yield great structural insights on how a P450 may interact with a ligand in terms of static snapshots of ligand-bound and ligand-free structures with the accompanying structural changes, it does not provide a time-dependent dynamic picture in context of ligand binding. For example, the crystal structures of CYP101 bound to most ligands are superimposable with minor changes due to freezing of protein conformations in forms that are easily crystallizable but may

not represent the most likely or diverse conformations observed in solution. This therefore does not reflect the entire range of ligand-dependent dynamic motion available to the protein that is observed for these same complexes in solution as observed in solution NMR measurements (Figure 7/8). Looking at the spectra with different ligands, clear differences can been in the differential linewidths and peak splittings when CYP101 is bound to different ligands. NMR data clearly indicates that CYP101 is sampling multiple conformations in solution in a time-dependent manner that is not reflected in crystal structures. It is likely that these dynamic motions can be used to effect ligand binding by the SRS in a specific manner as has been suggested by preliminary evidence from previous NMR and neutron scattering studies, however what is not clear from these studies is whether the protein dynamics change upon binding of a different ligand and if so, how and to what extent?

We aim to answer these questions in the current study by investigating the dynamic changes in CYP101 bound to various ligands using solution NMR spectroscopy. Previous dynamic work on CYP101 has been primarily carried out by comparing the dynamics of ligand-free oxidized CYP101 to only a single ligand-bound form i.e. camphor-bound CYP101. The current study aims to extend the dynamic investigation to a multitude of ligands and that too with differing physico-chemical properties in order to get an elaborate sense of the range of dynamics exhibited by the protein. This is the first such dynamic characterization for CYP101 and for that matter for any P450 in the superfamily and has potential to offer detailed insight into ligand binding dynamics at a level that goes beyond the current structural knowledge for this important family of enzymes. Since P450s are such an important drug target, this information is relevant to rational drug design, particularly in guiding flexible docking and screening of drugs using dynamic structural ensembles. NMR spectroscopy is a great technique for such dynamic characterization due to its ability to study these dynamic changes in solution on different timescales under physiological conditions. The main goal of the current study is therefore to look specifically at oxidized CYP101 dynamics on the slow timescales using NMR spectroscopy to try and get a comprehensive picture of the dynamic changes occurring upon binding of 4 different ligands that differ in their size and chemical properties.

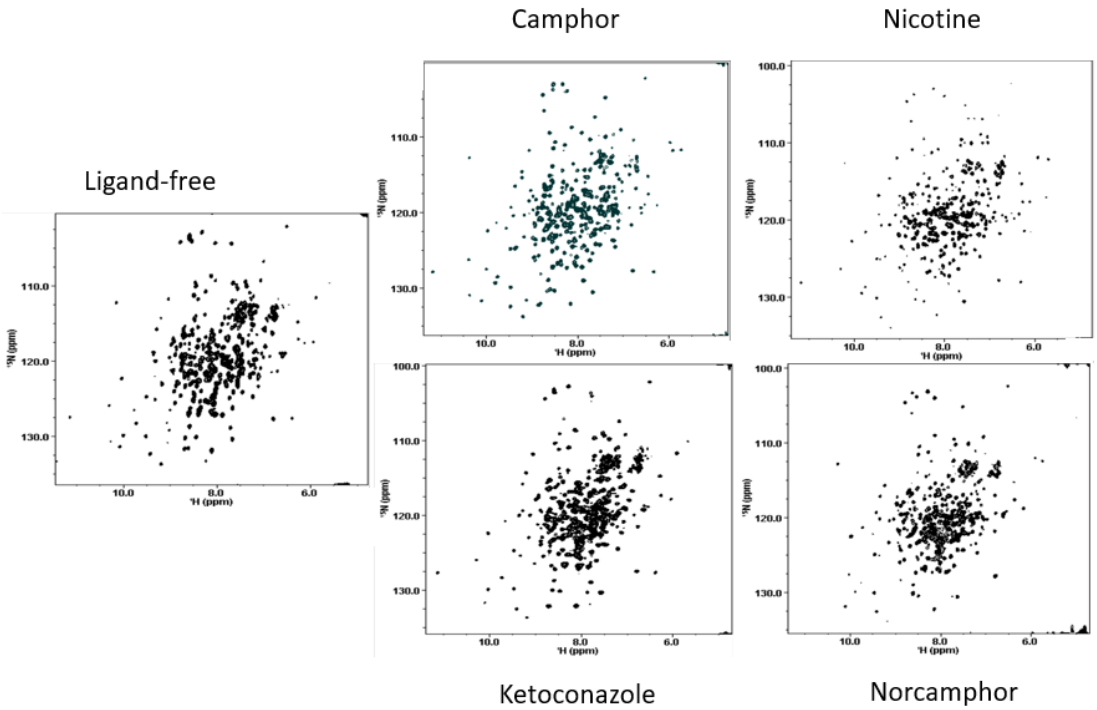


Figure 7: Solution NMR 1H-15N 2D HSQC-TROSY correlation spectra for oxidized CYP101 in ligand-free form and bound to various ligands

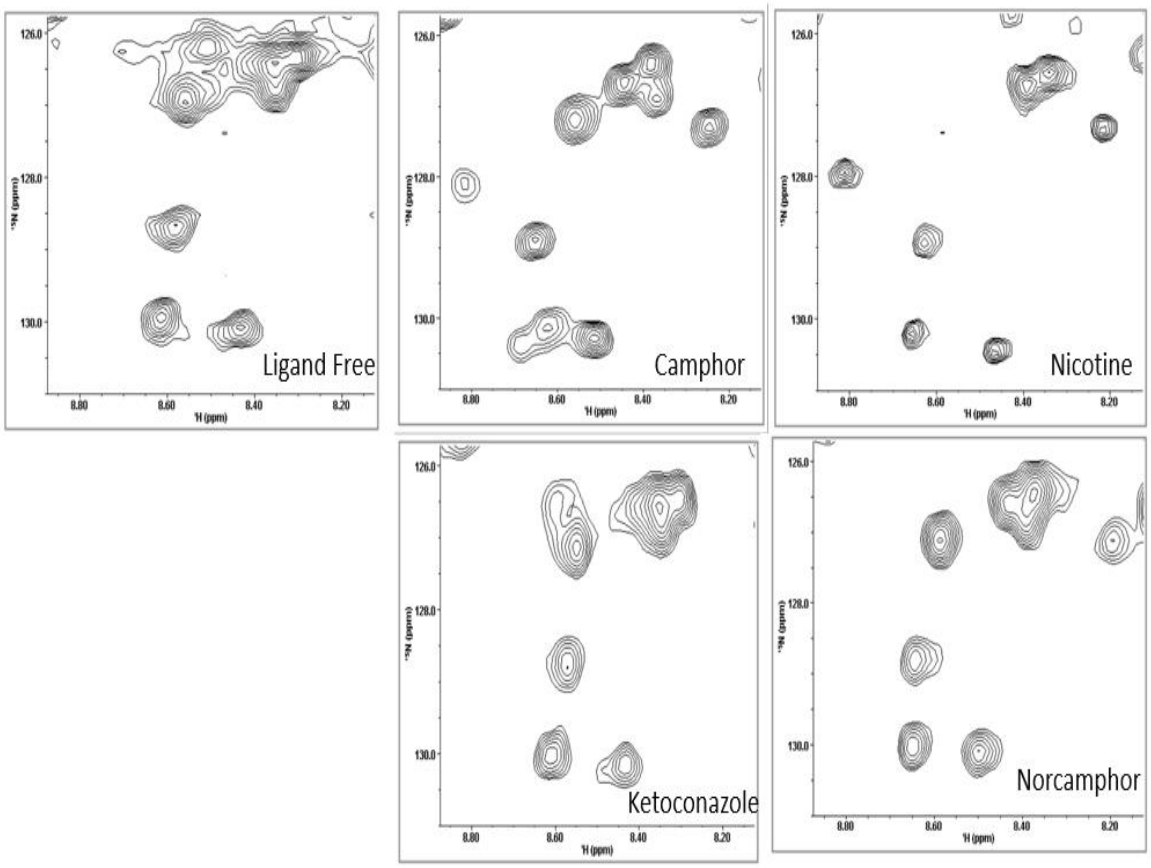


Figure 8: Zoomed in Portions of 2D HSQC-TROSY spectra for oxidized CYP101 in ligand-free and various ligand-bound forms showing differences in linewidth changes and peak splittings among the various spectra

CHAPTER TWO:
PREPARATION OF CYP101 FOR EXPERIMENTS

2.1 Cell growth and CYP101 protein expression

A pET24 vector-based plasmid (Novagen) that encodes for CYP101 with a C-terminal His_6 tag and Kanamycin antibiotic resistance was used to transform calcium competent BL-21 *E. coli* cells (Stratagene Inc). Transformed cells were plated on a LB medium plate containing 50 μg/mL of the antibiotic Kanamycin and incubated overnight at 37^0C. Colonies resistant to Kanamycin appeared on the plate in the morning, at which point a single colony of these cells was used to inoculate 50 mL of sterile LB broth containing 50 μL of 50 mg/mL stock of Kanamycin. Cells were incubated with shaking at 250 RPM and a temperature of 37 ^{0}C until cells reached an optical density of OD_{600}=0.6. Cells were then transferred to sterile centrifuge tubes and centrifuged for 10 minutes at 4^OC and 6000 RPM. The supernatant solution was decanted off and the remaining pellet was then suspended in 1L of enriched M9 media. The enriched M9 medium consisted of the following ingredients: 7.0g sodium phosphate dibasic anhydrous, 3.5g of potassium phosphate monobasic and 0.5g of sodium chloride/L, 1mL of 50 mg/mL kanamycin, 1mL of 1M magnesium sulfate, 1g of ammonium chloride, 4g of glucose, 75µL of 0.5 M iron(III) chloride, 100 µL of 1M calcium chloride, 1mL of trace metals (contents in g/L: 5 Na_2EDTA; 0.05 $FeCl_3$; 0.05 $ZnCL_2$; 0.01 $CuCl_2$; 0.01 $CoCl_2.6H_2O$; 0.01 H_3BO_3; 1.6 $MnCl_2.6H_2O$), and 25 µL of 2% thiamine. The ammonium chloride was replaced with ^{15}N labeled ammonium chloride and glucose replaced with $13C_6$-Glucose to achieve ^{15}N and ^{13}C uniform labeling respectively of the protein.

Cells were incubated with shaking at 250 RPM and 37°C in the enriched M9 medium until they reached OD_{600}=1, at which point the cells were induced to start protein expression with 1 mL of 1M isopropyl β-D-thiogalactopyranoside (IPTG). 35 mg/L of delta-amino levulinic acid hydrochloride was also added at time of induction to serve as a heme precursor. Cells were continued to shake for 12 hours at 37°C after induction, at which point they were harvested by centrifugation at 6000 RPM and the cell pellet stored at -80°C until use.

2.2 Protein Purification

Frozen cells were resuspended in 60 mL of phosphate buffer A (50 mM KPO_4, 50mM KCl, pH=7.4) for every 15 g of cells. Once fully suspended, the cell mixture was then put on ice and sonicated in a pattern of 20 seconds on and 1 minute off for a total of 4 times with a Branson Sonifier 250. The lysate was centrifuged for 15 minutes at 11000 RPM at 4^0C. The supernatant solution was then run through a second centrifugation cycle for 10 minutes at 11000 RPM at 4^0C to remove any remaining cell debris. The resulting supernatant solution was passed through a Ni^{2+} metal affinity (Talon Metal Affinity Resin, Clontech Laboratories) column, which captures CYP101 via the C-terminal His_6 tag. The captured protein on the column was washed with 20x the column volume with buffer A and eluted with 125 mM imidazole in buffer A. The eluate from Nickel column was then passed through a second column containing anion exchange resin Q Sepharose fast flow (GE Healthcare Life Sciences), washed with 20x the column volume witj buffer A and eluted with 250 mM KCl dissolved in buffer A. The eluted protein was then concentrated with a Millipore 30,000 MWCO centrifuge filter. The concentrated protein went through a final purification stage with a S100 size exclusion chromatography column (GE Life Sciences) running on an Amersham FPLC system. 1 mL fractions of CYP101 eluting from this column were collected and protein purity measured for them. Protein samples of oxidized CYP101 with a UV-Vis absorbance ratio of A_{391nm}/A_{280nm} greater than 1.4 were deemed to be at least 95% pure and were taken for further experiments. The concentration of the pure protein was determined via equation (2)

$$A_{418} = \varepsilon bc \qquad \textbf{\textit{Eqn}. 2}$$

Where A_{418} is the absorbance value at 418nm, which is where the heme absorbs in the absence of ligand, ε is the extinction coefficient, which is 100 M^{-1} cm^{-1} for CYP101, b is the path length of 1cm of the UV vis cuvette and c is the molar concentration of the protein. The 418 nM peak is sensitive to ligand binding and undergoes a blue shift upon substrate binding near the heme and a red shift upon inhibitor binding to the heme and can be used as a measure to indicate binding of a ligand (Figure 9). Interestingly, the blue shift can also be seen with the naked eye by observing a color change of the protein from red (ligand-free) to brown (substrate bound).

2.3 Selection of ligands to probe CYP101 dynamics

In order to comprehensively probe dynamic changes in CYP101 as a function of diverse ligand binding, a suite of 4 ligands was selected that differed in shape, size, chemical composition, binding affinities and binding site on the protein. The ligands used are shown in Table 1. Two substrates and two inhibitors were chosen, with varying affinities for CYP101. Camphor was the first substrate selected and is the natural substrate of CYP101, providing a base reference for the study as most previous structural and dynamic work on the CYP101 system has been on the camphor-bound form. Norcamphor was the other substrate chosen to be examined. Norcamphor is structurally very similar to camphor, keeping the same overall structure. However, there are two main differences between norcamphor and camphor: 1) lack of 3 methyl groups on norcamphor relative to camphor and 2) different positioning of oxygen on the ring structure [Table 1]. This structural difference is enough to reduce the binding affinity (K_d) of CYP101 for norcamphor by 2 orders of magnitude compared to that for camphor, as norcamphor is unable to make certain noncovalent interactions with residues in CYP101 as camphor does. Thus, use of these two different substrates allows us to probe the dynamics as a function of different substrate binding affinities. In the absence of Pdx, CYP101's reaction cycle does not proceed beyond the protein-substrate resting state [35]. This allows us to look at substrate binding without the risk of conversion to product and without the addition of another inhibitor that could affect the dynamics.

The two other ligands selected for this study are nicotine and ketoconazole, which differ greatly in size and chemical properties. Nicotine is a small ligand that binds directly to the iron in the heme with high affinity. Nicotine can act as either a substrate or inhibitor, depending on which P450 it is interacting with. In humans, P450s like CYP2A6 and 2B6 use nicotine as a substrate for C-oxidation, whereas other P450s such as CYP 2E1 see an inhibition by nicotine [49, 50]. While nicotine is hydroxylated by CYP101, it binds in a manner that is consistent with other CYP101 inhibitors. Ketoconazole is an synthetic drug that is a derivative of phenylpiperazine, and is used as an antifungal drug due to its ability to inhibit microsomal P450s. Ketoconazole also inhibits CYP101, likely by binding directly to the iron, based on analysis of other P450 crystal structures bound to ketoconazole [51-53]. Ketoconazole is a bulky, hydrophobic ligand, but binds with a

surprisingly similar affinity as nicotine. Ketoconazole also binds to and inhibits human P450s in similar ways as they do to CYP101, implying medical relevance of understanding of their effects on P450 dynamics [54]. Overall, the range of ligands selected should give a comprehensive picture of CYP101's dynamics with different types of ligands. The structure of CYP101 bound to 3 of the 4 ligands as well as the ligand-free structure has already been elucidated, providing a structural basis for the interpretation of dynamic changes upon differential ligand binding.

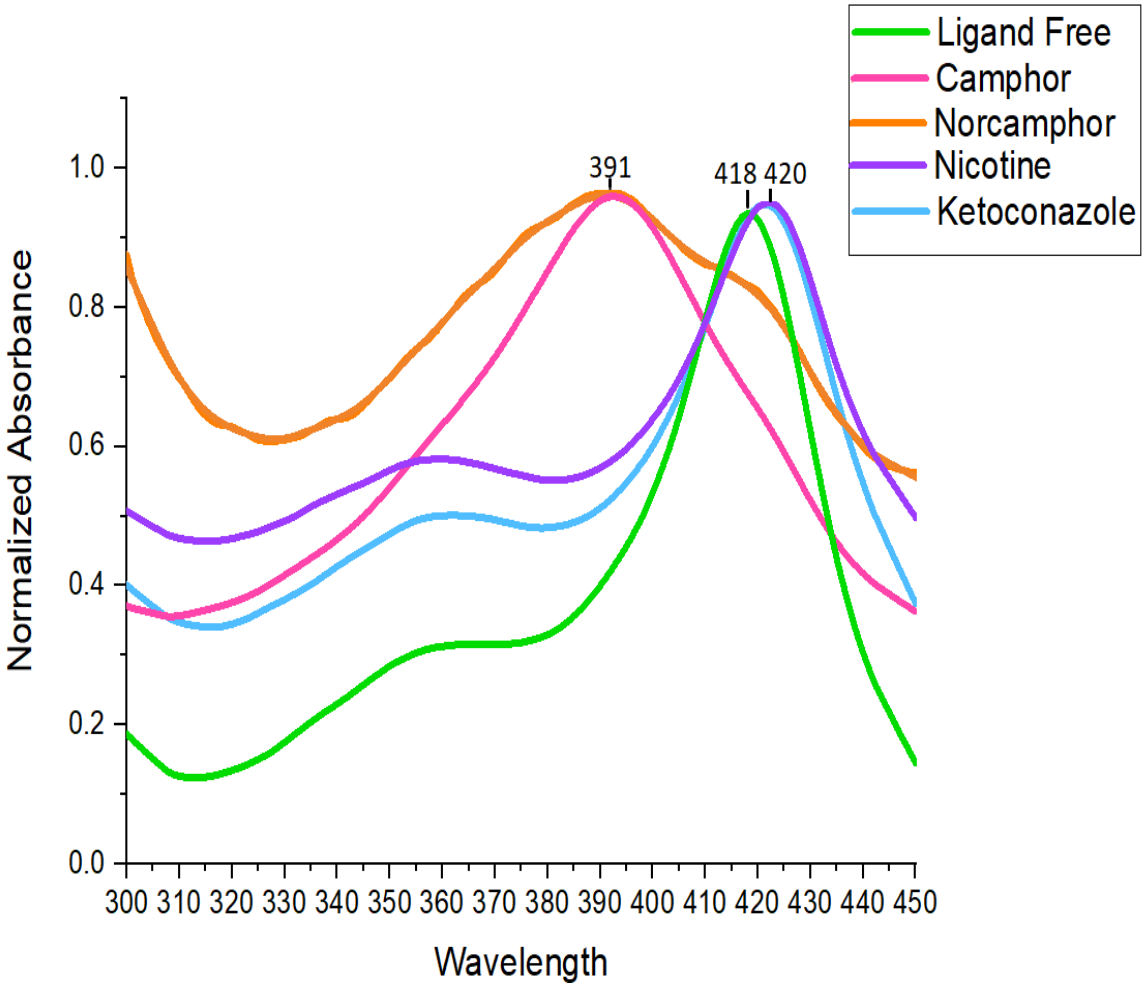
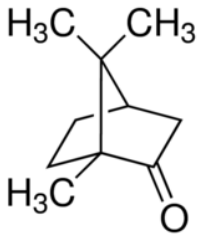
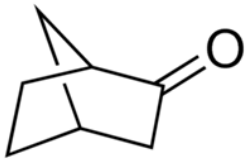
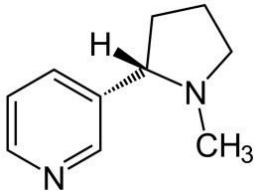
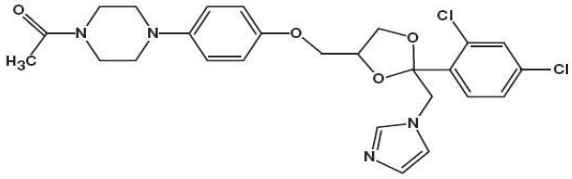


Figure 9 : UV-Vis spectra of CYP101 bound to each ligand

Table 1: List of ligands used to study dynamics of CYP101

Ligand	K_d	Classification	Structure
Camphor	1 μM [55]	Substrate, natural	
Norcamphor	150 μM [55]	Substrate	
Nicotine	10 μM [56]	Inhibitor	
Ketoconazole	.5 μM [52]	Inhibitor	

2.4 Preparation of NMR samples

The final sample preparation for the NMR experiments was similar for both resonance assignments and amide exchange experiments. The buffer solution consisted of 50 mM potassium phosphate (pH=7.4), 50 mM KCl, and trace amounts of β-mercaptoethanol. Higher KCl concentration of 150 mM was used to ensure ligand solubility for ketoconazole and prevent aggregation effects. Spectra collected with 150 mM KCl concentrations as well in the absence of β-mercaptoethanol did not show any detectable changes relative to the spectra in buffer containing 50 mM KCl and/or β-mercaptoethanol indicating that the change in conditions do not affect the spectral patterns seen once ligands are added. Protein concentration between 0.15 mM and 0.4 mM were used for various experiments in order to maximize signal intensity and resolution without protein aggregation. The hydrophobic ligand ketoconazole was dissolved in methanol before being added to the protein solution, while the other 3 ligands were dissolved in the same buffer as the protein and used directly at the appropriate concentration. Each ligand was added into the ligand-free protein until saturation was reached, based on achievement of maximum shift in the protein heme resonance in UV-Vis spectra upon ligand binding.

2.5 Preparation of Selective Isotope labeled samples of CYP101

CYP101 samples for ligand-free and all 4 ligand-bound forms were prepared with selective ^{15}N labeling of different amino acid types and used for 2D ^{1}H-^{15}N HSQC-TROSY spectral data collection. The protein samples were prepared and purified as described above, but with one major difference in that a single ^{15}N isotopically labeled amino acid was added into the enriched M9 media along with 19 unlabeled amino acids and other ingredients. The amino acids were added as an amino acid mix in the followings amounts (g/L): 0.5 A, 0.3 R, 0.8 D, 0.4 N, 0.25 C, 0.5 G, 2.0 S, 0.3 H, 0.0 P, 0.5 M, 1.0 E, 0.5 Q, 0.3 K, 0.3 F, 0.3 Y, 0.2 W, 0.5 L, 0.4 I, 0.5 T, and 0.5 V. The unlabeled amino acid was replaced with the appropriate ^{15}N labeled amino acid depending on the amino acid type labeling desired. For example, for preparing a sample labeled selectively with ^{15}N glycine, the unlabeled glycine in the amino acid mix was

replaced with the same amount of ^{15}N labeled glycine, while the rest of the amino acids were added in unlabeled form. Selective labeling in this fashion allowed the collection of a 2D NMR ^{1}H-^{15}N HSQC-TROSY spectrum that will only contain the resonances for the labeled amino acid type. The following amino acid types were labeled in this fashion: alanine, glycine, isoleucine, leucine, lysine, phenylalanine, and valine. These amino acid types were selected since they form the bulk of the amino acids found in CYP101 and have little to no scrambling tendency of their labels to other amino acids during protein expression, thus allowing unambiguous selective identification of their peaks in NMR spectra. NMR spectra were collected and processed for each of these samples in ligand-free and all 4 ligand-bound forms in a similar manner to the uniformly labeled samples.

2.6 Two- and Three-dimensional NMR experiments for resonance assignments

All NMR experiments were run on a Varian 600 MHz spectrometer equipped with a cold probe. ^{1}H-^{15}N 2D correlation spectra with a HSQC-TROSY pulse sequence were collected on ^{15}N uniformly labeled CYP101 samples to detect 2D amide resonances for ligand-free and all ligand-bound forms of the protein. A typical ^{1}H-^{15}N 2D experiment consisted of 16 scans acquired with 1024 complex points in the direct dimension and 128 increments in the indirect dimension. Depending on the protein concentration, the number of scans were varied form sample to sample to acquire spectra with similar sensitivity.

^{1}H-^{15}N-^{13}C 3D correlation NMR experiments to assign backbone resonances were carried out for ligand-free protein and the 4 ligand-bound forms. Standard 3D ^{1}H-detected TROSY versions of HNCA, HN(CO)CA and HNCO pulse sequences available on the Varian spectrometer were used to collect data sets for ligand-free and ligand-bound forms (camphor, nicotine, and ketoconazole) of CYP101. 3D ^{1}H-detected TROSY version of HNCACB pulse sequence was also used to collect data for camphor-bound and ligand-free forms. Only the HNCO and HN(CO)CA data sets were collected for the norcamphor-bound form of CYP101 as the HNCA peak correlations are likely similar enough to the camphor-bound form that it would be possible to assign

resonances in norcamphor-bound form by comparison. A typical 3D experiment consisted of 16 scans acquired with 1024 complex points in the direct proton dimension, 48 increments in the indirect carbon dimension and 32 increments in the indirect nitrogen dimension. The number of scans for each experiment was varied depending on sample concentration of a particular CYP101 form and to obtain sufficient sensitivity in peak correlations for data analysis.

The spectral data from all NMR 2D and 3D experiments described above was processed using NMRPipe software and analyzed using NMRViewJ software [57, 58]. The experiments were processed using sinusoidal and exponential window functions with a mild baseline correction. Linear prediction was applied in indirect dimensions with zero filling to double the number of points in the indirect dimensions and final spectra phase corrected before using for data analysis and resonance assignments.

2.7 Collection of Amide exchange data

H_2O/D_2O amide exchange (HDX) experiments were performed for ligand-free and all 4 ligand-bound forms of the CYP101 to monitor slow motions on the s-min timescale. Ligand-bound samples were prepared with saturating concentrations of ligand present. In order to measure exchange of protonated amides with deuterium, a fully protonated ^{15}N uniformly labeled sample was prepared and a 2D reference spectrum acquired initially. After the spectrum was collected, the protein was taken out of the NMR tube and lyophilized for a minimum of 12 hours to remove the hydrogenated solvent and taken to the NMR room. Just before the start of the HDX experiment, the lyophilized sample was quickly rehydrated with D_2O and brought up to a similar volume as the sample used to collect the reference spectrum. The lyophilized protein was checked for aggregation and stability by comparing the first collected spectra after exchange process was started to the reference spectrum collected before lyophilization and seeing similar peak characteristics for the peaks that were not exchanged out. Reference and all of the amide-exchanged 2D NMR ^{1}H-^{15}N HSQC-TROSY spectra were collected in a similar fashion to that described for resonance assignments with the number of scans reduced to 8 to obtain optimal spectra with sufficient signal intensity

but with frequent enough sampling of the amide exchange landscape to map out the exchange rates. After the exchange process was started, amide-exchanged spectra were collected every 28 minutes for approximately 40 hours to categorize both the fast and slow exchange rates. The first time point was collected within 5 minutes of the protein's exposure to D_2O to allow for detection of the fast exchange resonances. Samples were checked for stability by UV-Vis spectroscopy to ensure that the samples were stable through the length of the HDX experiments. All spectra were processed using NMRPipe and NMRViewJ in a similar manner to the ones used for resonance assignments and then taken up for analysis.

CHAPTER THREE:
SEQUENCE-SPECIFIC NMR RESONACE ASSIGNMENTS FROM MULITDIMENSIONAL HETERONUCLEAR EXPERIMENTS

One of the most critical steps in NMR data analysis is to assign peaks in a sequence-specific manner for the protein. For small proteins (<10 kDa), this process traditionally entails collection of multidimensional NMR spectra using ^{1}H detected pulse sequences such as 2D COSY and 2D NOESY due to the high sensitivity of protons relative to other nuclei. This approach however fails for larger proteins that contain thousands of protons causing resonance overlap among the multitude of proton signals observed for such proteins. Efforts to resolve this resonance overlap are critical in resonance assignment of larger proteins such as CYP101. In our case, sequence-specific backbone N-H resonance assignments are needed for the oxidized ligand-free and the 4 ligand-bound forms of CYP101 to characterize the dynamic differences between the forms, which required a strategy described below.

3.1 Three-dimensional Heteronuclear NMR experiments for resonance assignments in CYP101

Over the last three decades, ^{1}H-detected pulse sequences have evolved to include heteronuclei such as ^{13}C and ^{15}N which makes it easier to employ pulse sequences for resonance assignment of large proteins that rely on building sequential connections among adjacent residues via heteronuclear signal editing, allowing backbone and side-chain assignments to be made with more ease. These sequences correspond to a suite of multidimensional (2D, 3D and 4D) heteronuclear experiments which allow correlations to be made between ^{1}H from backbone and side-chains of amino acids in ^{13}C,^{15}N uniformly labeled proteins through heteronuclear scalar couplings to the specific heteronuclei that they are directly attached to (e.g. N, Cα, Cβ etc.). Several 3D heteronuclear experiments such as HNCA, HN(CO)CA, HNCO and HNCACB are available to build such correlations, originating from ^{1}H magnetization which is then transferred to other nuclei via spin manipulation allowing correlation of multiple nuclei within an amino acid and also with nuclei in adjacent amino acids. For example, in the HNCA experiment, the pulse sequence starts with the ^{1}H magnetization

of the backbone amide of an amino acid, then transfers to the attached ^{15}N via heteronuclear scalar coupling and then to the $^{13}C\alpha$ of its own amino acid as well as the $^{13}C\alpha$ of the previous amino acid in the uniformly labeled ^{13}C, ^{15}N protein. Thus, this experiment will show the N, H and Cα chemical shift correlations for a residue(i) as well as the $^{13}C\alpha$ chemical shift of the residue(i-1) previous to it in the sequence. This can be used to trace assignments along the protein sequence in a sequential manner. The HN(CO)CA experiment, on the other hand, manipulates spins in such a way that only correlations from the N-H chemical shifts of a certain amino acid to the previous residue's Cα shift (i-1) via the intervening carbonyl group (CO), but not the self-residue's Cα shift (i), are shown. This helps in resolving ambiguities about which resonances belong to self-amino acid and which belong to previous amino acids in the sequence. Similarly, the HNCO experiment will correlate the N-H chemical shifts of an amino acid to the carbonyl carbon chemical shift of the previous amino acid with which it forms a peptide bond with, again connecting one amino acid with the previous amino acid. HNCACB experiments correlate similar H, N and C chemical shifts as the HNCA, however offer even more refinement in determination of the self and previous amino acid's identity by extending it to the Cβ shift for the self and previous amino acids along with the Cα shifts. The Cβ shifts are quite different for different amino acids and also differ in sign of peak intensity, which allows better discrimination of the identity of amino acids by amino acid type. The combination of HNCA, HNCO, HNCOCA and HNCACB experiments not only helps determine the identity of specific amino acids but allows building of connectivities for self and previous residues in a sequence-specific manner that can be walked backwards on to determine assignments for backbone N-H amide groups and Cα for each residue in the protein (Figure 10).

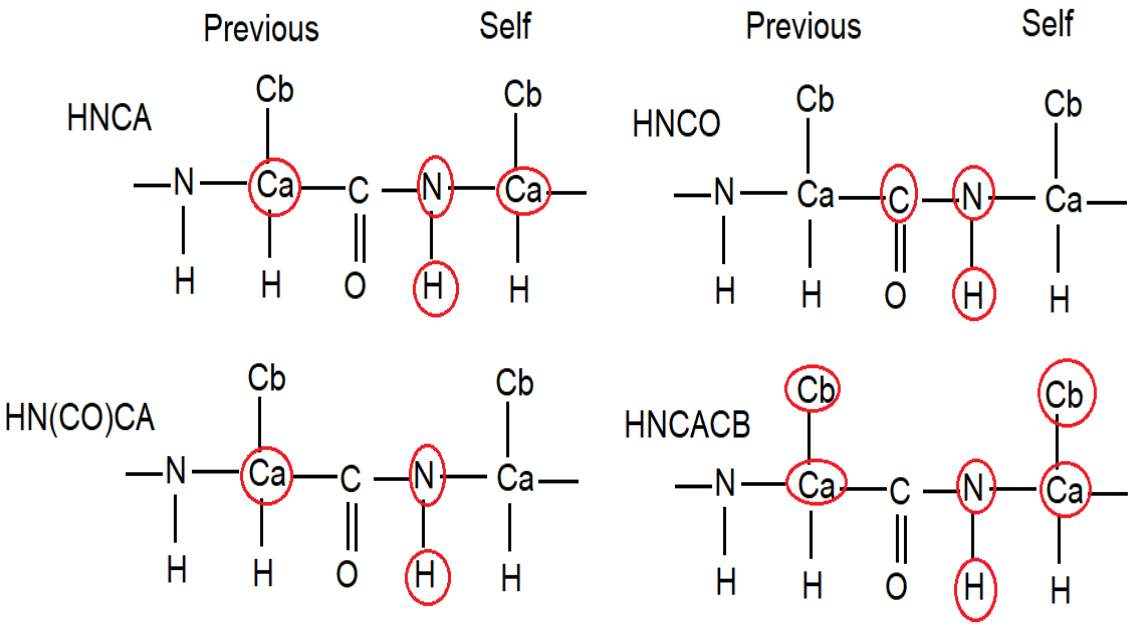


Figure 10: 3D NMR heteronuclear resonance assignment experimental scheme. Red circles indicate the chemical shifts of nuclei that will be visible in a certain experiment. Note that the N and H chemical shift in the HN(CO)CA experiment is correlating to the self.

3.2 Sequence-specific Assignments of various forms of CYP101

In the case of CYP101, multidimensional heteronuclear NMR experiments such as those described above have been performed previously to obtain sequence-specific resonance assignments in various redox forms of CYP101. Two published data sets are available for the reduced form of CYP101 from Pochapsky's group, which includes N, H and Cα chemical shifts (BMRB 19740 and 17415) [59, 60]. The reduced CYP101 data set makes available approximately 300 assignments for CYP101. Additionally, two published data sets of NMR assignments (BMRB 5759 and 19038) are also available for the oxidized camphor-bound form of CYP101 from two different research groups [60, 61]. These data sets obtained independently have approximately one hundred backbone N-H assignments, with significant overlap between the two sets. The large difference in number of assignments between oxidized and reduced CYP101 can be attributed to the presence of redox state of the heme group. In the oxidized Fe^{3+} form at physiological temperature, the heme is paramagnetic and broadens out resonances within an 8-12 Angstrom distance, depending on the orientation of the heme group's side prosthetic group and the fold of the protein polypeptide around the heme group. This causes the resonances to vanish from the multidimensional spectra in a random manner and breaks up the sequential connectivity in 3D NMR assignment experiments. This is not an issue in the reduced form of CYP101, as the iron gets converted to the Fe^{2+} state and switches from paramagnetic to diamagnetic, which does not have the same resonance broadening effect. Comparing the reduced and oxidized assignments of the protein will typically show large differences in the N and H chemical shifts but somewhat similar values for the Cα chemical shifts. This can be exploited to transfer some of the resonance assignments from the oxidized to the reduced CYP101 data sets, which is what was done by the two research groups in obtaining assignments for oxidized CYP101.

In order to characterize protein dynamics throughout the protein for oxidized ligand-free and the 4 ligand-bound forms of CYP101, resonance assignments are needed for all 5 forms of the protein. However, the published data sets only provide

assignments for the camphor-bound form, hence the assignment process had to be carried out for the other forms. The chemical shifts in the spectra for each of the forms are distinct enough that transfer of backbone N-H assignments from one form to another was possible only for a few outlying peaks in spectra, however could not be accomplished with confidence for the regions in the middle of the spectra where there is significant overlap. Another compounding problem is that the number of available assignments in the camphor-bound form are not sufficient to provide adequate coverage for the dynamic regions of the protein, especially the SRS regions. This requires use of traditional sequential assignment strategy involving acquisition of 3D heteronuclear NMR spectra and analyzing the data from them independently for each of the CYP101 forms, which was not performed in the previously published work. 3D NMR data sets were therefore collected for each CYP101 form as described in the experimental section in Chapter 2 and the assignment process undertaken from the resulting data (Figure 11). The assignments obtained with this strategy were further confirmed and/or augmented with peak inferences made from selectively labeled spectra.

There are several challenges in establishing NMR assignments for a protein such as CYP101. First, CYP101 has a large number of prolines (35) in its structure and because of proline's unique amino acid structure lacking hydrogen in its backbone amide group, the N-H resonances for proline do not show up in ^{1}H-detected 2D and 3D spectra. This can break up sequential connectivity when trying to work backwards in the 3D datasets. Second, the paramagnetism of the heme group hampers establishing sequential connections, especially in regions close to the heme group such as the SRS and I helix. As discussed previously, the residues in the SRS are the most flexible of all regions in the protein and therefore it is crucial that maximum number of assignments are available in order to characterize the dynamic changes in these regions in a comprehensive manner. While certain residues are not visible for the SRS, the majority of the region is remote enough from the heme that approximately 41-48 of the 55 important residues can be assigned depending on the ligand form. While focus in the

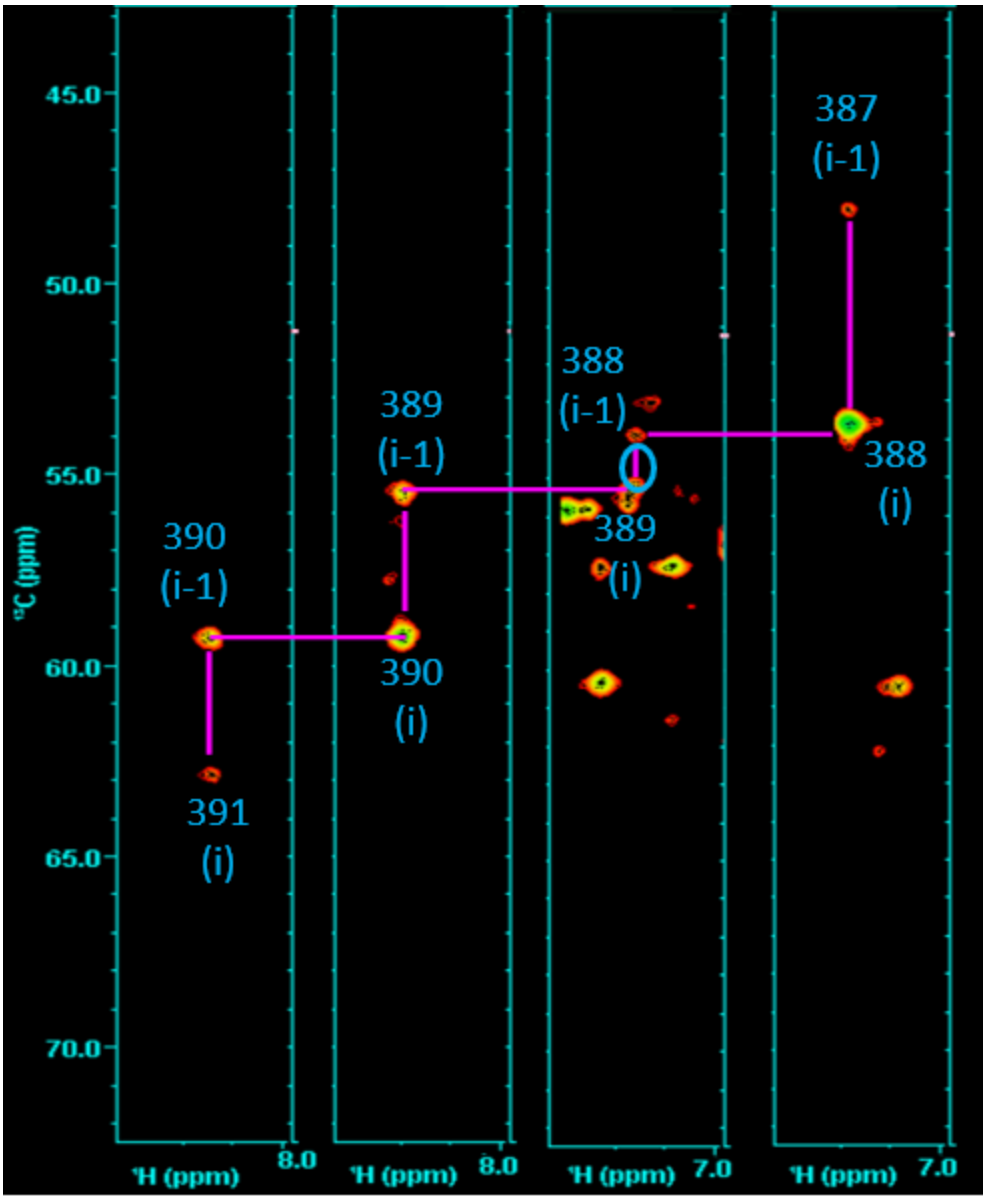


Figure 11: Example of sequential assignment of a stretch of residues 387-390 from ketoconazole-bound CYP101 using a 3D HNCA spectrum

current project was primarily directed towards assigning the resonances for SRS in the protein, however additional assignments have been also been made in other regions of the protein. The assignment process for the various CYP101 forms in this project also resolved discrepancies or errors in assignments between the two previously published data sets.

The assignment process was started using the published assignments for the 2D HSQC-TROSY spectra as a starting point. Since the published data is available only for the camphor-bound form, assignments were made for this form first by using the published assignments outlying resonances as a reference. Based on similarity of H, N and Cα chemical shifts, N-H assignments were transferred from the previous camphor-bound data set to the current data set where there was sufficient confidence and then verified independently via the 3D dataset connectivity, accepted Cα and Cβ chemical shifts, and their presence in selective labeled spectra. This independent verification also allowed to identify approximately 8 assignments that were erroneous in previous publications. A similar strategy was followed to transfer assignments from the camphor-bound form to other ligand-bound forms as well as the ligand-free form. As seen with previously published reduced and oxidized CYP101 spectra, the Cα chemical shifts tend to stay within a smaller range than the N and H chemical shifts, which allowed for comparison of the 3D datasets and transfer of assignments between the various CYP101 forms. Again, where there was difference in chemical shifts observed during the transfer process, independent verification as described above was undertaken.

Based on this assignment process, assignments totaling between 138-152 were obtained per CYP101 form (Table 2). This represents an increase of almost 50% assignments compared to previously published list and more importantly a majority of the resonance assignments in the SRS region which were only partially made in previous work [35, 59, 61]

Table 2 Summary of 3D NMR experiments on various forms of CYP101 and details of resonance assignments available (x indicates experiment performed for a certain CYP101 form)

	Ligand Free	Camphor	Norcamphor	Nicotine	Ketoconazole
HNCA	X	X		X	X
HNCO	X	X	X	X	X
HNCACB	X	X			
HN(CO)CA	X	X	X	X	X
Total N-H Assignments	138	152	139	149	140
Total SRS N-H assignments	41	48	44	47	42

Several of the assignments were already established in previously published data sets, but some assignments are novel to this work. Overall, the most assignments were obtained for the camphor-bound spectra. As the camphor-bound form's chemical shifts were used as a reference for assigning the other forms, this represents the maximum limit for this dataset. The nicotine-bound form has a similar number of assignments to camphor, due to its fairly narrow peaks that allowed good resolution to see peak connectivity. Ligand-free, norcamphor-, and ketoconazole-bound form assignments suffered from dynamic line-broadening, causing peaks to not appear with sufficient intensity in the spectra, thereby losing the connectivities during the assignment process. For all ligand-bound and ligand-free forms, new assignments were added than present in previous data sets for the F and G helices, FG loop as well as the N terminal loop that makes up the bottom lip of the SRS. Additional assignments have also been made to the β5 sheet and ends of the B' helix, making for a thorough coverage of the SRS. For regions outside the SRS, higher coverage of assignments has been obtained, particularly in the three corner regions of the triangular structure. Combined together, this provides adequate coverage of the entire protein to make

interpretation of dynamic changes in the various CYP101 forms from both a local and global perspective.

3.3 NMR spectra of ^{15}N selectively labeled CYP101 samples

^{15}N selectively labeled CYP101 spectra for different amino acid types were also collected to verify and augment the resonance assignments obtained from the 3D NMR sequential assignment strategy. This involved preparing the protein samples using a single ^{15}N labeled amino acid with the remaining 19 amino acids unlabeled so that the 2D ^{1}H-^{15}N HSQC-TROSY spectrum will only have the chemical shifts for the labeled amino acid visible (Figure 12). Similar ^{15}N selective labeled spectra were collected for all 4 ligand-bound and ligand-free forms of oxidized CYP101 for the following 7 amino acids: Alanine, Glycine, Isoleucine, Leucine, Lysine, Phenylalanine, and Valine. As expected, the selectively labeled spectra in all cases contained considerably less number of peaks compared to the corresponding uniformly labeled spectra indicating that the selective labeling was successful. For 4 of the amino acids labeled - Alanine, Lysine, Phenylalanine and Valine, the number of peaks in the selectively labeled spectra generally matched the number of amino acids of that amino acid type in the amino acid sequence of CYP101. About 4 Alanine, 7 Valine, and 6 Phenylalanine peaks were missing due to paramagnetism in every ligand-bound spectra. The ligand-free, ketoconazole-bound and norcamphor-bound spectra showed approximately 3-4 fewer peaks than camphor- and nicotine-bound due to broadening from dynamics. However, for the remaining 3 amino acid types - Glycine, Leucine and Isoleucine, additional peaks were observed in the spectra than were expected, which was the result of scrambling of the intended labeled amino acid to other amino acids during protein expression. For example, the ^{15}N label on the amino group of Glycine is known to scramble to serine and threonine due to shared metabolic pathways which leads to interconversion of amino acid backbone between these three amino acids.

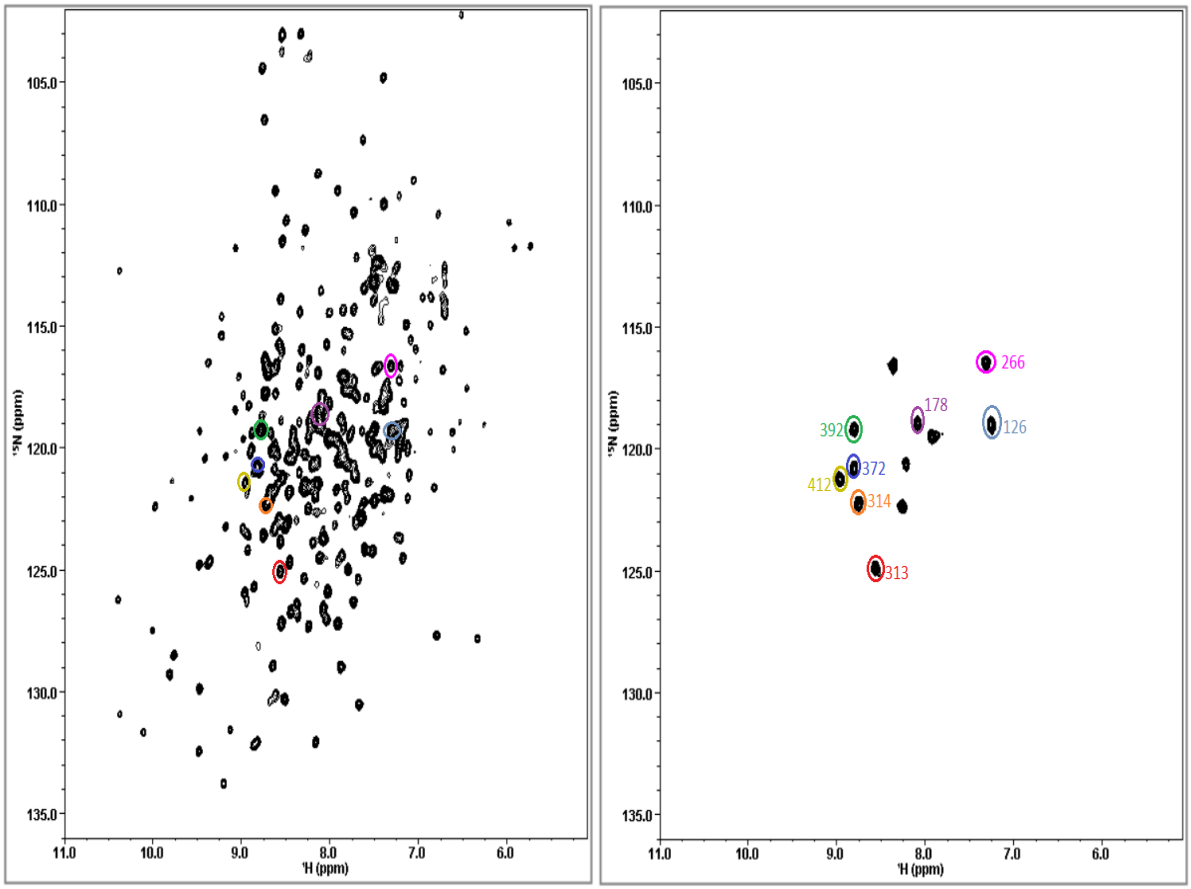


Figure 12: Comparison of 2D ^{1}H-^{15}N HSQC-TROSY of camphor-bound CYP101 uniformly labeled (left) and selectively labeled with ^{15}N Lysine (right). Only the ^{15}N lysine peaks are seen, identifying some of the peaks in the uniformly labeled spectrum by amino acid type

This actually was useful in our case, since it allowed the analysis of all three amino acid types without having to prepare a separate sample for serine and threonine. The three amino acid types could be separated based on their unique Cα and Cβ chemical shifts. Leucine is known to scramble to both Isoleucine and Valine, however since no scrambling as seen in the Valine spectra and very little scrambling seen in the isoleucine spectra, the peaks from these spectra were used to identify the peaks for leucine. Comparison of the selectively labeled CYP101 spectra with uniformly labeled CYP101 spectra allowed for resolution of ambiguities in the 3D spectra, helping in assignment of several of the peaks for each amino acid type. There were some peaks that remained unassigned for each amino acid type as sequential connectivities could not be established for them. However, at least the amino acid type is known for them and would be useful in future attempts at completing assignments of the protein.

CHAPTER FOUR:
MODELING OF CYP101-KETOCONAZOLE COMPLEX STRUCTURE VIA MOLECULAR DOCKING

Interpretation of dynamic changes between the various CYP101 forms from a functional perspective requires structural information on all forms. Currently, the crystal structures of CYP101 in the ligand-free form and 3 out of the 4 ligand-bound forms (camphor, norcamphor and nicotine) are available. From the preliminary NMR data, the ketoconazole-bound form of CYP101 is found to be fairly flexible and thus is an important complex to study in terms of dynamics. However, there is no published structure for CYP101 bound to ketoconazole, which limits the dynamic interpretation that can be done for this complex. Attempts to crystallize CYP101 bound to ketoconazole based on published crystallization conditions for ligand-free and other ligand-bound crystal structures of CYP101 have been unsuccessful, preventing elucidation of ketoconazole-bound CYP101 structure via Xray crystallography.

NMR spectroscopy offers another avenue for solving the ketoconazole-bound structure, however, can be quite challenging. Traditionally, solution NMR methodology has been used to solve *de novo* structures of proteins and protein complexes that are 25 kDa in size or below, and CYP101 is ~ 44kDa, well above that limit. This large size can broaden spectral linewidths due to decreased molecular tumbling, which makes it difficult to collect intermolecular restraints in form of NOEs and residual dipolar couplings. In addition, due to the large number of amino acid residues (414) in CYP101, there is bound to be significant overlap of resonances because of spectral crowding that could interfere with the ability to determine the intermolecular restraints with sufficient accuracy. Deuteration of CYP101 can potentially be used to overcome these limitations, but at the expense of losing protons in the protein which mainly contribute to the intermolecular NOEs. The problem is further compounded by the limited availability of resonance assignments for oxidized CYP101 due to paramagnetic effect from the heme, especially in the SRS region which can significantly reduce the number of intermolecular constraints available for this important region. Based on these considerations, utilization of traditional NMR methods to solve the complex structure of

ketoconazole-bound CYP101 was not deemed to be deserving of the required effort. Instead, use of molecular docking methods was considered to obtain the structure of this complex.

There are several options currently available in terms of molecular docking software such as DOCK, ROSETTA, AUTODOCK VINA and HADDOCK to accomplish structural modeling of the complex. After test runs of docking using all of these software and evaluating the pros and cons, HADDOCK was chosen to enable the docking of the complex [62]. There are several advantages of using HADDOCK compared to the other docking software. The biggest advantage is that the docking protocol allows use of NMR-derived or other forms of direct restraints to facilitate docking of two molecules, which is not possible in the other docking software. It also allows use of indirect restraints in form of flexible and non-flexible residues allowing flexible docking that permits sampling of multiple conformations. An optimal structure can be calculated using energy minimization and short molecular dynamics simulation as the final step in the procedure. This circumvents the need for having to use a separate molecular dynamics program to further equilibrate the docked structures, which would be necessary, for example in AUTODOCK VINA. The flexible docking feature is especially important for this complex, as preliminary NMR data shows that CYP101 is very flexible, especially in the ketoconazole-bound form. Availability of NMR assignments and other indirect restraints such as chemical shift perturbations from comparison of ketoconazole-bound spectra with ligand-free form spectra, allowed us to define which residues can be made flexible and which ones can be kept fixed during the docking, which permits a guided docking process for the complex rather than random rigid body docking, increasing the probability of obtaining feasible structures for the complex. HADDOCK outputs the final docked and equilibrated structures after water refinement with a ranking based on calculated energy of favorable interactions and RMSD, which allows the user to evaluate the validity of the docked structures.

4.1 Docking Protocol for the Ketoconazole-CYP101 complex

The docking of ketoconazole was performed using the ligand-free structure of CYP101 published previously (PBD id: 3L62) [40]. The ligand-free structure was chosen as the initial structure for docking as NMR chemical shift perturbation data suggested that the ketoconazole-bound structure conformation resembled it the closest relative to the other ligand-bound CYP101 structures. Since the PDB file for the ligand-free structure is missing structural coordinates for the disordered regions such as the B' helix and a N terminal fragment of 10 residues, these regions were rebuilt in the molecular operating environment (MOE) software. MOE was also used to calculate the partial charges on the protein and do energy minimization after rebuilding of the structure. As a double-check, partial charges were also calculated in Chimera, and no appreciable difference in assignment of partial charges was seen between the two programs, despite the fact that they use different methods to calculate the charge [63]. The initial three-dimensional structure for ketoconazole molecule was generated using CHEMDRAW (Figure 13). Partial charges and hydrogens were added to this structure using CHIMERA followed by energy minimization of the structure in MOE. Energy parameter and topology files for ketoconazole to be utilized in the docking process were also generated in CHEMDRAW.

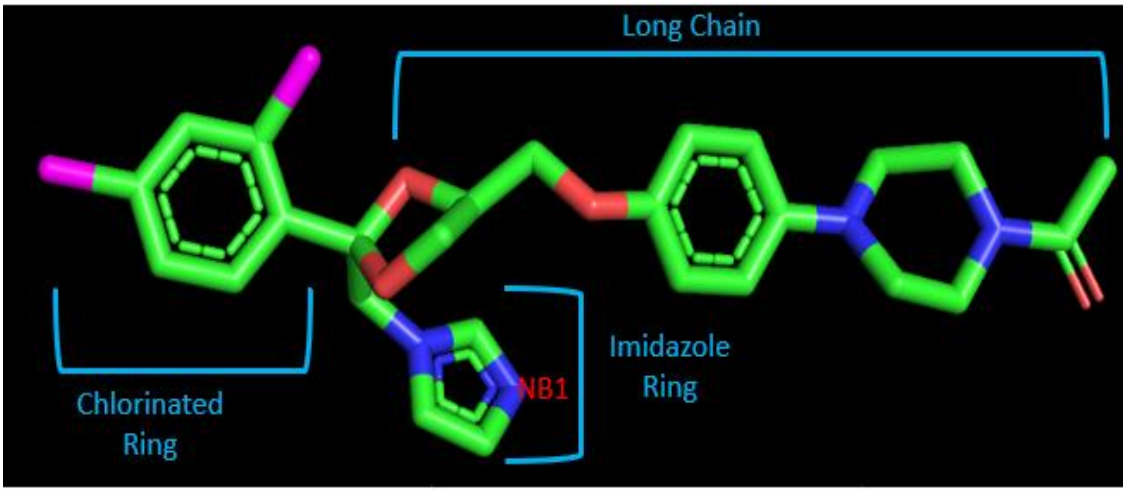


Figure 13: Ketoconazole initial structure generated from CHEMDRAW

With the initial structures for the protein and ligand generated for docking, restraints were defined in HADDOCK to bring the structures together in a rigid body docking process. Since it is possible in HADDOCK to specify which specific interactions the ligand can form with the protein, a restraint of 0.5 Angstroms (with an upper and lower bound of 1 Angstroms) between the NB1 nitrogen in imidazole ring of ketoconazole structure (Figure 13) and the heme iron was specified based on previously available information on the ligating site of ketoconazole with CYP101 in other published P450-ketoconazole complex structures [52, 64]. Similar bond formation is seen for the other inhibitors such as nicotine when it interacts with CYP101. The general binding site for ketoconazole on CYP101 for initial rigid body docking was designated as the opening formed by the SRS regions which is known to provide an entry channel to the heme active site. The regions in SRS such as the B' helix and tip of the F-G loop were set as fully flexible whereas the semi-flexible interfaces were defined as the rest of the FG loop, the F/G/H helices, and part of the β5 sheet (380-400), based on the NMR chemical shift perturbations observed between the ketoconazole-bound and ligand free spectra. Other minimal loose distant restraints were defined for the

flexible branched ring structures of ketoconazole to guide the structure into the binding pocket since the unconstrained rings in initial docking runs were observed to generate steric clashes with the residues in binding pocket. These included a restraint of 3 Angstroms from the chlorinated ring of ketoconazole to Phe98 and Phe193, and from the oxygen on the longer chain to Lys392 and Ser393 and were equally weighted for the docking run. These restraints were defined based on similar interactions of ketoconazole rings with the protein observed in the same general location as deduced from previous P450-ketoconazole complexes [51, 65, 66]. Energy parameters and topology for the protein were generated by HADDOCK during the docking process. The docking was performed on a local Linux server and took approximately 20 hrs per run (Figure 14).

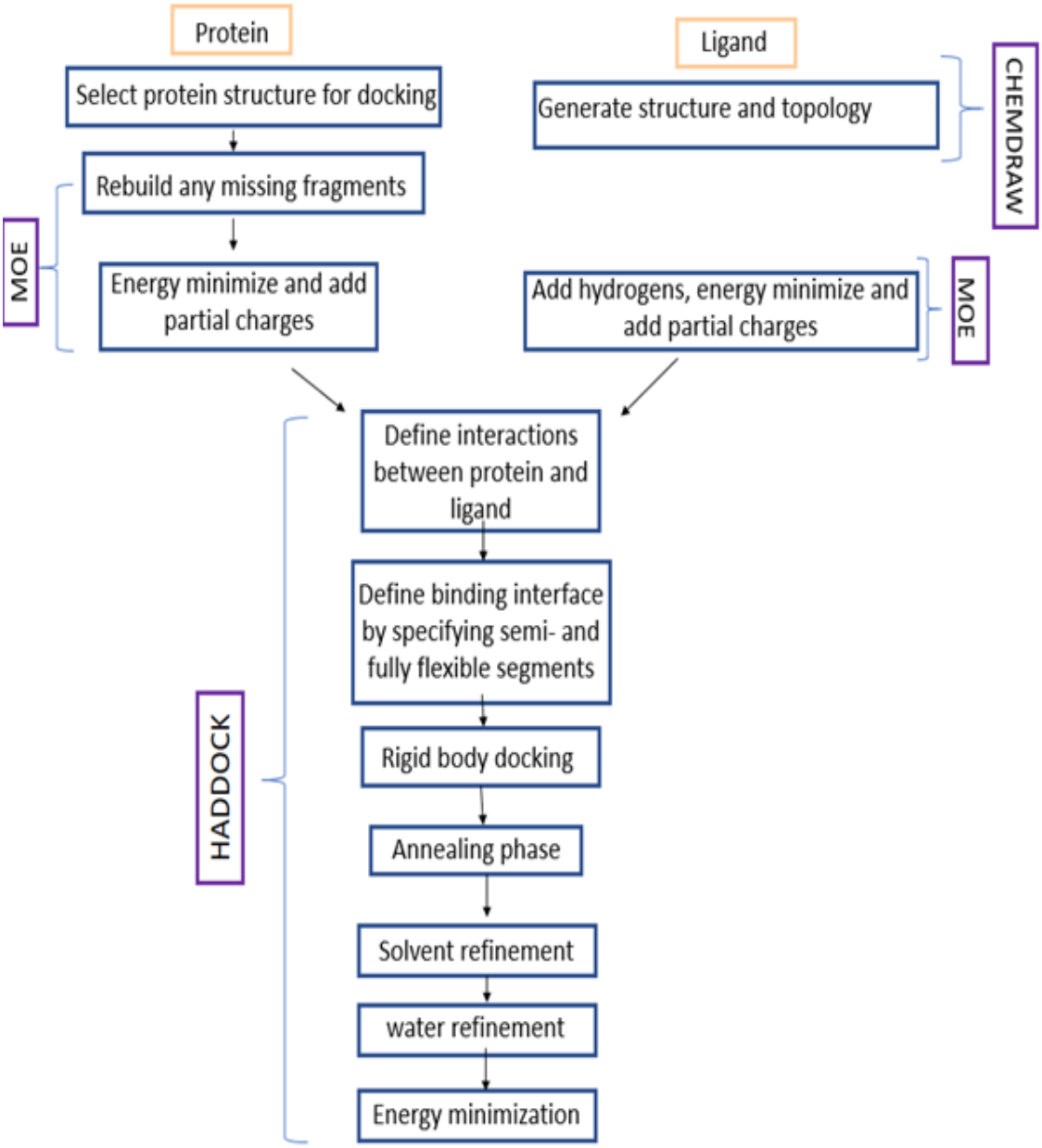


Figure 14: Flow chart showing the docking protocol for CYP101-ketoconazole complex

For each run, 1000 structures were generated during rigid body docking, and 200 of these structures were chosen based on HADDOCK's automated scoring system for refinement in the annealing phase where the rest of the protein was allowed to be semi-flexible. All 200 structures then underwent explicit solvent refinement and water refinement, following which the structures were then ranked based on RMSD comparison, energy analysis, and restraint violations, as well as other factors, and the lowest energy ranked structure was chosen as the representative structure and further analyzed.

This lowest energy structure has the imidazole nitrogen in ketoconazole coordinated to the iron in the heme group as expected (Figure 15). The ketoconazole ligand orientation within the binding pocket of CYP101 was reproduced well in a way that's both consistent with the NMR data as well as other published ketoconazole structures with the chlorinated phenyl ring generally oriented towards the B' helix and sits in a hydrophobic pocket defined by Tyr96, Phe98 and Phe193 (Figure 16) The longer chain of ketoconazole orients in the opposite direction and is positioned in a pocket between the β5 and FG loop making very weak contacts with the basic side-chains from the B' helix and β5 loop via the oxygen and nitrogen atoms in the long chain. Apart from the lowest energy structure, several other structures with similar or slightly increased free energies generated during the run that are generally clustered around this lowest energy structure in the binding pocket of CYP101. The chlorinated phenyl ring and the longer chain were found to have various flipped orientations for the phenyl rings with slight displacement about the lowest energy structure. This ensemble of structures is generally consistent with the weak nature of interactions expected between the protein and ketoconazole ligand that has a predominant hydrophobic contribution based on NMR chemical shift perturbations and ITC data discussed in Chapter 6.

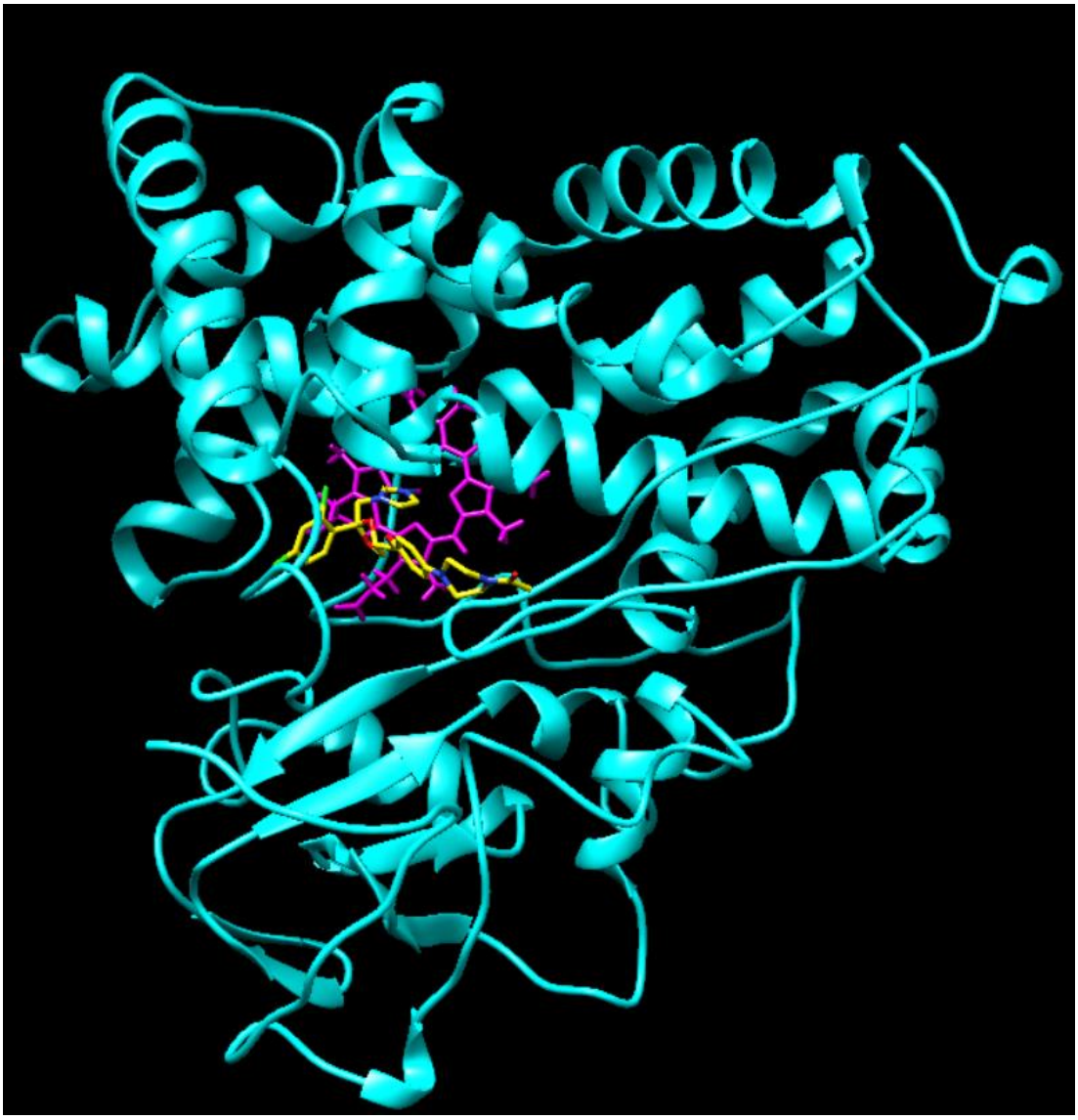


Figure 15: Docked structural model of ketoconazole-bound CYP101 generated with HADDOCK

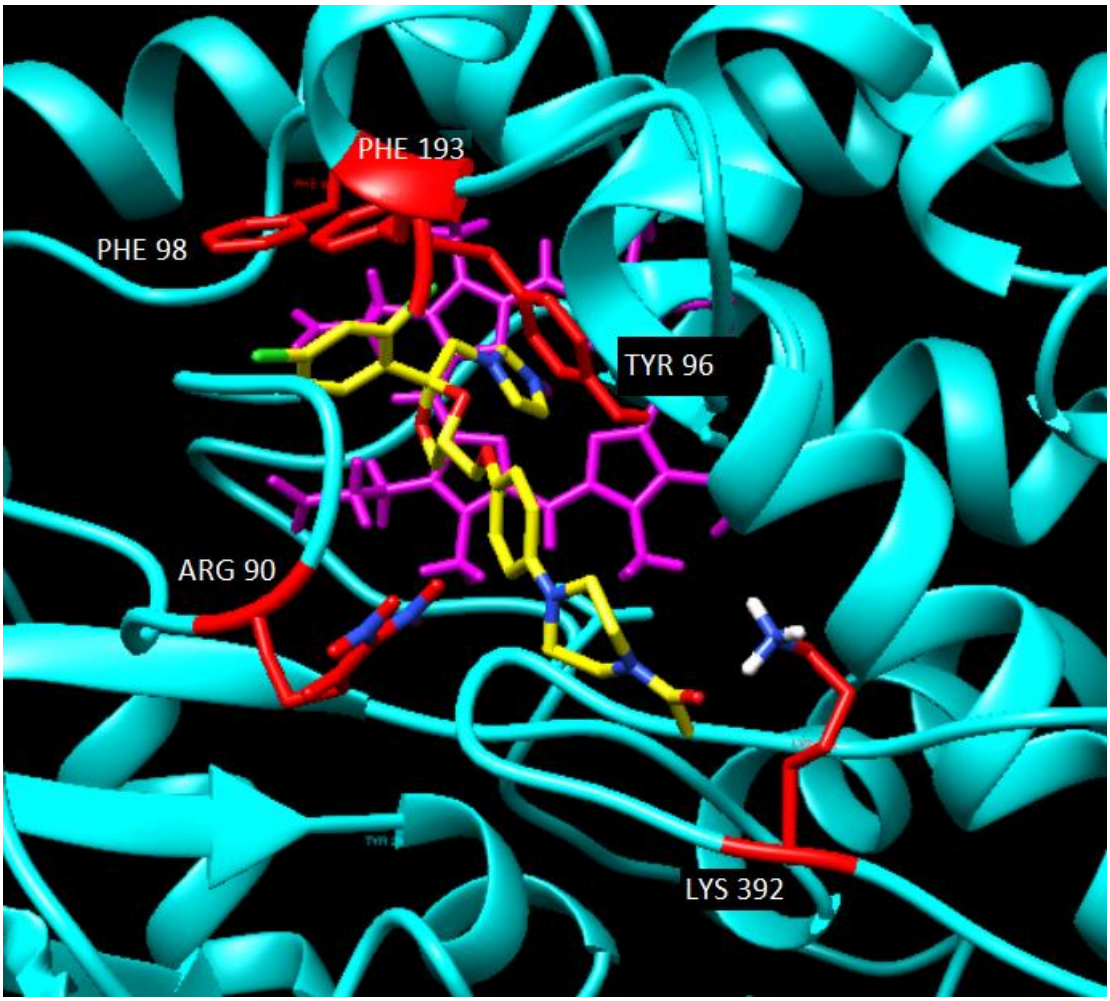


Figure 16: Structural view of docking site of ketoconazole in the SRS of CYP101, showing the interactions of chlorinated phenyl ring and long chain with specific residues in CYP101

Overall and residue-wise RMSD for the protein part was calculated between the ligand-free and lowest energy docked structure. The residue-wise RMSD is shown in a color-coded manner in Figure 17 and when compared to the NMR chemical shift perturbations for corresponding residues in the protein color-coded in a similar manner, show a good match between the two representations, lending confidence in the docked structure produced by HADDOCK. There are no major changes observed in protein structure outside of SRS as expected based on these regions not defined as flexible in HADDOCK and in line with the minimal NMR chemical shift perturbations for most protein regions outside of the SRS. The SRS on the other hand, shows large structural deviations in the new docked structure relative to the ligand-free structure, which is again in line from the large NMR chemical shift perturbations observed for these regions, especially the B' helix. This indicates that the docked structure of the complex reproduces characteristics seen in NMR data fairly well and can be used as a good structural model for dynamic interpretation of amide exchange data.

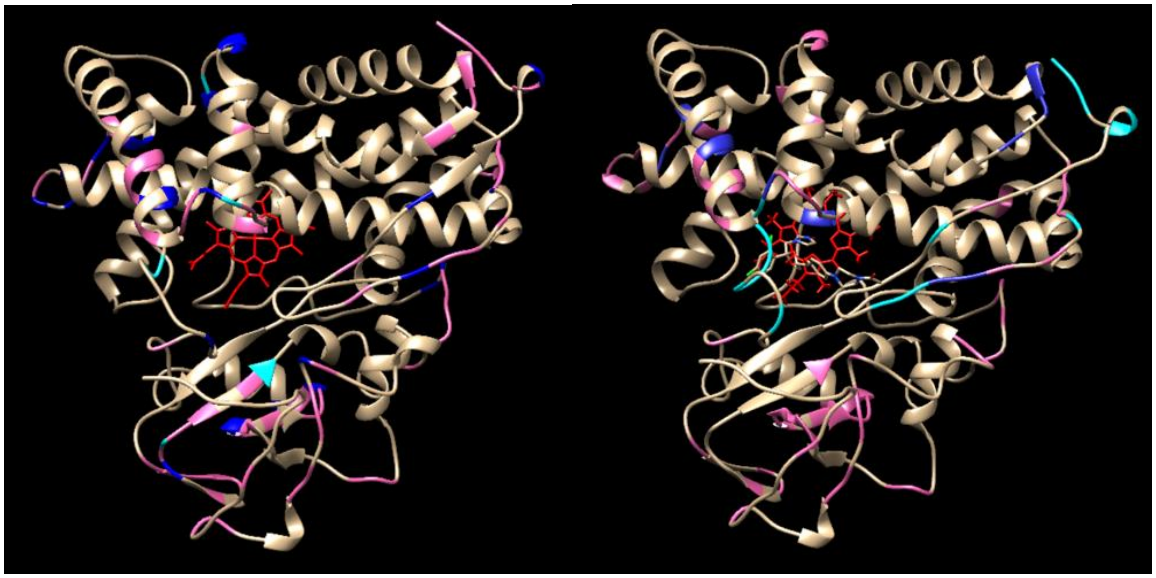


Figure 17: Comparison of NMR chemical shift perturbations between ligand-free and ketoconazole-bound CYP101 (left) and RMSD between ligand-free and docked ketoconazole-bound CYP101 structure (right). Color code for Chemical shift perturbation and RMSD difference: Small -Pink, Medium-Blue, Large-Cyan.

CHAPTER FIVE:
HYDROGEN-DEUTERIUM EXCHANGE EXPERIMENTS TO MONITOR LIGAND-DEPENDENT DYNAMIC CHANGES IN CYP101

5.1 Ligand-dependent dynamic changes in CYP101 from preliminary 2D NMR spectra

The 2D ^{1}H-^{15}N correlation spectra for each ligand at saturating levels with CYP101 reveals a different pattern of dynamics when compared to the other ligands. This can be visualized by looking at the changes in overall peak patterns (Figure 7, 8). Comparing the camphor-bound spectrum to the ligand-free spectrum, the peaks in ligand-free spectra are generally broader and show multiple peak splittings for several resonances in the spectra, which is not observed for the camphor-bound spectrum. This indicates that the protein has increased differential dynamics and is sampling more conformations in the absence of a ligand than when bound to camphor. This difference becomes even more pronounced when comparing the ligand-free spectrum to nicotine-bound spectrum. The nicotine-bound spectrum shows presence of sharp, similar intensity single peak patterns throughout the spectrum with very noticeable peak splittings for the vast majority of peaks indicating that the protein overall does not sample multiple conformations and is not as dynamic in presence of nicotine (Figure 18). The norcamphor-bound spectrum, on the other hand, shows an intermediate pattern, more dynamic than camphor or nicotine, but also not as dynamic as ligand-free or ketoconazole-bound spectra. The ketoconazole-bound spectrum looks most similar to the ligand-free spectrum and shows similar increased dynamics throughout the protein. These findings are surprising when considering ligand properties. For example, ketoconazole is a large size ligand with several polar groups and would be expected to bind the protein with multiple ligand-protein contacts based on its high affinity to

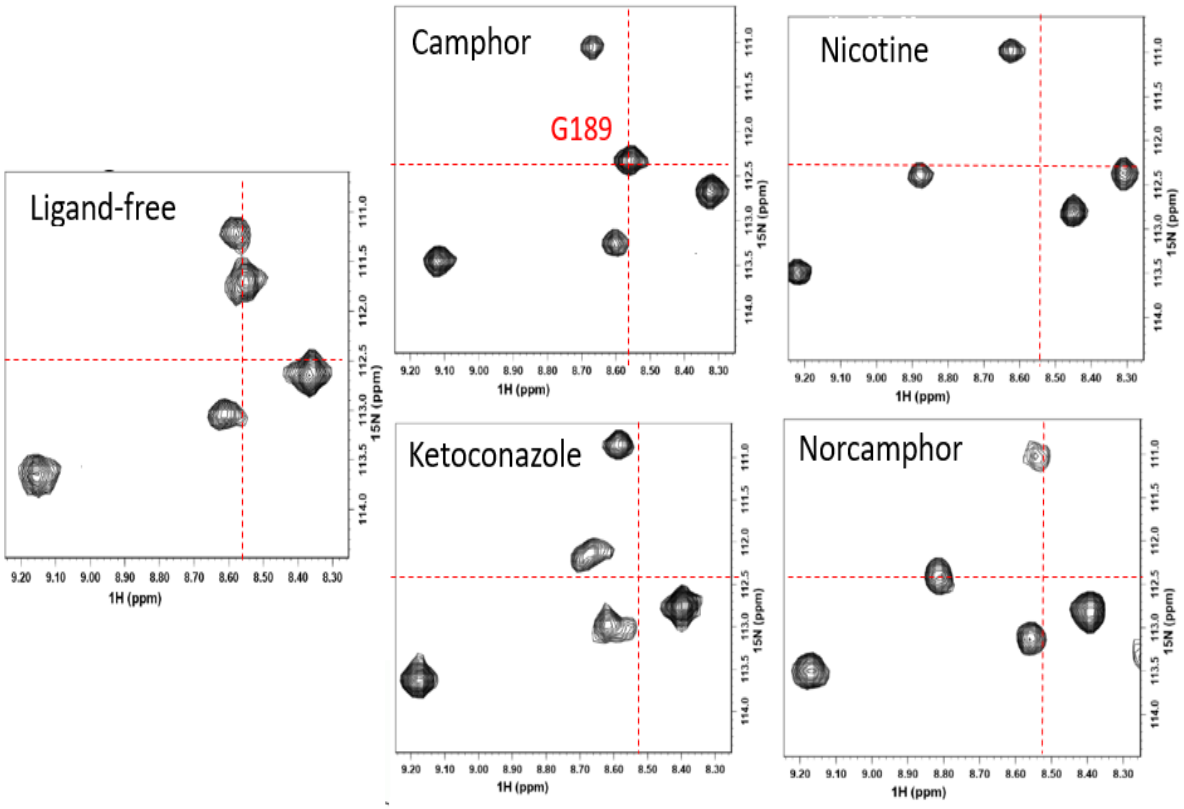


Figure 18: Zoomed in portions of 2D ^{1}H-^{15}N HSQC-TROSY of CYP101 in ligand-free and various ligand-bound forms, showing the different peak patterns in terms of chemical shift changes, line-broadening and peak splittings. The amino acid G189 in CYP101 is used as an example to illustrate these variations

CYP101, locking the protein down to a reduced set of dynamic conformations compared to ligand-free form. However, the opposite is observed where the dynamic pattern and number of conformations stays comparable to the ligand-free protein, in contrast to any of the other ligands examined. Similar discrepancy is observed in the case of nicotine, which is a small ligand. One would expect the ligand to not make any significant protein-ligand contacts given its size and allow the protein to freely sample multiple conformational states. However, the affinity of nicotine is surprisingly high and most of it is derived by directly coordinating with the heme. In spite of this, the ligand still reduces dynamics throughout the protein and is the least dynamic of all the ligand-bound forms examined here.

The main conclusion from the above observations is that dynamics for CYP101 can vary depending on which ligand is present, and not necessarily in the expected manner, with some inducing greater dynamic changes and some smaller. This is in contrast with how superimposable the camphor-, nicotine-, and norcamphor-bound crystal structures are, indicating that the crystal structures are not able to capture the multitude of conformations the protein can undergo when in solution. Although there are small differences observed in these structures in the SRS regions, virtually no structural or dynamic differences are detected in regions outside of SRS, which has led to the longstanding notion that only the structural flexibility of the SRS determines binding of a particular ligand. This is not supported by our NMR observations that the protein as a whole, responds to the binding of a ligand and this event leads to significant dynamic differences throughout the protein. A detailed investigation of ligand-dependent dynamics of CYP101 by solution NMR can help resolve the question as to what the basis of the dynamic differences observed in the NMR spectra with various ligands is and which regions are responsible for these dynamics. This is particularly important in light of previous observations that fast collective motions span the entire protein in CYP101, including regions remote from the SRS, and facilitate ligand binding via translating into slower motions[47]. However, this was demonstrated only for one ligand and it is not clear if recognition of other ligands follows a similar mechanism or is timescale dependent. An important first step would be to characterize the dynamic changes with different ligands and whether there is a distinct pattern for each ligand on

different timescales that would indicate whether only the SRS or the entire protein responds to the binding event.

There are different NMR experiments to characterize motions on different timescales. While the 2D ^{1}H-^{15}N NMR spectra indicate the presence of dynamic motions on all timescales, these only serve as a qualitative measure of protein dynamics that appear in a combined manner in these 2D spectra and the motions on different timescales need to be characterized separately to figure out which type of motions contribute to what extent to the dynamics of a protein. Motions on the slower timescales (s-min) are the easiest to characterize and utilize measurement of amide exchange rates that monitor slower breathing motions of the protein. These experiments require exchange of amide protons with deuterium and the determination of these exchange rates in a residue-dependent manner allows interpretation of opening and closing motions of the protein especially water access channels, facilitating exchange of interior amide protons. These motions not only report on the overall global flexibility of the protein but also report on motional changes in key residues in local regions of the protein. The global motions are usually linked to local functional motions of the protein in terms of a dynamic hierarchy, which permits drawing inference about some of the faster motions and coordination between different regions of the protein in facilitating ligand binding [3]. Dynamic differences in hydrogen exchange patterns also indicate rearrangement of hydrogen bonding patterns and different solvent exposure of protein areas, suggesting a likely role for water molecules in dynamics and thermodynamics of ligand binding. With these goals in mind, backbone amide exchange measurements were carried out for the ligand-free and 4 ligand-bound forms of CYP101 to map the dynamic differences in CYP101 on slow dynamic timescales in a ligand-dependent manner.

5.2 Measurement of Amide Exchange Rates for Various Forms of CYP101

Amide exchange (HDX) experiments can be used to quantify the slow time scale motions of the protein by measuring the peak intensities of backbone amide protons in a 2D ^{1}H-^{15}N correlation spectrum as a function of time when exposed to D_2O. In this experiment, the fully protonated lyophilized protein is dissolved in D_2O and the rate of

exchange of amide protons with deuterium is monitored over a period of time (Figure 19). The protons attached to carbon are not exchangeable due to the nonpolarity of the bond and thus their exchange rates cannot be measured. The amide protons exchange differentially based on various factors with the main ones being presence of hydrogen-bonding and increased or decreased dynamic motions which leads to different solvent exposure. This causes the peaks corresponding to these amide protons in the 2D spectrum to rapidly or slowly decrease in intensity at different rates. Identical 2D spectra can be measured at different time intervals from which peak intensities can be extracted and plotted as a function of time to determine exchange rates for individual amide protons.

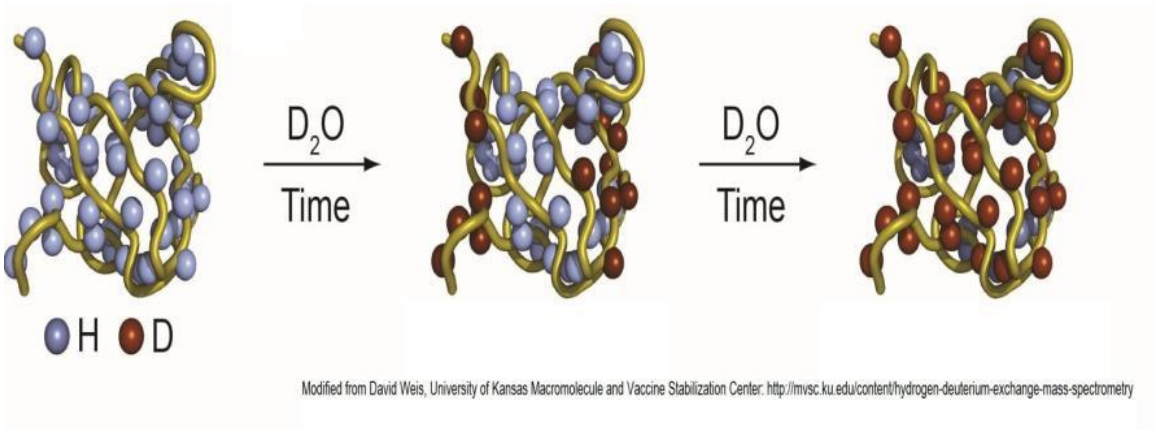


Figure 19: Scheme for Hydrogen-Deuterium Exchange experiment [67]

Depending on their exchange rates, the resonances can be categorized into three different groups: Fast, intermediate, or slow exchange. Fast exchange is typically designated by comparing the reference time point with no D_2O added to the time point at the end of the first collected spectrum. If the resonance has disappeared from the spectrum by that time point, it is considered to undergo fast exchange. In this case, the timepoint of the end of the first collected spectrum corresponded to approximately 28

minutes. Spectra were collected every 28 minutes defining the time interval of the timepoints. A resonance was considered to be undergoing intermediate exchange if it was visible after 28 minutes but not after approximately 16 hours and undergoing slow exchange if it remained visible after the 16th hour up to the final time point collected around 40 hours. 2D ^{1}H-^{15}N spectra for all 5 forms of CYP101 were acquired in an identical manner at identical time intervals and total time periods.

Cross-peak intensities and volumes in each spectra for all the assigned and visible peaks were measured with a standard integration routine in NMRViewJ and their time-course was fitted to equations 1 and 2 to calculate exchange rates as a function of time.

$$\ln I = \ln I_o - R_{ex}t \qquad \textbf{\textit{Eqn. 3}}$$

$$\ln V_0 = \ln V_0 - R_{ex}t \qquad \textbf{\textit{Eqn. 4}}$$

where I and V is the peak intensity, I_0 and V_0 is the initial peak intensity, R_{ex} is the exchange rate constant, and t is time of exchange.

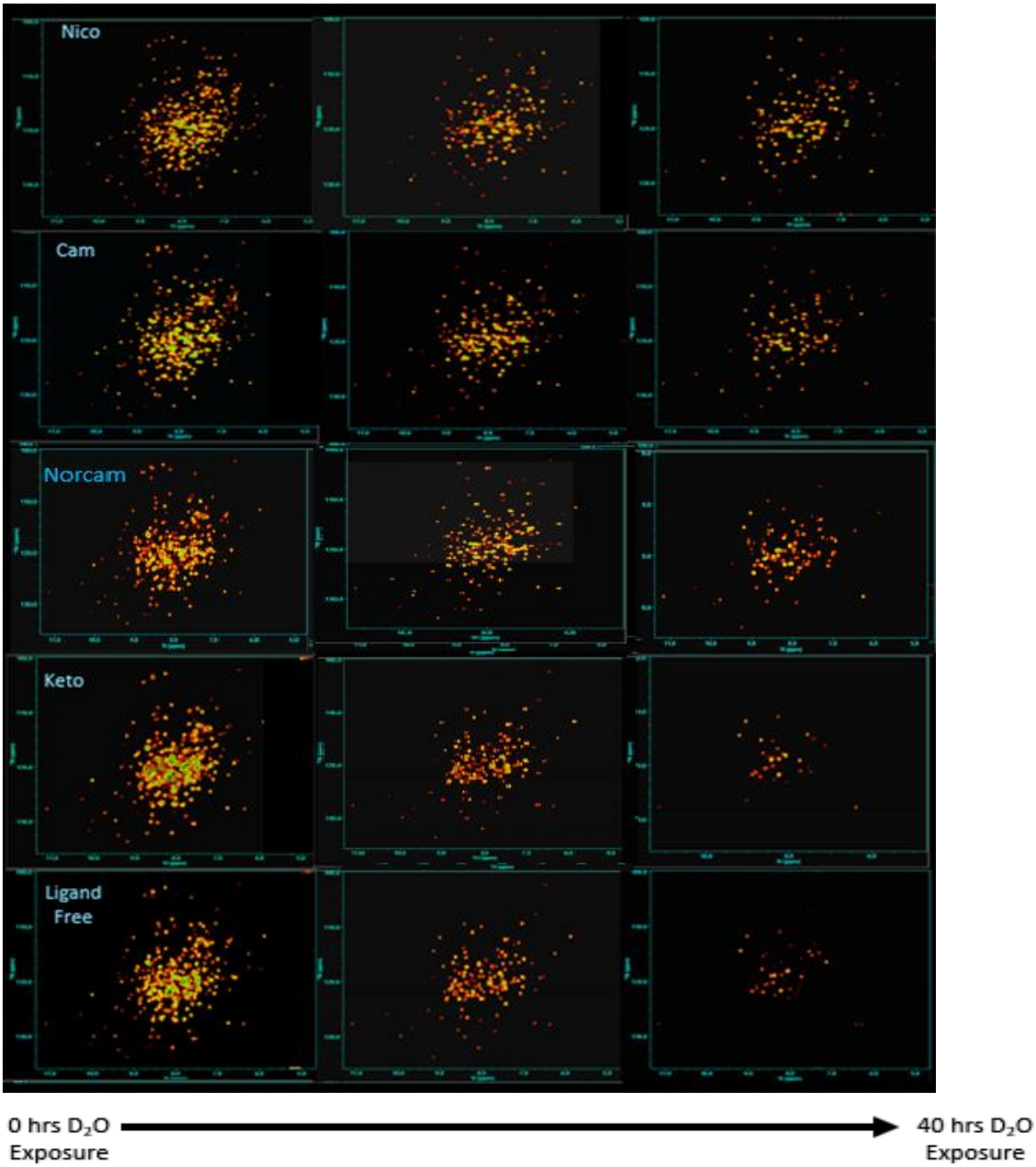


Figure 20: 2D ^{1}H-^{15}N HSQC spectra of CYP101 bound to various ligands showing amide exchange at different timepoints starting with no D_2O added at 0 hrs to 40 hours after D_2O exposure.

Table 3. HDX exchange categorization table (Actual exchange rate constants are reported in Table A.4/5 in Appendix section)

Type	Number	Secondary Structure	Camphor	Nicotine	Ligand Free	Ketoconazole	Norcamphor
LEU	11	N term Loop	Fast	Slow	Slow	Intermediate	Slow
ALA	12	N term Loop		Fast	Fast	Fast	Fast
HIS	17	N term Loop	Fast	Fast	Fast	Fast	Fast
VAL	18	N term Loop	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
LEU	22	N term Loop	Slow	Slow	Intermediate	Fast	--
VAL	23	N term Loop	Slow	Slow	Intermediate	Slow	Slow
PHE	24	N term Loop	Slow	Slow	Intermediate	Slow	Slow
PHE	26	**N term Loop**	Fast	Slow	Fast	Fast	Intermediate
ASP	27	**N term Loop**	--	Fast	Fast	Fast	Fast
TYR	29	**N term Loop**	Slow	Slow	Fast	Slow	Slow
ASN	30	**N term Loop**	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
ALA	36	N term Loop	Fast	Fast	Fast	Intermediate	Fast
GLY	37	A Helix	Slow	Slow	Fast	Intermediate	Fast
ASN	49	Aβ1 Loop	Fast	Slow	Slow	Slow	Slow
VAL	50	Aβ1 Loop	Intermediate	Slow	Intermediate	Intermediate	Fast
VAL	54	β1 Sheet	Slow	Slow	Intermediate	Slow	Slow
TRP	55	β1 Sheet	Slow	Slow	Intermediate	Intermediate	Slow
THR	56	β1 Sheet	Slow	Slow	Intermediate	Intermediate	Slow
CYS	58	**β1 Sheet**	Slow	Fast	Slow	Intermediate	Slow
ASN	59	**β1 Sheet**	Intermediate	Fast	Fast	Fast	Fast
GLY	60	**β1 Sheet**	Fast	Fast	Fast	Fast	Fast
GLY	61	**β1 Sheet**	Slow	Slow	Intermediate	Intermediate	Slow
HIS	62	β1 Sheet	Slow	Slow	Intermediate	Slow	Slow
ILE	64	β1 Sheet	Slow	Slow	Intermediate	Slow	Slow
ALA	65	β1 Sheet	Slow	Slow	Intermediate	Intermediate	Slow
THR	66	β1 Sheet	Slow	Slow	Intermediate	Slow	Slow
ARG	67	B Helix	Slow	Slow	Intermediate	Intermediate	Slow
GLY	68	B Helix	Fast	Fast	Fast	Fast	Fast
GLN	69	B Helix	Slow	Slow	Intermediate	Intermediate	Fast
LEU	70	B Helix	Fast	Fast	Fast	Fast	Fast
CYS	85	BB' Loop	Slow	Slow	Slow	Fast	Slow
PHE	87	BB' Loop	Intermediate	Slow	--	--	Fast
ILE	88	BB' Loop	Fast	Slow	--	--	--
ARG	90	**B' Helix**	Intermediate	Fast	Fast	Fast	Fast
GLU	91	**B' Helix**	Fast	Fast	Fast	--	Fast
ALA	92	**B' Helix**	Fast	Intermediate	--	--	Intermediate
GLY	93	**B' Helix**	Fast	Fast	Fast	Fast	Fast
GLU	94	**B' Helix**	Slow	Slow	Slow	Slow	Slow
ALA	95	**B' Helix**	Fast	Fast	Intermediate	Fast	Intermediate
TYR	96	**B' Helix**	Intermediate	Fast	Fast	Fast	Fast
PHE	98	**B'C Loop**	Fast	Slow	--	--	--
VAL	123	C Helix	Slow	Slow	Intermediate	Fast	Slow
VAL	124	C Helix	Fast	Slow	Fast	Intermediate	Slow
ASP	125	C Helix	Fast	Slow	--	Intermediate	Slow
LYS	126	C Helix	Slow	Fast	Intermediate	Intermediate	Slow
SER	141	D Helix	Intermediate	--	Fast	Fast	Fast
GLN	145	D Helix	Intermediate	Fast	Intermediate	Slow	Intermediate
GLY	146	DE Loop	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
GLN	147	DE Loop	Slow	Slow	Slow	Slow	Intermediate
CYS	148	DE Loop	Slow	Slow	Slow	Slow	Slow
ASN	149	E Helix	Fast	Fast	Intermediate	Intermediate	Fast
PHE	150	E Helix	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
THR	151	E Helix	Fast	Fast	Intermediate	Intermediate	Intermediate
GLU	152	E Helix	Fast	Intermediate	Intermediate	--	--
ASP	153	E Helix	Slow	Slow	Slow	Slow	Slow
LEU	165	E Helix	Slow	--	Intermediate	Slow	--
LEU	166	E Helix	Slow	Fast	--	Fast	--
ALA	167	E Helix	Slow	Slow	Intermediate	Slow	Slow
GLY	168	E Helix	Intermediate	Intermediate	Intermediate	Intermediate	Fast
GLU	171	EF Loop	Fast	Fast	Fast	Intermediate	Slow
GLU	172	EF Loop	Fast	Fast	Fast	Intermediate	Fast
LYS	178	**F Helix**	Slow	Fast	Intermediate	Intermediate	Intermediate
TYR	179	**F Helix**	Intermediate	Intermediate	--	--	Intermediate
THR	185	**F Helix**	Slow	Slow	Intermediate	Fast	Fast
ASP	188	**FG Loop**	Fast	Intermediate	Intermediate	Slow	Fast

Table 3 continued

GLY	189	**FG Loop**	Intermediate	Fast	Intermediate	Fast	Fast
SER	190	**FG Loop**	Fast	Slow	Slow	Intermediate	Slow
MET	191	**FG Loop**	Fast	Slow	Fast	Slow	Slow
THR	192	**G Helix**	Intermediate	Intermediate	Fast	Intermediate	Slow
PHE	193	**G Helix**	Fast	Fast	--	--	Slow
ALA	194	**G Helix**	Intermediate	Slow	Intermediate	Fast	Fast
LYS	197	**G Helix**	Slow	Slow	--	Intermediate	--
GLU	198	**G Helix**	Slow	Slow	Intermediate	Slow	Fast
ALA	199	**G Helix**	Slow	Slow	Fast	Fast	Fast
LEU	200	**G Helix**	Slow	Slow	Intermediate	Intermediate	Slow
ILE	207	G Helix	Fast	Intermediate	Intermediate	Intermediate	Intermediate
ILE	208	G Helix	Slow	Slow	Slow	Slow	Slow
LYS	214	G Helix	Slow	Slow	Fast	Intermediate	Intermediate
GLY	216	GH Loop	Fast	Fast	Fast	Fast	Fast
ALA	219	H Helix	Slow	Slow	--	--	--
ILE	220	H Helix	Slow	Slow	Fast	Fast	Fast
VAL	223	H Helix	Slow	Slow	Intermediate	Intermediate	Slow
ALA	224	H Helix	Slow	Slow	Intermediate	Intermediate	Fast
ASN	225	H Helix	Intermediate	--	--	--	Fast
GLY	226	β2 Sheet	Intermediate	Intermediate	Intermediate	Fast	Intermediate
GLN	227	β2 Sheet	Intermediate	Slow	Intermediate	Slow	Slow
VAL	228	β2 Sheet	Intermediate	Intermediate	Fast	Fast	Intermediate
ASN	229	β2 Sheet	Slow	Slow	--	--	--
GLY	230	β2 Sheet	Fast	Fast	Intermediate	Fast	Fast
ARG	231	β2 Sheet	Slow	Slow	Slow	Slow	Slow
ILE	233	β2 Sheet	Intermediate	Intermediate	Intermediate	Fast	Intermediate
THR	234	I Helix	Intermediate	Slow	Intermediate	Fast	Fast
ALA	265	I Helix	Slow	Fast	Slow	Intermediate	Slow
LYS	266	I Helix	Fast	Slow	Intermediate	Slow	Slow
SER	267	I Helix	Intermediate	Slow	Slow	Slow	Slow
ARG	271	J Helix	Slow	Fast	Slow	Slow	Slow
GLN	272	J Helix	Slow	Intermediate	Slow	Slow	Intermediate
GLU	273	J Helix	Fast	Fast	Fast	Fast	Slow
LEU	274	J Helix	Intermediate	Slow	Intermediate	Slow	Slow
GLU	279	JK Loop	Fast	Fast	Fast	Fast	Fast
ARG	280	K Helix	Intermediate	Intermediate	Intermediate	Slow	Intermediate
ASP	304	β3β4 Loop	Intermediate	Fast	Slow	Intermediate	Fast
TYR	305	β4 Sheet	Slow	Slow	Intermediate	Intermediate	Fast
GLU	306	β4 Sheet	Fast	Fast	Fast	Fast	Fast
PHE	307	β4 Sheet	Slow	Intermediate	Intermediate	Intermediate	Slow
HIS	308	β4 Sheet	Fast	Fast	Fast	Fast	Slow
GLY	309	β4 Sheet	Fast	Slow	Fast	Fast	Fast
VAL	310	β4 Sheet	Slow	Slow	Intermediate	Intermediate	Slow
GLN	311	β4 Sheet	Slow	Intermediate	Intermediate	Fast	Slow
LEU	312	β4 Sheet	Slow	Slow	Intermediate	Intermediate	Slow
LYS	313	β3β4 Loop	Slow	Slow	Intermediate	--	Slow
LYS	314	β3β4 Loop	Slow	Slow	Intermediate	Slow	Slow
GLY	315	β3 Sheet	Fast	Fast	Intermediate	Fast	Fast
ASP	316	β3 Sheet	Slow	Slow	Intermediate	Slow	Slow
GLN	317	β3 Sheet	Intermediate	Intermediate	Fast	Fast	Intermediate
LEU	319	β3 Sheet	Slow	Slow	Fast	Intermediate	Slow
LEU	320	β3 Sheet	Slow	Slow	Intermediate	Slow	Slow
ALA	333	β3L loop	Fast	Fast	Slow	Fast	Slow
MET	336	β3L loop	Fast	Fast	Intermediate	Intermediate	Intermediate
HIS	337	β3L loop	Slow	Fast	Fast	Fast	Fast
VAL	338	β3L loop	Slow	Fast	Intermediate	Fast	Fast
ASP	339	β3L loop	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
PHE	340	β3L loop	Fast	Fast	Fast	Fast	Fast
SER	341	β3L loop	Fast	Fast	Intermediate	Intermediate	Fast
ARG	342	β3L loop	Fast	Intermediate	Intermediate	Fast	Intermediate
VAL	369	L Helix	Fast	Intermediate	Fast	Fast	Fast
LYS	372	L Helix	Slow	Slow	Slow	Fast	Fast
ASP	380	Lβ5 Loop	Intermediate	Intermediate	Intermediate	Slow	Intermediate
PHE	381	Lβ5 Loop	--	Fast	Intermediate	Intermediate	Intermediate
SER	382	β5 Sheet	Slow	Slow	Slow	Slow	Slow
ILE	383	β5 Sheet	Slow	Slow	Slow	Slow	Intermediate
ALA	384	β5 Sheet	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
GLY	386	β5 Sheet	Fast	Slow	Fast	Fast	Fast
ALA	387	β5 Sheet	Fast	Intermediate	Intermediate	Fast	Fast
GLN	388	β5 Sheet	Intermediate	Intermediate	Fast	Slow	Intermediate
ILE	389	β5 Sheet	Fast	Fast	Intermediate	Fast	Fast

Table 3 continued

GLN	390	β5 Sheet	Fast	Fast	Fast	Fast	Fast
HIS	391	β5 Sheet	Fast	Fast	Fast	Fast	Fast
LYS	392	β5 Sheet	Fast	Slow	Intermediate	Slow	Slow
SER	393	β5 Sheet	Slow	Slow	Intermediate	Slow	Intermediate
GLY	394	β5 Sheet	Slow	Slow	Intermediate	Intermediate	Intermediate
VAL	396	β5 Sheet	Intermediate	Intermediate	--	--	--
GLY	398	β5 Sheet	Slow	Slow	Intermediate	Intermediate	Intermediate
VAL	399	β5 Sheet	Intermediate	Fast	--	--	--
GLN	400	β5 Sheet	Slow	Slow	Slow	Slow	Slow
ALA	401	β5 Sheet	Intermediate	Fast	--	Slow	Fast
LEU	402	β5 Sheet	Slow	Slow	Slow	Slow	Slow
VAL	405	β5 Sheet	Slow	Slow	Slow	Slow	Slow
ALA	409	C term Loop	Slow	Fast	Intermediate	Slow	Slow
THR	410	C term Loop	Fast	Fast	Fast	Fast	Fast
LYS	412	C term Loop	Fast	Slow	Slow	Fast	Slow
ALA	413	C term Loop	Fast	Slow	Slow	Intermediate	Fast
VAL	414	C term Loop	Fast	Fast	Slow	Intermediate	Fast
		Total	152	149	138	139	140
		Fast	50	38	40	49	51
		Intermediate	36	37	72	50	31
		Slow	66	76	26	40	59

5.3 Ligand-dependent dynamic changes in the substrate recognition regions of the protein structure

The amide exchange experiments show that there are significant differences in overall slow timescale dynamics of the protein in a ligand dependent manner (Figure 20). Comparing the endpoint of the exchange process for various CYP101 forms after 40 hours of D_2O exposure, it is clear that the number of resonances remaining in the spectra varies greatly based on the ligand. The nicotine-bound protein spectrum shows the largest number of peaks remaining after 40 hours followed closely by camphor-bound protein, indicating an overall slower exchange pattern compared to the other forms. A large fraction of the peaks (~50%) are still visible after 40 hrs of exchange. This is in contrast to the camphor-, ligand-free, ketoconazole- or norcamphor-bound spectra, which only have approximately 20-40% of the resonances still visible depending on the ligand. The intermediate exchange regime follows a similar trend with only about 25% of peaks showing intermediate exchange for nicotine, camphor and norcamphor, whereas the other forms show a much higher percentage (40-50%) of intermediate exchange. All CYP101 forms show a similar fast exchange trend with between 25-35% of the peaks undergoing fast exchange indicating that the fast exchange results mostly from direct solvent-exposed residues in all forms.

Table 4. Overall amide exchange type percentages for assigned peaks in various CYP101 forms

Exchange type	Camphor	Nicotine	Ligand Free	Ketoconazole	Norcamphor
Fast	32%	25%	28%	35%	36%
Intermediate	25%	25%	52%	36%	28%
Slow	43%	50%	20%	29%	36%

While these exchange percentages were determined considering the protein as a whole, when the exchange patterns are considered region-wise such as those for SRS for example, there are considerable differences between the various CYP101 forms (Figure 21). While the exchange patterns are largely similar for the B' helix in all CYP101 forms ranging from intermediate to fast exchange, the F-G loop/helices and the β5 sheet show considerable differences. The similar exchange pattern for the B' helix is not surprising since the helix is mostly solvent exposed and although the crystal structures point to more disorder in the ligand-free form than the camphor-, norcamphor- and nicotine-bound forms where the helix stays largely ordered, the increased exposure to solvent predisposes this helical region to intermediate to fast exchange. Thus, no specific inferences can be made from exchange patterns in this region although it is one of the most dynamical regions of the protein. The region preceding the B' Helix (residues 80-90), which is shielded somewhat better from the solvent compared to the helix itself, however, can give some insights in the helix's dynamics. Comparison of exchange patterns in this region shows that the while mostly slow exchange is observed for this region in the nicotine-bound form, a combination of fast and intermediate exchange is observed for the camphor-bound and norcamphor-bound forms, whereas predominantly fast exchange is observed for the ketoconazole-bound and ligand-free forms. This suggests that the B' helix is likely more dynamic in forms like ketoconazole-bound and ligand-free CYP101, which is in line with the major chemical shift perturbations and more dynamic peak pattern changes from 2D NMR spectra for these two forms compared to other CYP101 forms.

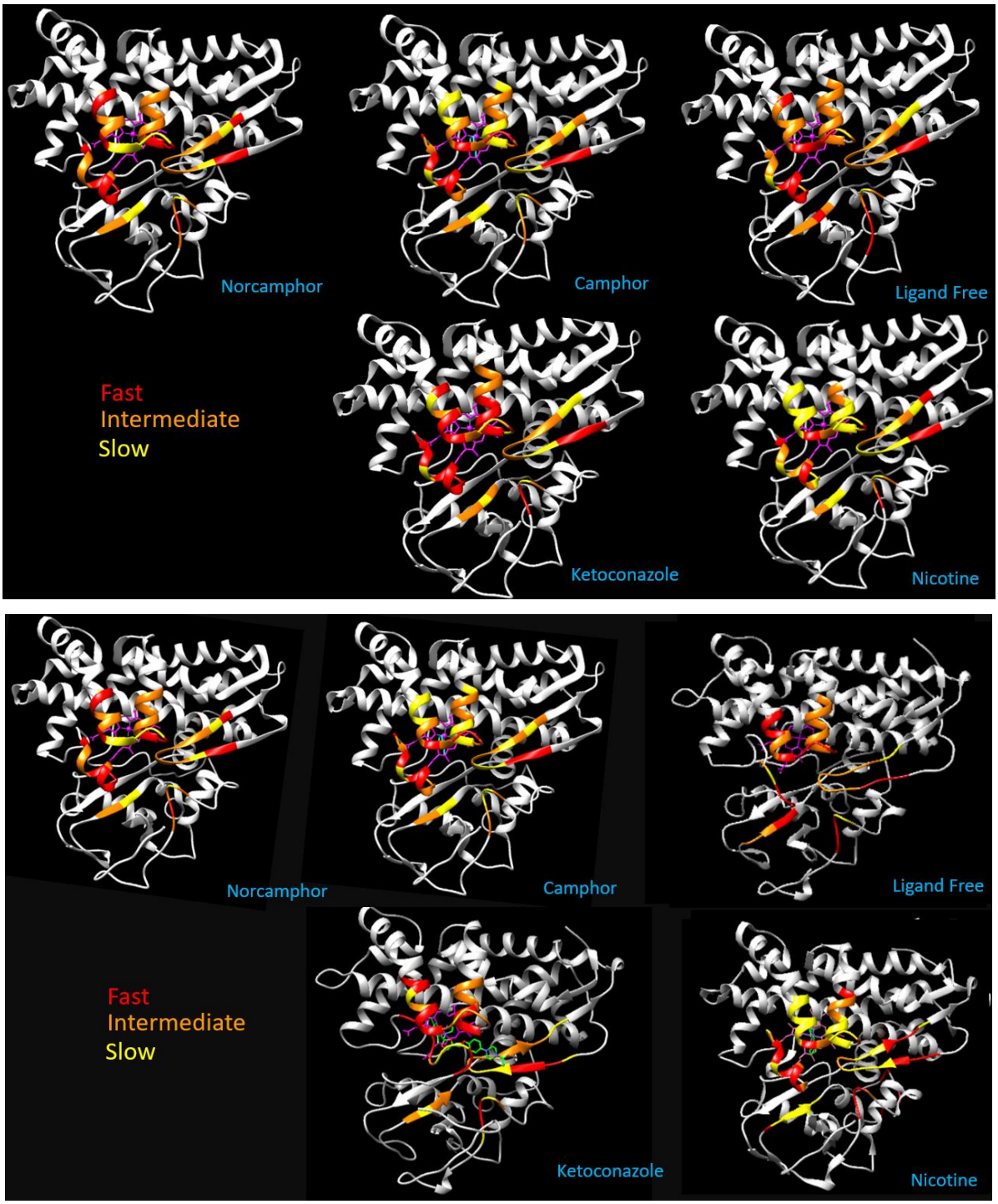


Figure 21: Comparison of the SRS amide exchange patterns for the various CYP101 forms on the same structure (2CPP) (top) and on their various crystal/docked structures (bottom)

The F-G loop formed as an intervening segment by the F and G helices is one of the most flexible regions in all P450s and is known to adapt to presence of different ligands in the binding pocket by changing conformations based on the numerous crystal structures published for different P450s so far. In CYP101, the F-G loop is observed to move by as much as 9-10 Angstroms distance to convert between the open (ligand-free) and closed (ligand-bound) conformations upon binding of certain ligands such as camphor and nicotine with slightly smaller displacement for norcamphor. The exchange pattern for the F-G loop matches this conformational tendency in the crystal structures where the loop exhibits predominantly slow exchange in the nicotine-bound form, while in the camphor- and norcamphor-bound forms the loop exhibits a mix of slow and intermediate exchange. In contrast, the F-G loop in ligand-free form is observed mostly as intermediate exchange, which suggests it is sampling between the open and closed forms as has been suggested from previous structural and dynamic studies[68]. The most surprising pattern of dynamics for this loop is from the ketoconazole-bound form, which shows it as a mix of slow and intermediate exchange. This is likely due to the fact that the placement of the bulky ligand such as ketoconazole in the binding pocket hinders solvent access to this region. This would also explain the mostly open conformation of F-G loop in the ketoconazole-bound form since it is unable to move freely down to wrap around and close the binding pocket like it does in case of other ligands, which makes it less flexible relative to ligand-free form even though it is an open conformation like the ligand-free form.

Like the F-G loop, the F and G helices themselves also play a role in binding of ligands as they move in tandem with the loop. The regions of the helix close to the F-G loop show predominantly slow exchange for the nicotine-bound form matching what is observed for the F-G loop. The camphor-bound form shows a mix of slow and intermediate exchange for the helices, implying it is more dynamic than the nicotine-bound form, but still favoring the closed conformation. The norcamphor-bound form is even more dynamic with the helices showing mostly intermediate exchange. In contrast, the ligand-free form shows a mix of intermediate and fast exchange which matches what is seen for the F-G loop. The ketoconazole-bound form exhibits the highest dynamics of all the forms with mostly fast exchange for the F helix and a mix of slow,

intermediate and fast exchange for the G helix. This suggests that the F and G helices are moving more rapidly in the ketoconazole form in tandem with the F-G loop even though they are bound to a large ligand. Based on the docked model structure for the CYP101-ketoconazole complex, the F-G loop does not dip far down to make any interactions with the ligand, but the sidechain of Phe193 is part of the hydrophobic pocket formed by several aromatic rings that surrounds the chlorinated phenyl ring and thus restricts the motion of the G helix partially, explaining the dynamic behavior of this helix.

The β5 region (381-404) largely shows a mix of exchange patterns for all CYP101 forms. The nicotine-bound protein shows the slowest exchange of all forms followed closely by the camphor-bound whereas the ligand-free, ketoconazole and norcamphor-bound forms show higher rates of exchange. Like the F-G loop, this region of the SRS is also known to undergo a conformational change contributing to the switch from the open to closed conformation upon ligand binding. Since the protein is observed predominantly in the closed conformation for nicotine-bound and camphor-bound forms, solvent access is likely restricted to this region in these forms explaining the observed slow exchange. In contrast, the ligand-free, norcamphor-bound and the ketoconazole-bound forms sample more of the open conformations allowing better solvent access and hence more intermediate to fast exchange is observed for these forms.

Finally, the last flexible region of the SRS is a short segment of the N-terminal loop (25-31) which lies at the bottom of the binding pocket. This region also shows variable amide exchange properties depending on the ligand. The exchange in this region is slowest for the nicotine-bound form of the protein but is mostly intermediate for the norcamphor and camphor-bound forms, whereas fast exchange is observed primarily for the ketoconazole-bound and ligand-free forms of the protein. This is in line with what is seen for the other regions of the SRS, where restriction to solvent access due to being more in closed conformation versus an open conformation will lead to the observed changes for the specific forms.

In summary, the SRS is most dynamic in the ligand-free, ketoconazole-bound, followed by the norcamphor-bound form which is less dynamic, while the camphor-bound and nicotine-bound forms show the slowest dynamics. This points to the

presence of differential dynamics within these regions depending on the ligand that may have an important role in how CYP101 is able to recognize diverse ligands, and it will be interesting to see if a similar pattern can be seen in the fast timescales as well.

5.4 Amide Exchange results show ligand-dependent dynamic changes in the regions not associated with ligand binding

While there are considerable dynamic differences observed for the flexible SRS regions, what is surprising is that noticeable ligand-dependent dynamic differences between the various forms are also observed for regions outside of the SRS (Figure 22). In particular, the C helix shows different dynamics between the various forms with nicotine-bound form again showing the slowest exchange followed closely by camphor-bound form whereas the ligand-free, norcamphor-bound and ketoconazole-bound forms showing more a mix of slow, fast and intermediate exchange. The redox partner of CYP101, Pdx, is known to bind to this region and has a structural effect on CYP101 upon binding. Pdx binding forces CYP101 into a more open conformation than when bound to camphor, so it makes sense that the dynamics would also be modulated for this region [69].

The I helix also shows differences in dynamics in a ligand-dependent manner. This is the helix that contains the active site residues around the heme and NMR resonances can only be observed for monitoring of dynamics on the edges of the helix due to the proximity to the paramagnetic heme. The two sides of the helix show opposite trends in exchange patterns with nicotine- and camphor-bound forms showing slow to intermediate exchange whereas norcamphor-bound, ligand-free and ketoconazole-bound forms showing intermediate exchange to fast exchange on one side (residues 233-240) which then reverses on the other side (residues 260-268) with all three forms slowing down the exchange considerably but the other forms increasing

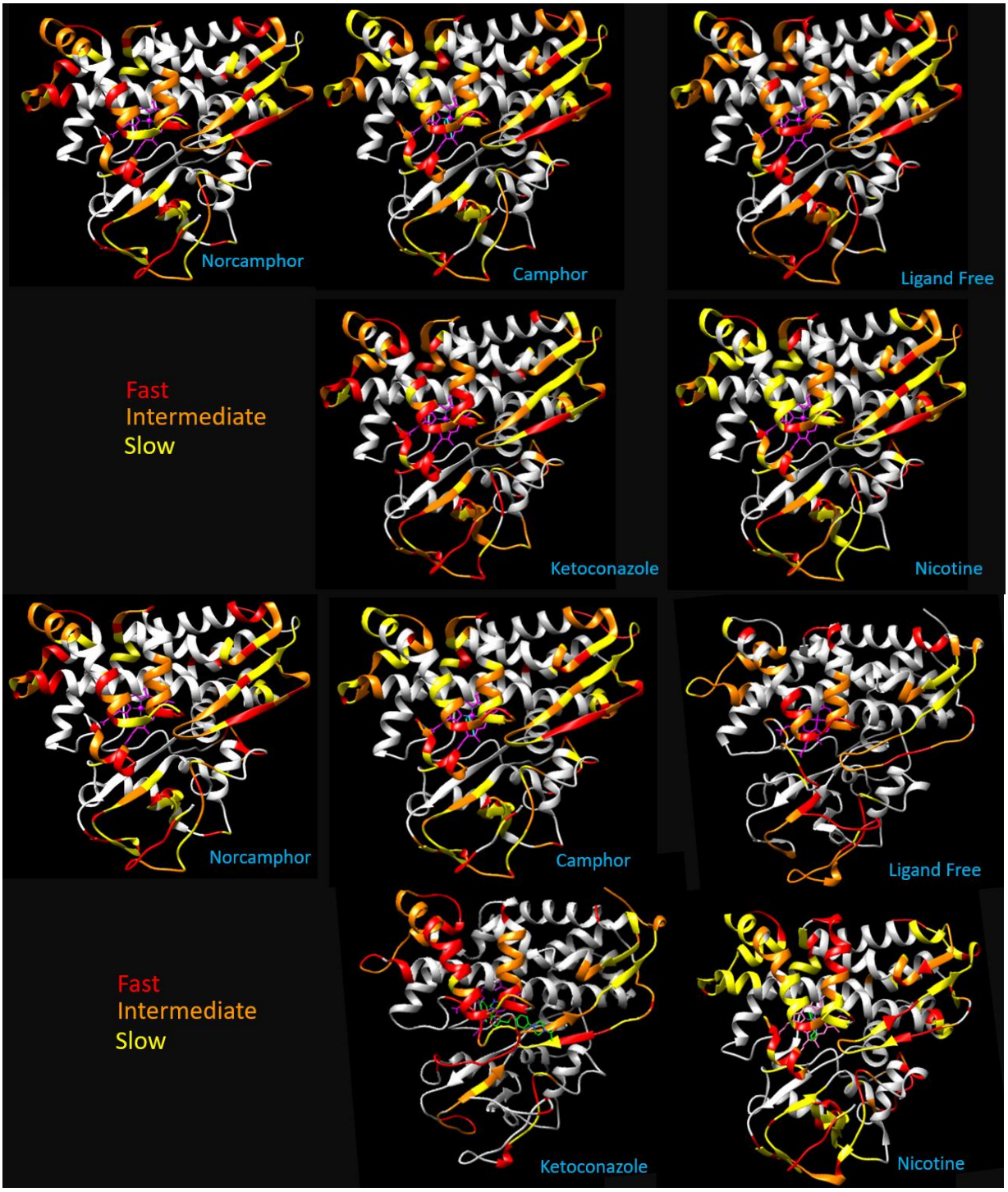


Figure 22: Comparison of the overall amide exchange patterns for the various CYP101 forms on the same structure (2CPP) (top) and on their various crystal/docked structures (bottom)

their exchange rates. The β2 sheet, which leads into the I helix, doesn't show any major changes between ligands. This indicates that the I helix responds differently to different ligands, which might be due to different buckling of the structure in the middle kink portion of the helix as a response to different ligands that would cause the dynamics at the ends of the helix to change.

Another interesting region that shows dynamic differences based on the ligand is the C terminus, which makes up the top right corner of the protein. Surprisingly, fast exchange is observed for this region for the camphor-bound form, but slow exchange is observed in the ligand-free protein, a reversal of the usual trend. The trends for the nicotine-bound, ketoconazole-bound and norcamphor-bound forms still stay the same though as observed for the other regions. The region for this is not clear, but it is possible that binding of different ligands changes the structure for this region in different ways relative to the ligand-free form, which is then reflected also in the change in dynamics for these forms.

Other regions of the protein such as the E helix, J helix and the loops in the protein leading in to and out of some of the important secondary structure elements in the protein show no significant changes between ligand free and the ligand bound forms. They all show an exchange pattern comprised of a mix of fast and slow exchange similar to the ligand-free form, implying that the dynamics in these regions do not contribute significantly to ligand binding. One notable exception is the loop region between the G and H helices on the top left corner of the structure which interestingly shows slow exchange for nicotine-bound and camphor-bound forms but remains as intermediate to fast exchange for the other forms. Recent studies with camphor-bound form has shown that this region might provide a second weak binding site for camphor [69] which would restrict the dynamics and solvent access for this region leading to the observation of slow exchange. There is no evidence currently of nicotine also binding to this secondary site, but the exchange could be slow for the nicotine-bound form generally in line with other regions of the protein even without binding. The other ligands, norcamphor and ketoconazole are likely not binding to this site as well on the basis of their exchange rates.

The dynamic differences observed for regions outside the SRS suggests that the dynamics of these regions may help modulate substrate binding as well, something which was suggested by a recent study using quasielastic neutron scattering experiments, which indicated opening and closing motions of the substrate access channel that are controlled by correlated interdomain motions of regions away from the access channel [47]. These domain motions have their basis in fast, collective motions that pervade throughout the protein and the slower timescale motions characterized by the amide exchange process may be a manifestation of these. If the fast, collective motions do indeed change depending on the ligand, they would be reflected in dynamic differences in the exchange rates of these regions and would explain the role of dynamics in facilitating ligand access to the active site. This is an intriguing possibility and it remains to be seen whether similar trends hold on a region-wise basis even on other timescales. If so, that would support our hypothesis of dynamic-modulated ligand recognition in CYP101 and possibly other P450s, that can be controlled in an allosteric manner.

CHAPTER SIX: THERMODYNAMICS OF CYP101-LIGAND INTERACTIONS

The dynamic differences observed in different CYP101 forms indicates that the mode of interaction of each ligand with the protein is different and may require distinct protein-ligand contacts to achieve the required affinity. All ligands, except norcamphor, display similar high affinity for CYP101 however are very different in their physico-chemical properties. Thus, the thermodynamic basis of binding may be rooted in the distinct nature of protein-ligand contacts. It would be interesting to investigate whether protein dynamics contributes to the binding and whether there is any thermodynamic benefit to modulating the dynamics of CYP101.

The enthalpic (ΔH) and entropic (ΔS) components of the free energy of binding (ΔG) can be quite informative in elucidating the driving force for protein-ligand interactions. A robust way of determining these components in protein-ligand binding is isothermal calorimetry (ITC). ITC measures the heat input or output of a binding reaction at a constant temperature by titrating the protein with ligand, giving a direct and accurate value for the ΔH of the association. ΔG can be found independently by fitting the titration curve which gives the affinity and stoichiometry of binding. ΔS can then be calculated from the Gibbs equation at that temperature giving a full thermodynamic representation of the system. ITC measurements were carried out for the various CYP101-ligand forms in order to understand the thermodynamic driving force for association of these ligands with oxidized CYP101 and the role of protein dynamics in this process.

6.1 Isothermal Calorimetry of CYP101-ligand complexes

Titration experiments were carried out for the protein with all 4 ligands as part of an ITC setup to calculate the ΔH, binding affinity and ΔS for each protein-ligand association. The ITC experiments were performed on a Malvern MicroCal VP-ITC instrument. Ligands (camphor, nicotine and norcamphor) were prepared as stock solutions in filtered potassium phosphate buffer and were diluted appropriately in the same buffer as protein to ensure complete saturation of the protein based on the

expected binding affinity (Table 5). A reverse titration of ketoconazole was carried out with the protein, since ketoconazole tends to aggregate at high concentrations. A stock solution of ketoconazole in DMSO solvent was initially made and then this stock ketoconazole solution was diluted in buffer to reach the appropriate concentration (Table 5) which contained only 1% DMSO in the end. Both ligand and protein samples were filtered and degassed prior to the experiment. The cell and syringe were both thoroughly cleaned and rinsed with filtered and degassed buffer to ensure no solvent mismatches. All titrations were done at 37 ^{O}C to maintain consistency with the NMR experiments.

Table 5 Summary of ITC experiments

Ligand	Protein [mM]	Ligand [mM]	Kd	Δ G	Δ H	TΔS
Cam	.03	.5	1.98 uM +/- 1.12	-7.72	-2.72 +/- 0.28	-4.96
Norcam	.04	1	176 uM +/- 120	-5.10	+0.69 +/- 0.6	-5.79
Nico	.02	.1	.156 uM +/- 0.1	-9.25	-11.95 +/- 0.5	+2.70
Keto	.07	.01	.566 uM +/- 0.31	-8.50	-4.32 +/- 0.52	-4.18

The protein and ligand were allowed 30 minutes to equilibrate to the temperature before the first injection. For the first injection, only 2 μL of titrant was added, and another 27 injections with 10 μL each of titrant were made with 4 minute spacing between injections, totaling 272 μL volume added to the cell whose volume was 1.45 mL. The protein-ligand mix in the cell was stirred at a speed of 304 rpm. The camphor and norcamphor titrations included the reducing agent TCEP at a concentration of 0.1

mM in the buffer to prevent protein degradation during the course of the experiment (~2.5 hrs).

The raw data sets were processed using NITPIC [70]. Reference titrations were performed by injection of each ligand into the appropriate buffer to account for the heat of dilution in the protein-ligand titration. The baseline was manually adjusted after the heat of dilution subtraction and initial automated baseline correction to ensure the best fit for the peak shapes. The isotherm was then saved and analyzed further in SEDPHAT for fitting and extraction of thermodynamic parameters.[71] The isotherms were fitted using the A + B → AB Hetero-Association configuration assuming a one-site binding model. After removing the first injection data point and other data points that were more than 2 standard deviations off the line likely due to air bubbles, the isotherm curve was best fit by a combination of Marquardt nonlinear least-squares analysis and Simplex fitting until both models converged. This converged fit was then used to calculate the binding affinity in terms of K_d and enthalpy of binding (ΔH) for each ligand-protein association, from which the entropy of binding was then derived (Table 5). Final heat of injection, residual and isotherm plots were generated with GUSSI [72].

6.2 Thermodynamic parameters of CYP101-ligand complexes

Binding of ligands with diverse physico-chemical properties to oxidized CYP101 was investigated by ITC to examine the thermodynamic basis of differential ligand binding. Analysis of thermodynamic parameters obtained from the ITC measurements reveal an interesting pattern of ligand binding to CYP101 (Table 5). Camphor and norcamphor bind with significantly different affinities and with K_d values similar to what was observed in spectroscopic measurements [55]. The norcamphor binding is much weaker than that of camphor, which is expected since norcamphor is lacking the hydrogen-bond with Tyr96 side-chain that camphor is known to form and also two methyl groups present on camphor that might reduce the extent of hydrophobic contacts formed by camphor in the binding pocket of CYP101. This would explain the slightly exothermic nature of camphor binding versus that of norcamphor which is slightly endothermic in nature. However, the entropy change is more favorable for the binding process of both ligands and could result from solvent reorganization effects for both

protein and ligand. The ligand-free form of CYP101 has several bound water molecules lining the hydrophobic interior of the binding pocket [69]. One of the water molecules is ligated to the sixth coordination site on the heme and is displaced by both camphor and norcamphor binding as part of the high spin transition of the heme active site upon ligand binding. This along with possible release and/or reorganization of other water molecules in the binding pocket could possibly result in a more favorable entropy for the binding process. Since norcamphor binds weakly to CYP101 mainly due to lack of hydrogen-bond to Tyr96 of the protein, the flexibility of the protein is not reduced as much as that observed for camphor which would explain the higher average amide exchange rates for norcamphor-bound form relative to the camphor-bound form.

In contrast, ketoconazole binding is a mix of enthalpy and entropy driven process. The docked model structure of the ketoconazole-CYP101 complex supports this observation. The enthalpy change upon ketoconazole binding is largely resultant from the interaction of the imidazole nitrogen with the heme iron. Although weak contacts between the polar oxygen and nitrogen atoms of ketoconazole with surrounding polar side-chains in CYP101 can contribute to this enthalpy change, they are not going to be significant based on the conformational mobility of the long chain in the ketoconazole structure as observed in the docked model. The weak interaction of the chlorinated phenyl ring with the phenyl side-chains of CYP101 in the hydrophobic interior again may not contribute significantly to the enthalpy change, however may actually contribute to the entropy change again due to the release and/or reorganization of water molecules in the interior of the protein. Also, ligand desolvation upon binding could contribute to this effect. The observation from NMR chemical shift perturbation and the docked structure that ketoconazole does not make any specific contacts since the protein is largely found in a conformation similar to the ligand-free form suggests that binding of large ligands such as ketoconazole does not require major change in CYP101 conformation to accomplish tight binding and can instead utilize minimal enthalpic interactions and favorable entropy from solvent reorganization for that purpose. Similar ITC experiments with other P450s have similarly shown that ketoconazole binding is a mix of enthalpy and entropy changes to varying degrees with entropy generally being the more dominant driving force [64].

Finally, the binding of nicotine is largely driven by enthalpy change due to ligation of nicotine imidazole nitrogen with heme iron and hydrogen-bonding interaction with surrounding water molecules. The nicotine ligand is a small polar molecule and thus not displace the water molecules significantly from the hydrophobic interior of the protein but instead binds to them, which would account for the unfavorable entropy of binding observed in the ITC experiments.

An interesting observation made from the ITC experiments is the similar binding affinities for all ligands other than norcamphor, which indicates the free energy of binding (ΔG) remains largely unchanged for these ligands even though relative contribution of enthalpy and entropy vary, suggesting an enthalpy-entropy compensation mechanism. This mechanism is generally explained by assuming that if a change in the ligand causes tighter van der Waals contacts or H-bonds, it inevitably leads to reduced flexibility in one or both of the components reducing the overall conformational entropy which compensates for the enthalpy increase. Vice-versa, if the overall conformational entropy increases, it results in increased conformational flexibility with a concomitant reduction in enthalpy change. This mechanism has typically been invoked to explain change in conformational dynamics of the protein. In the case of camphor and norcamphor, an increase in overall entropy is observed which is also the driving force for the association of these ligands with the protein. However, amide exchange studies show that the conformational dynamics of the protein are actually reduced overall relative to the ligand-free form. Conformational entropy of the protein therefore is not likely the major factor in this observed increase in entropy for the system. Another factor that can change the entropy of the system is the water hydrating the system and if the ligand-protein interaction changes the amount of bound water either on the ligand or the protein, that could contribute to the compensatory mechanism [73] In case of camphor and norcamphor, they are both hydrophobic ligands with a predominantly non-polar character and thus their interaction with the hydrophobic interior of the protein can cause release of bound water changing the entropy of the system. The dynamics of the protein can still be altered in this case but in the opposite direction as release of bound water could change the dynamic landscape of the protein where the protein samples a more restricted conformation due to a tighter

rearrangement of protein-ligand contacts in the interior of the protein (e.g. nonpolar contacts), assuming the entropy of the ligand bound to the protein has not changed significantly relative to its free state. In case of camphor, the formation of hydrogen bond between camphor oxygen atom and Tyr96 side-chain likely adds to this restriction on flexibility. This hydrogen-bond is not seen in norcamphor-bound CYP101 explaining its slightly higher flexibility relative to camphor-bound form.

The binding of nicotine to CYP101 also supports this explanation of role of water in the enthalpy-entropy compensation mechanism observed in CYP101. The crystal structure of nicotine bound to CYP101 apart from directly binding to the heme also shows direct interaction of nicotine with multiple crystallographic waters [Figure 5] that is not observed in the case of camphor or norcamphor. This in principle should considerably reduce the flexibility of nicotine-bound protein and decrease the enthalpy and entropy change associated with the liganded system. This is indeed what is observed from the ITC data, where the nicotine binding shows a large decrease in enthalpy of the system which is accompanied by reduced system entropy. The conformational dynamics of the protein in the nicotine-bound form is also reduced significantly and is the least among all the ligands examined here.

The binding of ketoconazole is intermediate in terms of compensatory mechanism between that of camphor/norcamphor and nicotine. The enthalpy and entropy change are equally favorable as observed from the ITC data and with NMR spectra showing little to no change in the conformational flexibility of CYP101 upon ketoconazole binding. As discussed earlier, while a large ligand such as ketoconazole is potentially capable of forming several intermolecular contacts with the protein, the docked structure and minimal NMR chemical shift perturbations in this case suggest otherwise. Therefore, the binding of ketoconazole is not enthalpy dominated but is a mix of both favorable enthalpy and entropy change, with the main enthalpic contribution coming from the direct interaction of ketoconazole with the heme. This interaction should not perturb the protein structure much and therefore the flexibility of ketoconazole-bound CYP101 is not affected significantly but is similar to that of the ligand-free form. Such differential enthalpy-entropy compensation has also been seen previously in other P450s such as CYP2B4 when studied with a series of inhibitors of

varying chemical characteristics, where CYP2B4 changes conformation and becomes more compact in the presence of some inhibitors, but not with all [74].

An important aspect that is not included in the above discussion is the fact that the dynamic changes in the protein when binding to camphor, norcamphor and nicotine ligands are manifested over the entire protein. Especially, in the case of nicotine-bound form, the dynamic changes are quite dramatic and lead to almost complete loss of flexibility throughout the protein. Such an observation can be explained by utilizing the concept of enthalpy-entropy transduction in which local thermodynamic forces such as ligand binding enthalpies can be transduced into different global thermodynamics via conformational selection [75] In this mechanism, proteins that can bind a series of ligands are prone to sample different conformations of similar free energy but different enthalpy and entropy. As a consequence, the same protein can bind different ligands with similar free energies but with different entropies and enthalpies. In this way, this transduction mechanism can lead to a highly linear enthalpy-entropy compensation mechanism, which is what is likely observed in our case. It should be noted that this transduction mechanism does not preclude the role of water in this compensation mechanism. An added characteristic that results from this transduction is selective stabilization of certain conformational states that can lead to complete transformation of protein dynamics at a global level. Our observations support such a transduction mechanism in CYP101 based on dramatic global dynamic changes and conformational selection observed by NMR for nearly all ligands studied so far. This has experimentally not been demonstrated for any protein previously to the best of our knowledge, although the concept and relevant theory from analysis of simulations has been around for a few years[75]. More studies are needed to confirm and characterize this phenomenon in CYP101 and other P450s. If validated, this will open up a new paradigm in the area of P450 ligand recognition with major implications in drug design in human P450s.

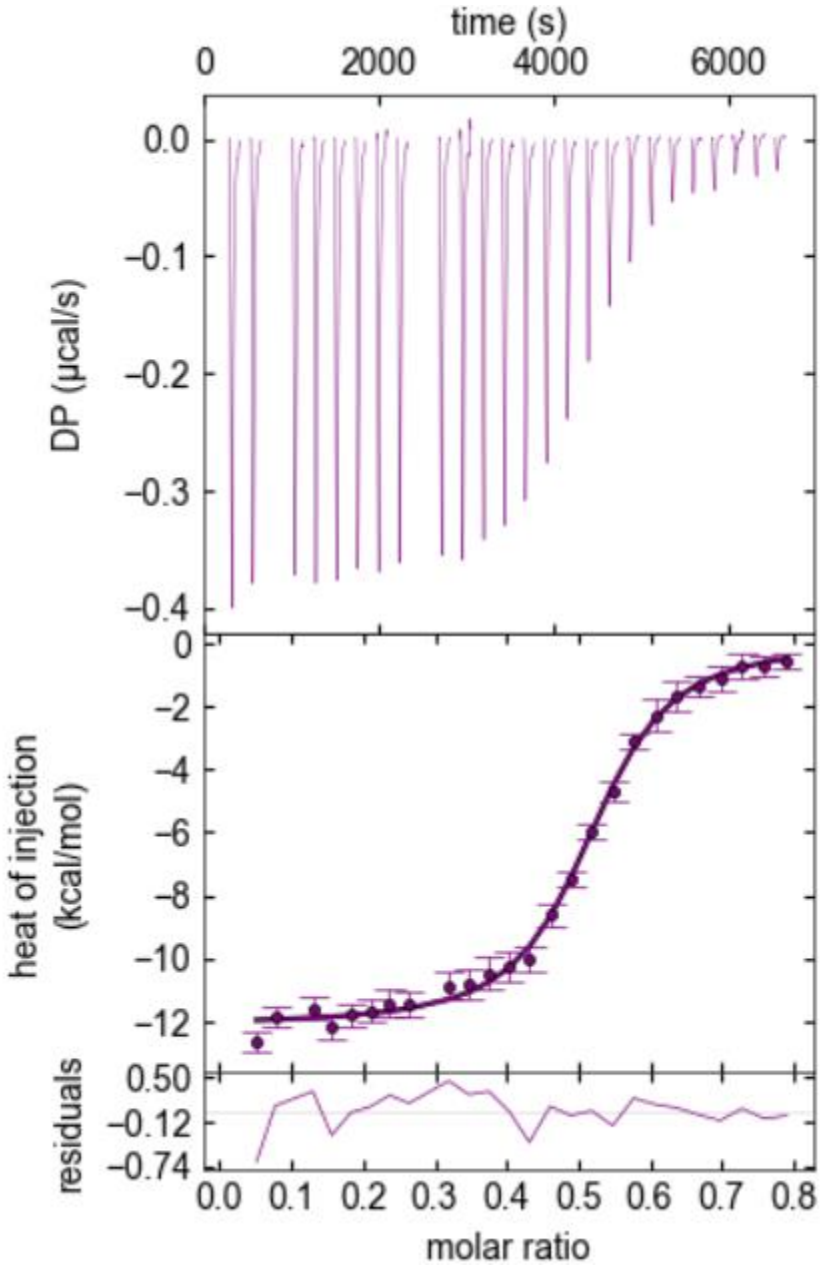


Figure 23: ITC curves and fit for binding of nicotine to CYP101

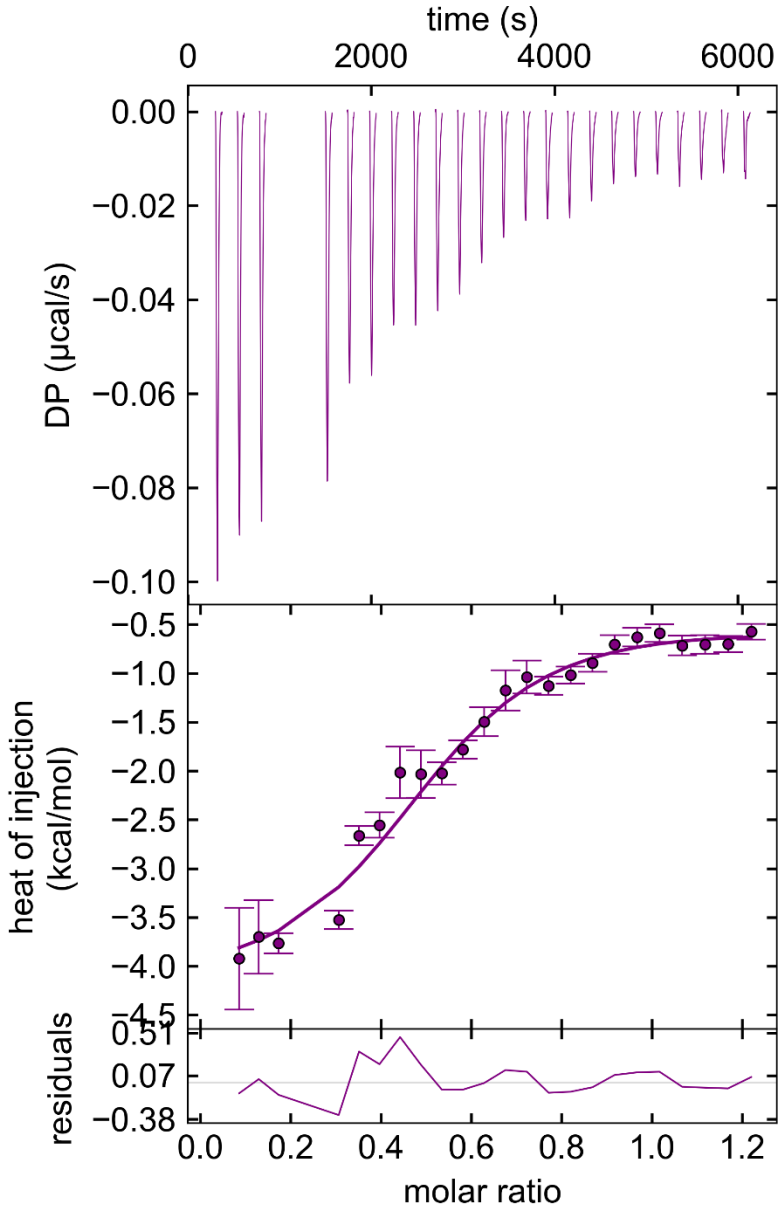


Figure 24: ITC curves and fit for binding of ketoconazole to CYP101

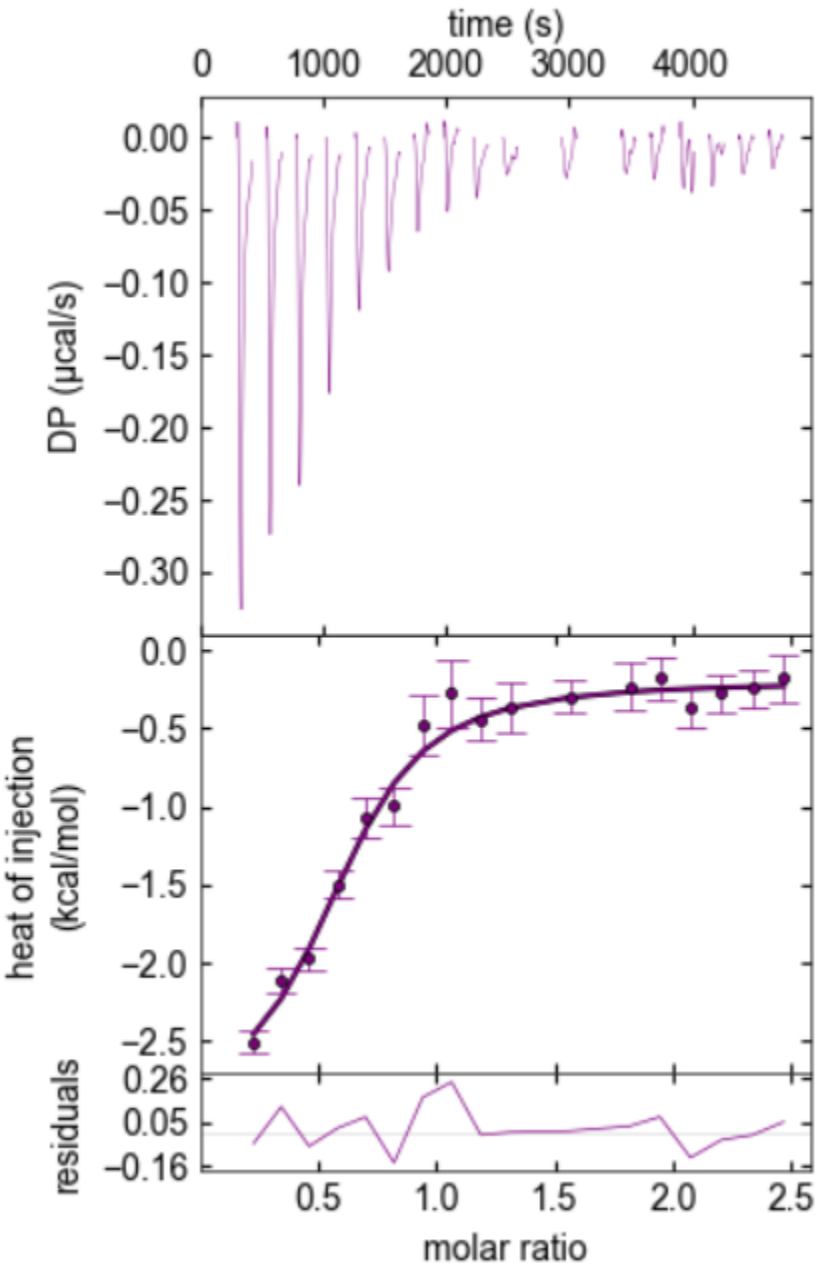


Figure 25: ITC curves and fit for binding of camphor to CYP101

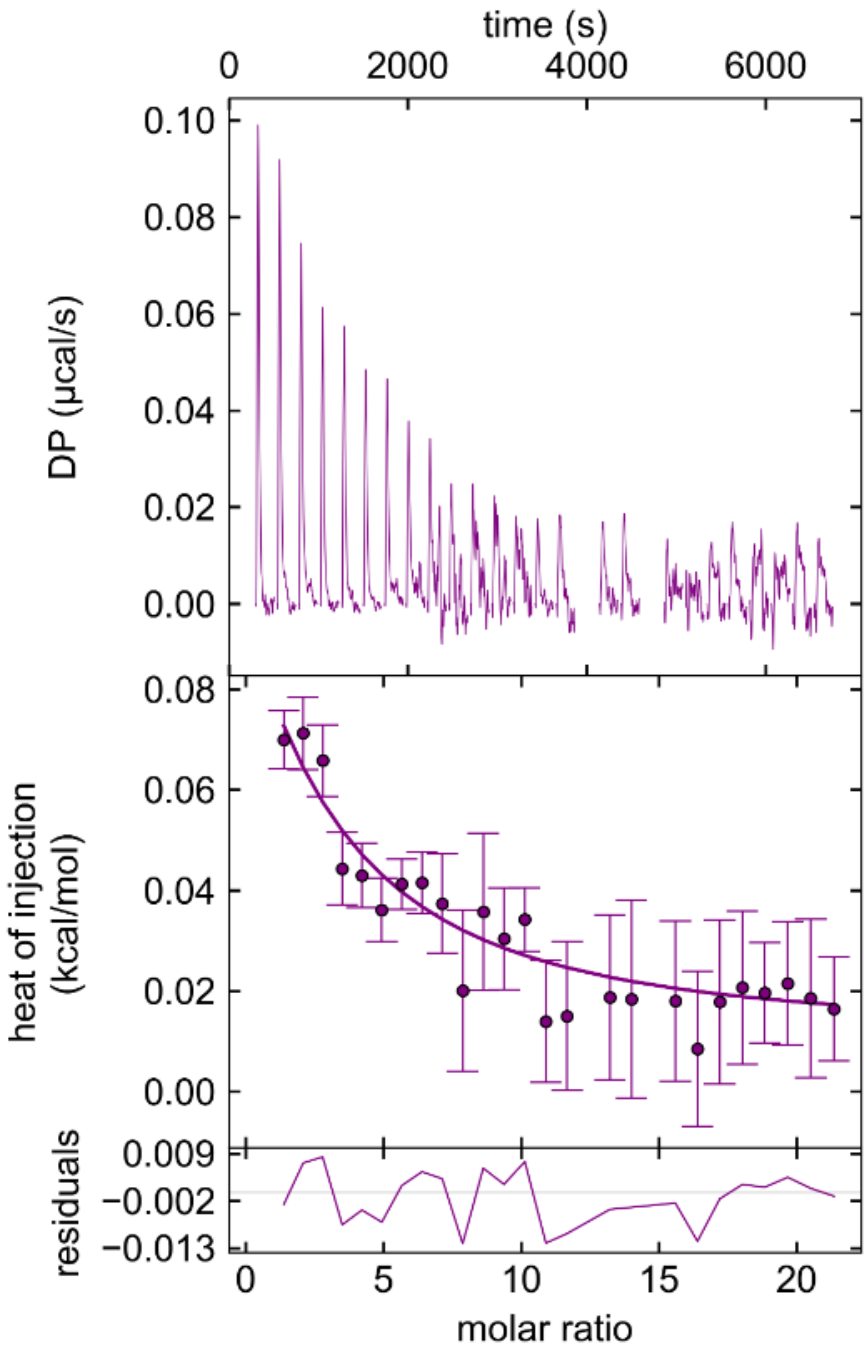


Figure 26: ITC curves and fit for binding of norcamphor to CYP101

CHAPTER SEVEN:
CONCLUSIONS AND FUTURE DIRECTIONS

7.1 Summary of Project and Conclusions

Understanding of protein dynamics in context of protein function is an important biological goal. The cytochrome P450 system offers a valuable paradigm to understand protein dynamics in context of differential ligand binding where a vast multitude of P450s are able to recognize millions of diverse ligands. In this research project, an investigation of the dynamic changes that happen in a model P450 system, CYP101, upon binding a suite of different ligands was undertaken in an effort to address this aspect. Solution NMR spectroscopy was chosen as the method of choice since it is able to characterize the dynamic processes on a variety of timescales. NMR data from amide exchange measurements helped elucidate the slow timescale motions of CYP101 in complex with various ligands and comparison with the dynamics of ligand-free form allowed us to map the dynamic differences resulting from ligand binding. An important finding from this comparison was that the dynamic differences are not just restricted to the flexible SRS regions but are observed throughout the protein. This validates the preliminary observations from 2D NMR HSQC spectra where considerable effects were seen in terms of line-broadening, peak intensities and peak splittings throughout the NMR spectra indicating that the protein responds to the ligand event as a whole. This finding also strengthens results from previous neutron scattering studies on the same system which demonstrated that collective motions on ps-ns timescales span across the protein and facilitate opening and closing of the ligand access channel. Thus, the role of dynamics in the protein is clearly linked to its function.

The project required overcoming multiple challenges, both in terms of methodology and also in interpretation of diverse sets of data. A large number of isotopically labeled samples, both uniform and selective labeled, were made and used within multidimensional NMR experiments to collect and assign resonances in CYP101 as a first step before acquisition of dynamic data. This included several 3D NMR experiments on five different CYP101 forms, 2D NMR experiments on multiple

selectively labeled samples again on five different CYP101 forms, followed by collection of several sets of amide exchange data on all of these forms. All of these experiments resulted in a total close to 200 individual spectra that were collected, processed and analyzed over the course of this project. This not only allowed augmenting of the existing backbone and sidechain resonance assignments available for CYP101 in ligand-free and camphor-bound forms, but also helped provide additional assignments for the three other forms of CYP101 not previously available. These assignments provided substantial coverage in monitoring important regions of the protein such as the SRS as well as other regions outside of the protein that are dynamic and were critical in obtaining dynamic insights on these regions from the amide exchange studies. In addition to the NMR experiments, molecular docking methods were employed to obtain a reasonable structural model for the CYP101-ketoconazole complex which was then used in the interpretation of the dynamic data for this complex.

Finally, ITC experiments were also performed to gain insights into the thermodynamic basis of ligand binding to CYP101 for all 4 ligands examined in this project. The dissection of thermodynamic parameters for each ligand binding to CYP101 clarified the thermodynamic driving force for each binding event and revealed interesting patterns in terms of enthalpy-entropy compensation mechanisms for this series of ligands. A valuable aspect of this study was the insight that water associated with the protein may be involved in this compensation mechanism, something which is overlooked frequently in structural studies of ligand binding in the P450 area. Based on the ITC data and its correlation with the observed protein dynamics from NMR experiments, a novel enthalpy-entropy transduction mechanism was suggested to explain the differential change in dynamics upon ligand binding, which creates a fascinating link between dynamic and thermodynamic characteristics of the protein. This framework provides an integrated view of protein dynamics and thermodynamics and would be very useful in practical aspects of manipulating ligand binding to flexible proteins such as P450s, especially in context of drug design in the case of human P450s.

7.2 Future Directions

Only slow timescale motions for CYP101 from amide exchange measurements were characterized in this project. However, as the 2D NMR spectra of CYP101 with different ligands clearly show that the dynamics of CYP101 is occurring on multiple timescales depending on the ligand bound. Therefore, a logical next step in the future would be to do a dynamic characterization with the same ligand-bound forms on other timescales, such as fast and intermediate, using NMR spectroscopy. The availability of an expanded set of resonance assignments from the current project should benefit this characterization and speed up the process of dynamic analysis of data from these new experiments. Such a complete characterization of dynamic differences on multiple timescales in this system will go a long way in garnering additional insights into the dynamic mode of differential ligand binding and help test whether enthalpy-entropy transduction mechanism is operating on these timescales as well. ITC experiments could also be conducted in the presence of various osmolytes to confirm the role of the water network in ligand binding[76]. An efficient way to assess the enthalpy-entropy transduction mechanism would be to perform titration experiments on CYP101 with the different ligands to monitor dynamic changes as a function of ligand concentration. This would provide evidence for whether individual ligands go through a different conformational selection pathway to accomplish this transduction. This can easily be followed by observing chemical shifts as they transition through different populated states starting from the ligand-free state to reach the final conformation most compatible with a certain ligand. Such experiments are currently in progress and preliminary results (data not shown) show strong evidence for such a mechanism in a ligand-dependent manner. When completed, these experiments will help make important progress towards a long-sought goal of linking the conformational dynamics with thermodynamics of a protein

APPENDICES

APPENDIX

Table A.1 NMR Chemical shift Assignments for oxidized Cyp101 in complex with the substrates camphor and norcamphor

All chemical shifts are referenced with respect to DSS

Type	Number	Secondary structure	Camphor H (ppm)	Camphor N(ppm)	Camphor Cα(ppm)	Norcamphor H (ppm)	Norcamphor N (ppm)	Norcamphor CA (ppm)
LEU	11	N term Loop	7.63	122.24	53.8	7.6	122.14	53.9
ALA	12	N term Loop	8.617	130.2	51.3	8.64	130.1	
HIS	17	N term Loop	7.71	112.2	56.7	7.72	112.1	56.6
VAL	18	N term Loop	6.8	127.8	60.9	6.82	127.5	
LEU	22	N term Loop	6.88	119.3	54.5			
VAL	23	N term Loop	7.1	121.1	63.78	7.11	120.89	63.6
PHE	24	N term Loop	9.15	133.7	58.7	9.23	133.75	
PHE	26	N term Loop	8.3	125.25	58.5	8.32	125.2	61.6
ASP	27	N term Loop	8.34	133.14		8.14	133.4	
TYR	29	N term Loop	7.89	118.8	60.3	7.95	119.5	60.1
ASN	30	N term Loop	6.74	116.84	52.4	6.83	116.64	
ALA	36	N term Loop	7.809	124.96	53.6	7.81	125	53.4
GLY	37	A Helix	7.41	109.9	43.2	7.4	109.8	
ASN	49	Aβ1 Loop	8.267	116.9	54.12	8.27	116.8	53.6
VAL	50	Aβ1 Loop	7.749	126.3	61.35	7.72	125	
VAL	54	β1 Sheet	8.56	113.8	58.79	8.5	113.72	58.7
TRP	55	β1 Sheet	8.92	123.54	56.1	8.96	123.44	56.3
THR	56	β1 Sheet	7.9	120.4	58.6	7.9	120.38	
CYS	58	β1 Sheet	8	119.56	59.55	7.99	119.46	59.3
ASN	59	β1 Sheet	9.22	115.3	56.3	9.29	115.6	56.2
GLY	60	β1 Sheet	8.7	104.3	45.675	8.83	104.4	45.8
GLY	61	β1 Sheet	7.45	112.2	45.068	7.5	112.4	44.8
HIS	62	β1 Sheet	7.11	122.07	56.25	7.14	122	
TRP	63	β1 Sheet			57.0118			
ILE	64	β1 Sheet	10.4	126.2	63.011	10.37	126.2	62.8
ALA	65	β1 Sheet	7.89	128.9	51.7	7.8	128.9	51.1
THR	66	β1 Sheet	8.6	109.46	62.8	8.56	109.337	62.8
ARG	67	B Helix	6.49	115.14	52.98	6.38	115.1	
GLY	68	B Helix	10.48	112.7	47	10.3	112.57	47.8
GLN	69	B Helix	8.91	120.1	59.7	8.89	120.3	59.7
LEU	70	B Helix	6.43	117.4	59.7	6.32	117.46	
CYS	85	BB' Loop	5.74	111.7	54.75	5.78	111.87	
PHE	87	BB' Loop	5.85	117.6	54.3	6.1	118.58	

Table A.1 continued

ILE	88	BB' Loop	5.78	121.88	58.02			
ARG	90	B' Helix	9.38	124.77	61.06	9.37	125.4	61.3
GLU	91	B' Helix	9.378	116.53	60.56	9.39	117	
ALA	92	B' Helix	7.2	123.7	54.776	7.23	123.2	
GLY	93	B' Helix	7.3881	104.83	48.082	7.52	104.84	
GLU	94	B' Helix	8.94	121.97	59.62	8.94	121.73	
ALA	95	B' Helix	6.88	120.48	53.4	6.9	119.9	52.5
TYR	96	B' Helix	7.3282	121.91	60.37	7.25	121.6	
PHE	98	B'C Loop	5.78	121.88	57.9			
PRO	122	C Helix			66.3			66.3
VAL	123	C Helix	7.22	116.6	65.48	7.14	117.2	65.33
VAL	124	C Helix	7.46	122	67.4	7.45	121.9	67.4
ASP	125	C Helix	8.2	118.6	58.1	8.2	118.9	58
LYS	126	C Helix	7.26	119	58.9	7.19	119.26	
SER	141	D Helix	7.62	113.3	57	7.62	113.3	
GLN	145	D Helix	8.33	114.445	58.2	8.35	114.24	57.5
GLY	146	DE Loop	6.5	102.27	44.81	6.54	102.14	
GLN	147	DE Loop	5.93	111.72	54.7	5.97	111.64	54.4
CYS	148	DE Loop	8.28	111.05	57.36	8.29	110.98	57.1
ASN	149	E Helix	9.18	120.3	52.7	9.18	120.2	
PHE	150	E Helix	9.32	124.8	63.8	9.32	124.8	63.8
THR	151	E Helix	7.25	109.5	65.1	7.59	111.5	
GLU	152	E Helix	6.27	118.9	58.31			58.3
ASP	153	E Helix	8.6	115.1	56.2	8.68	115.1	
MET	164	E Helix			58.3			
LEU	165	E Helix	7.7	121.19	58.66			
LEU	166	E Helix	8.2	123.6	58.7			58.5
ALA	167	E Helix	8.67	119.2	53.65	8.68	119.47	
GLY	168	E Helix	7.92	109.44	47.521	7.8	110.58	
GLU	171	EF Loop	8.75	123.1	60.1	8.66	123.43	59.9
GLU	172	EF Loop	9.42	120.37	59.7	9.4	120.24	
LYS	178	F Helix	8.1	119.03	59.8	8.14	119.58	59.3
TYR	179	F Helix	7.14	118.7	61.8	7.59	119	
THR	185	F Helix	7.63	107.32	59.5	7.24	107.2	
ASP	188	FG Loop	8.37	127.1	53.9	8.34	127	53.5
GLY	189	FG Loop	8.5	110.66	45.5	8.45	110.09	45.3
SER	190	FG Loop	8.61	120	61.3	8.7	120	61
MET	191	FG Loop	8.16	120.7	55.9	8.17	120.829	55.5
THR	192	G Helix	8.7	117	62.1	8.66	116.9	
PHE	193	G Helix	8.52	123.77	61.59			62.5

Table A.1 continued

ALA	194	G Helix	8.39	119.4	56.08	8.455	119.61	
GLU	198	G Helix	7.72	117.2	58.2	7.66	117.24	58
ALA	199	G Helix	7.373	121.68	55.13	7.23	121.7	55.1
LEU	200	G Helix	7.85	124.4	58.3	7.88	125.3	
PRO	206	G Helix			65.7			
ILE	207	G Helix	6.6	119.3	55.8	6.65	119.37	55.5
ILE	208	G Helix	8.27	121.4	63.8	8.28	121.66	
LYS	214	G Helix	8.34	116.6	60.4	8.4	116.67	
PRO	215	GH Loop			64.29			63
GLY	216	GH Loop	8.52	111.517	43.9	8.53	111.4	
THR	217	GH Loop	8.092	117.97?	62			
ALA	219	H Helix	8.55	118.8				
ILE	220	H Helix	7.89	115.2	65.4	7.8	114.3	
VAL	223	H Helix	7.85	119.06	67.4	7.79	119	67.2
ALA	224	H Helix	9.04	117.1	55	9	117.07	54.5
ASN	225	H Helix	7.2	113	53.4	7.24	113.2	53.4
GLY	226	β2 Sheet	7.72	110.2	45.65	7.7	109.9	46.2
GLN	227	β2 Sheet	8.092	118.58	55.1	7.99	118.9	55.01
VAL	228	β2 Sheet	8.922	118.17	60.42	8.92	118.14	
ASN	229	β2 Sheet	8.9	121.9	no ca?			54.3
GLY	230	β2 Sheet	8.55	103.7	45.1	8.6	103.66	46.1
ARG	231	β2 Sheet	7.98	119.5	53.23	7.97	119.5	
ILE	233	β2 Sheet	8.237	127.2	62.4	8.24	127.15	62.4
THR	234	I Helix	9.09	120.1	61.38	9.04	119.9	
ALA	265	I Helix	8.21	121.2	54.86	8.21	121.2	
LYS	266	I Helix	7.13	116.49	57.7	7.27	116.38	
SER	267	I Helix	7.09	115.53	63.7	7.02	115.86	
HIS	270	J Helix						59.3
ARG	271	J Helix	7.6	117.9	61.5	7.62	117.87	61.4
GLN	272	J Helix	8.35	117.4	59.2	8.39	117.2	59
GLU	273	J Helix	7.47	119.23	60.4	7.5	119.7	60.3
LEU	274	J Helix	7.03	115.8	56.6	7.07	115.69	
GLU	279	JK Loop	9.98	122.35	58.8	10.04	122.37	58.6
ARG	280	K Helix	8.53	121.2	57	8.59	121.24	
ASP	304	β3β4 Loop	8.46	121.79	55.7	8.46	121.7	55.3
TYR	305	β4 Sheet	8.43	126.6	55.6	8.43	126.63	55.3
GLU	306	β4 Sheet	7.69	130.54	57.28	7.66	130.3	
PHE	307	β4 Sheet	8.4933	130.335	55.3	8.49	130.3	54.9
HIS	308	β4 Sheet	8.92	124.2	57.9	8.937	123.95	57.4
GLY	309	β4 Sheet	8.33	102.94	45.48	8.35	102.85	45.7

Table A.1 continued

VAL	310	β4 Sheet	7.537	124.17	61.93	7.51	124.1	61.49
GLN	311	β4 Sheet	8.37	126.35	56.83	8.38	126.2	56.6
LEU	312	β4 Sheet	8.65	128.9	54.32	8.65	128.75	54
LYS	313	β3β4 Loop	8.5706	124.97	54.36	8.46	124.7	53.9
LYS	314	β3β4 Loop	8.74	122.4	59.2	8.7	122.25	59
GLY	315	β3 Sheet	8.63	117.64	46.1	8.61	117.67	45.6
ASP	316	β3 Sheet	8.96	123.19	56.2	8.94	123.33	55.9
GLN	317	β3 Sheet	9.47	119.2	56.4	9.45	119.14	
ILE	318	β3 Sheet			58.8			
LEU	319	β3 Sheet	9.13	131.5	55.4	9.07	131.1	
LEU	320	β3 Sheet	8.83	132.2	52	8.77	131.6	
ALA	333	β3L loop	8.464	124.6	53	8.54	125.1	
MET	336	β3L loop	8.724	116.86	58.3	8.71	116.87	
HIS	337	β3L loop	8.21	123.8	56.3	8.06	123.6	
VAL	338	β3L loop	8.04	127.1	62.2	8.06	126.89	
ASP	339	β3L loop	10.15	131.6	52.58	10.1	131.87	52.1
PHE	340	β3L loop	9.496	124.71	52.48	9.49	124.9	52.7
SER	341	β3L loop	8.72	116.51	58.42	8.59	116.01	58.2
ARG	342	β3L loop	7.18	124.5	57.01	7.23	124.51	
ILE	368	L Helix			58			
VAL	369	L Helix	7.52	109.2	60	7.42	109.8	
LYS	372	L Helix	8.8	120.8	55.7	7.26	119	
PRO	379	Lβ5 Loop			64.78			
ASP	380	Lβ5 Loop	8.134	116.8	54.06	8.15	117	53.6
PHE	381	Lβ5 Loop	6.634	119.9	56.019	6.64	119.34	55.5
SER	382	β5 Sheet	9.04	111.8	57.8	9.08	111.79	57.3
ILE	383	β5 Sheet	8.51	123.01	61.3	8.51	123	61.1
ALA	384	β5 Sheet	8.157	132.1	52	8.17	132.12	
GLY	386	β5 Sheet	8.5	111.5	44.3	8.48	111.35	45.8
ALA	387	β5 Sheet	7.61	124.1	53.3	7.62	124.2	53.04
GLN	388	β5 Sheet	8.41	123	54.8	8.38	122.93	54.7
ILE	389	β5 Sheet	8.56	127.1	59.8	8.54	127.3	59.5
GLN	390	β5 Sheet	9.48	129.9	54.9	9.47	129.6	54.9
HIS	391	β5 Sheet	8.97	125.8	53.7	8.89	125.9	53.4
LYS	392	β5 Sheet	8.89	119.2	56	8.87	119.5	55.5
SER	393	β5 Sheet	8.73	117.68	56.1	8.84	117.86	56
GLY	394	β5 Sheet	6.78	110.4	46.5	7.02	111.42	
VAL	396	β5 Sheet	5.91	114.7	61.2			
SER	397	β5 Sheet			61.46			

Table A.1 continued

GLY	398	β5 Sheet	8.54	103.11	46.113	8.61	103.2	
VAL	399	β5 Sheet	8.06	119.8	63.04			
GLN	400	β5 Sheet	8.93	126.3	53.8	8.95	126.4	
ALA	401	β5 Sheet	7.355	117.5	52.6	7.37	118.2	52.6
LEU	402	β5 Sheet	8.63	120.1	55.9	8.76	119.87	
VAL	405	β5 Sheet	9.22	114.49	60.1	9.2	114.67	
ALA	409	C term Loop	8.68	121.79	55.1	8.7	122	
THR	410	C term Loop	8.15	108.76	62.2	8.15	108.66	
LYS	412	C term Loop	8.9257	121.35	55.9	8.98	121.1	
ALA	413	C term Loop	8.073	125.4	52.27	8.05	125.8	
VAL	414	C term Loop	7.9	127.1	64.8	7.91	126.94	
	Total		153			140		

Table A.2 A1 NMR Chemical shift Assignments for oxidized Cyp101 in ligand-free form

All chemical shifts are referenced with respect to DSS

Type	Number	Secondary structure	Ligand Free H (ppm)	Ligand Free N(ppm)	Ligand Free Cα(ppm)
LEU	11	N term Loop	7.8	122.6	54.1
ALA	12	N term Loop	8.6	130.16	50.7
PRO	16	N term Loop			64.12
HIS	17	N term Loop	7.66	112	56.3
VAL	18	N term Loop	6.77	127.5	60
LEU	22	N term Loop	7.05	119.2	54.4
VAL	23	N term Loop	7.064	120.7	62.9
PHE	24	N term Loop	9.2	133.83	58
PHE	26	N term Loop	8.33	126	58.5
ASP	27	N term Loop	8.234	133.14	50.1
TYR	29	N term Loop	7.97	118.06	59.2
ASN	30	N term Loop	6.81	116.21	52.06
ALA	36	N term Loop	7.74	124.8	52.8
GLY	37	A Helix	7.36	109.8	44
ASN	49	Aβ1 Loop	8.23	116.73	53.6
VAL	50	Aβ1 Loop	7.67	126.2	60.77
VAL	54	β1 Sheet	8.54	113.7	58.2
TRP	55	β1 Sheet	8.88	123.1	55.6
THR	56	β1 Sheet	7.82	122.6	58.9
CYS	58	β1 Sheet	8	119.4	58.97
ASN	59	β1 Sheet	9.28	115.6	55.7
GLY	60	β1 Sheet	8.76	104.1	44.72
GLY	61	β1 Sheet	7.46	112.2	43.8
HIS	62	β1 Sheet	7.12	122	55.74
ILE	64	β1 Sheet	10.3	125.9	62.39
ALA	65	β1 Sheet	7.7	128.77	50.93
THR	66	β1 Sheet	8.44	109.58	61.99
ARG	67	B Helix	6.2	114.65	52
GLY	68	B Helix	10.15	112.18	46.98
GLN	69	B Helix	8.71	120.1	58.9
LEU	70	B Helix	6.21	116.8	58.8
CYS	85	BB' Loop	5.77	117.84	55.7
ARG	90	B' Helix	9.29	126	62.7
GLU	91	B' Helix	9.2	113.47	59.3
ALA	92	B' Helix	7.1	123.22	

Table A.2 Continued

GLY	93	B' Helix	7.68	104.32	45.94
GLU	94	B' Helix	8.92	121.55	
ALA	95	B' Helix	7.15	119.56	53.8
TYR	96	B' Helix	7.43	120.9	59.6
VAL	123	C Helix	7.12	118.6	67.9
VAL	124	C Helix	7.44	121.8	66.8
LYS	126	C Helix	7.15	119.6	58.9
SER	141	D Helix	7.52	113.1	56.5
GLN	145	D Helix	8.24	113.83	57.74
GLY	146	DE Loop	6.5	102.04	44.4
GLN	147	DE Loop	5.86	110.71	54.1
CYS	148	DE Loop	8.22	110.74	56.7
ASN	149	E Helix	9	119.4	52.7
PHE	150	E Helix	9.32	124.8	62.7
THR	151	E Helix	7.07	109.14	65.1
GLU	152	E Helix	6.54	118.9	57.3
ASP	153	E Helix	8.68	115.3	56.07
LEU	165	E Helix	7.8	120.2	58.8
ALA	167	E Helix	8.74	120.2	52.5
GLY	168	E Helix	7.38	109.37	47.6
GLU	171	EF Loop	8.65	123.09	59.5
GLU	172	EF Loop	9.29	119.82	59
LYS	178	F Helix	8.03	119.5	58.9
THR	185	F Helix	6.72	107.96	
ASP	188	FG Loop	8.4	126.4	53.1
GLY	189	FG Loop	8.44	109.9	45
SER	190	FG Loop	8.54	119.7	59.77
MET	191	FG Loop	8.16	121.19	56.33
THR	192	G Helix	8.56	116.48	62.3
ALA	194	G Helix	8.37	119.8	55.3
ALA	196	G Helix	7.84	126.27	55.9
LYS	197	G Helix	8.15	118.1	54.3
GLU	198	G Helix	7.75	117.12	57.6
ALA	199	G Helix	7.11	122	55.6
LEU	200	G Helix	8.02	124.7	57.9
ILE	207	G Helix	6.61	118.96	63.9
ILE	208	G Helix	8.2	122.2	61.9
LYS	214	G Helix	8.4	117.15	58.4
GLY	216	GH Loop	8.49	111.45	43.3

Table A.2 continued

ILE	220	H Helix	7.67	113.9	
VAL	223	H Helix	7.55	118.98	66.3
ALA	224	H Helix	8.9	117.2	54.5
GLY	226	β2 Sheet	7.63	109.58	45.1
GLN	227	β2 Sheet	8.14	118.8	54.6
VAL	228	β2 Sheet	8.76	117.63	59.9
GLY	230	β2 Sheet	8.56	103.4	45.6
ARG	231	β2 Sheet	7.88	119	52.64
ILE	233	β2 Sheet	8.28	127.4	61.99
THR	234	I Helix	8.97	119.98	60.7
ALA	265	I Helix	8.23	121.67	55.3
LYS	266	I Helix	7.32	116.47	57.1
SER	267	I Helix	6.98	115.9	62.6
GLU	273	J Helix	7.46	119.1	59.7
LEU	274	J Helix	7.02	115.65	55.9
GLU	279	JK Loop	10.02	122	58.1
ARG	280	K Helix	8.56	121.2	56.5
ASP	304	β3β4 Loop	8.44	121.56	55
TYR	305	β4 Sheet	8.34	126.7	55
GLU	306	β4 Sheet	7.65	130.2	56.44
PHE	307	β4 Sheet	8.43	130	54.5
HIS	308	β4 Sheet	8.9	123.9	57.1
GLY	309	β4 Sheet	8.3	102.7	45
VAL	310	β4 Sheet	7.45	124	61.6
GLN	311	β4 Sheet	8.33	126.35	56.22
LEU	312	β4 Sheet	8.57	128.7	53.7
LYS	313	β3β4 Loop	8.49	124.51	53.6
LYS	314	β3β4 Loop	8.69	121.2	58.8
GLY	315	β3 Sheet	8.56	117.5	44.1
ASP	316	β3 Sheet	8.8	123.12	55.6
GLN	317	β3 Sheet	9.35	118.9	55.5
LEU	319	β3 Sheet	9.04	130.8	54.7
LEU	320	β3 Sheet	8.6	131.6	50.8
ALA	333	β3L loop	8.46	124.5	52.3
MET	336	β3L loop	8.7	116.8	57.8
HIS	337	β3L loop	8	124.5	55.9
VAL	338	β3L loop	8.01	126.9	61.5
ASP	339	β3L loop	10.07	131.4	51.87
PHE	340	β3L loop	9.28	125.9	51.8
SER	341	β3L loop	8.73	116.3	57.8

Table A.2 continued

ARG	342	β3L loop	7.22	124.53	56.3
VAL	369	L Helix	7.3	110.6	60
LYS	372	L Helix	8.93	121.29	55.2
ASP	380	Lβ5 Loop	8.1	116.77	53.5
PHE	381	Lβ5 Loop	6.57	119.23	55.26
SER	382	β5 Sheet	9.08	111.67	57
ILE	383	β5 Sheet	8.5	122.67	60.39
ALA	384	β5 Sheet	8.1	131.8	51.4
GLY	386	β5 Sheet	7.64	109.5	45.1
ALA	387	β5 Sheet	7.64	123.9	52.7
ILE	389	β5 Sheet	8.599	126.8	58.99
GLN	390	β5 Sheet	9.51	129.8	54.6
HIS	391	β5 Sheet	8.89	125.4	54.42
LYS	392	β5 Sheet	8.77	119.07	54.2
SER	393	β5 Sheet	8.86	117.9	55.4
GLY	394	β5 Sheet	6.92	110.92	
GLY	398	β5 Sheet	8.57	103.44	45.73
GLN	400	β5 Sheet	8.99	126.34	53.2
LEU	402	β5 Sheet	8.73	121	54.7
VAL	405	β5 Sheet	9.19	114.23	59.6
ALA	409	C term Loop	8.84	122.72	55.5
THR	410	C term Loop	8.1	108.21	61.8
LYS	412	C term Loop	8.95	121.22	55.3
ALA	413	C term Loop	8.04	125.6	51.7
VAL	414	C term Loop	7.87	126.88	63.9
	Total		138		

Table A.3 A1 NMR Chemical shift Assignments for oxidized Cyp101 in complex with nicotine and ketoconazole (keto)

All chemical shifts are referenced with respect to DSS

Type	Number	Secondary structure	Nicotine H (ppm)	Nicotine N(ppm)	Nicotine Cα(ppm)	Keto H (ppm)	Keto N (ppm)	Keto CA (ppm)
LEU	11	N term Loop	7.64	122.1	54.2	7.68	122.6	54.01
ALA	12	N term Loop	8.67	130.3	51.4	8.6	130.01	51.1
PRO	16	N term Loop			64.9			65.1
HIS	17	N term Loop	7.66	111.9	56.9	7.67	111.97	56.89
VAL	18	N term Loop	6.7	127.4	60.9	6.77	127.45	60.8
LEU	22	N term Loop	6.89	120	53.9	6.86	119.85	54.3
VAL	23	N term Loop	7.04	120.89	63.62	7.04	120.75	63.7
PHE	24	N term Loop	9.2	133.8	58.9	9.14	133.8	58.7
PHE	26	N term Loop	8.27	124.8	58	8.28	125.2	
ASP	27	N term Loop				8.188	132.28	
TYR	29	N term Loop	8	119	60.19	8	118	60.3
ASN	30	N term Loop	6.8	116.9	52.18	6.85	116.68	52.6
ALA	36	N term Loop	7.75	124.8	53.57	7.74	124.9	53.5
GLY	37	A Helix	7.37	109.7	43.4	7.31	109.7	44.2
ASN	49	Aβ1 Loop	8.1	115.5	54	8.24	116.92	54.1
VAL	50	Aβ1 Loop	7.72	126.2	61.41	7.6	125.93	61.5
VAL	54	β1 Sheet	8.47	113.8	58.36	8.49	113.72	58.9
TRP	55	β1 Sheet	8.82	122.1	55.8	8.87	123.4	56.3
THR	56	β1 Sheet	7.9	120.4	59.6	7.88	120.3	58.6
CYS	58	β1 Sheet	7.77	119.58	59.3	8.01	119.54	
ASN	59	β1 Sheet	9.29	115.47	55.98	9.29	115.81	56.4
GLY	60	β1 Sheet	8.83	104.5	45.1	8.78	104.3	45.9
GLY	61	β1 Sheet	7.49	112.24	44.44	7.51	112.19	44.9
HIS	62	β1 Sheet	7	121.85	56.3	7.11	122.49	56.41
ILE	64	β1 Sheet	10.34	126	62.5	10.26	125.88	63.1
ALA	65	β1 Sheet	7.67	129	51	7.75	128.7	51.3
THR	66	β1 Sheet	8.41	109.39	62.6	8.42	109.08	62.7
ARG	67	B Helix	6.248	115	52.2	6.24	114.65	52.51
GLY	68	B Helix	10.17	112.2	45	10.2	112.61	47.8
GLN	69	B Helix	8.58	120.2	59.75	8.74	119.88	59.6
LEU	70	B Helix	6.22	117.56	59.7	6.21	117.56	59.7
CYS	85	BB' Loop	5.66	111.97	55.1	5.76	112.17	55.3
PRO	86	BB' Loop			64.7			
PHE	87	BB' Loop	6.05	117.24	54.6			
ILE	88	BB' Loop	6.04	121.3	57.9			

Table A.3 continued

ALA	92	B' Helix	7.33	123.77	54.38			
GLY	93	B' Helix	7.54	105	49.9	7.76	104.665	45.9
GLU	94	B' Helix	9.01	121.82	59.2	8.89	121.7	59.1
ALA	95	B' Helix	7.09	120.4	53	6.85	121.5	54.2
TYR	96	B' Helix	7.2	120.51	60.2	7.25	119.65	60.2
PHE	98	B'C Loop	6.05	121.8	57.7			
VAL	123	C Helix	7.22	116.5	65.1	7.2	118.3	65.48
VAL	124	C Helix	7.42	122	67.4	7.48	121.4	67.4
ASP	125	C Helix	8.2	118.6	58.1	8.1	118.7	58.1
LYS	126	C Helix	7.09	120	58.9	7.19	119.45	59
SER	141	D Helix	7.58	113.5	56.6	7.65	113.3	57
GLN	145	D Helix	8.28	114.3	58.27	8.25	114	58.4
GLY	146	DE Loop	6.5	102.18	44.8	6.47	102.44	44.4
GLN	147	DE Loop	5.88	111.55	54.7	5.93	111.7	
CYS	148	DE Loop	8.24	110.95	57.54	8.24	111.04	57.48
ASN	149	E Helix	9.17	120.1	52.7	9.1	119.9	52.7
PHE	150	E Helix	9.32	124.9	63.8	9.4	124.6	63.8
THR	151	E Helix	7.25	109.5	65.1	7.05	109.2	65.9
GLU	152	E Helix	6.27	118.9	58.31			
ASP	153	E Helix	8.6	115.1	56.2	8.67	115.2	56.4
LEU	165	E Helix				7.73	120.9	57.38
LEU	166	E Helix	8.46	123.1	60.9	8.05	124.16	58.5
ALA	167	E Helix	8.76	120.05	53	8.67	119.17	53.4
GLY	168	E Helix	7.57	109.37	47.2	7.38	109.61	44
GLU	171	EF Loop	8.66	123.5	59.8	8.67	123.65	60.23
GLU	172	EF Loop	9.35	120.38	59.8	9.3	119.34	59.9
LYS	178	F Helix	8.11	119.55	59.5	7.9965	119.5	59.7
TYR	179	F Helix	7.12	119.55		7.59	119	
THR	185	F Helix	7.16	106.8		7.17	107.4	
ASP	188	FG Loop	8.39	127.2	53.3	8.36	126.79	53.6
GLY	189	FG Loop	8.72	110.8	45.33	8.57	110.3	47
SER	190	FG Loop	8.66	119.9	61.3	8.62	119.7	60.7
MET	191	FG Loop	7.87	120.9	55.9	8.27	121.6	56.1
THR	192	G Helix	8.9	117	61.3	8.57	116.24	61.93
PHE	193	G Helix	8.59	123.5	61.6			
ALA	194	G Helix	8.38	118.21	55.6	8.36	119.6	56.1
ALA	196	G Helix	7.67	125.5	55.9	7.4	125.2	55.97
LYS	197	G Helix	8.23	118.5	55.3	8.11	118.2	55.2
GLU	198	G Helix	7.84	117.2	59.4	7.77	117.12	58.2
ALA	199	G Helix	7.29	121.9	55.7	7.48	122.18	55.5

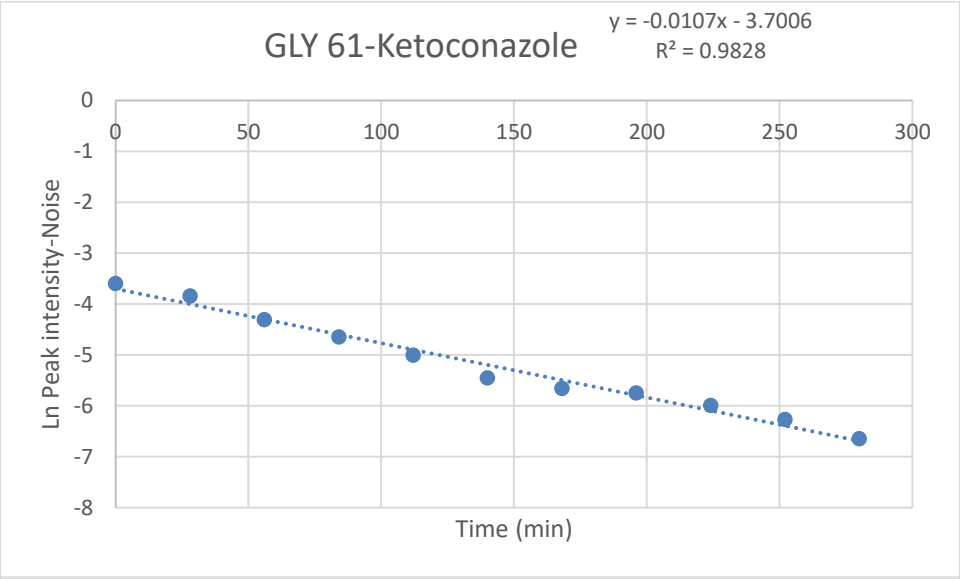
Table A.3 continued

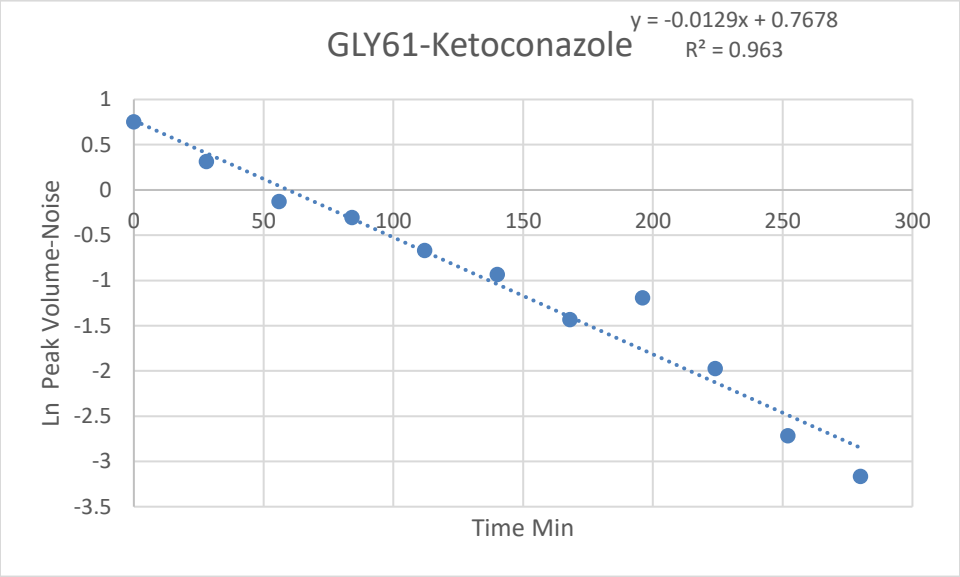
LEU	200	G Helix	7.97	124.1	58.7	8.02	124.33	58.6
ILE	207	G Helix	6.62	119.8	56	6.6	119.17	55.97
ILE	208	G Helix	8.21	121.46	60.93	8.18	122.16	63.3
LYS	214	G Helix	8.37	116.72	59.3	8.17	116.6	60.1
GLY	216	GH Loop	8.49	111.3	43.6	8.49	111.3	44.6
ALA	219	H Helix	8.44	118.42	53.2			
ILE	220	H Helix	7.53	114.8	65.4	7.6	114.3	63.8
VAL	223	H Helix	7.5	119.1	67.4	7.58	118.3	67.5
ALA	224	H Helix	8.88	116.87	55	8.92	117.4	54.9
GLY	226	β2 Sheet	7.62	110	46.2	7.67	109.63	46.2
GLN	227	β2 Sheet	7.99	118.2	55	8.1	118.26	55.1
VAL	228	β2 Sheet	8.857	118.17	60.3	8.7	118.2	60.48
ASN	229	β2 Sheet	8.94	121.7				
GLY	230	β2 Sheet	8.63	103.53	45.1	8.57	103	46.6
ARG	231	β2 Sheet	7.85	119.39	53.2	7.91	119.37	52.91
ILE	233	β2 Sheet	8.167	127.13	62.43	8.21	127.45	62.6
THR	234	I Helix	8.977	119.85	61.47	9.035	119.66	61.53
ALA	265	I Helix	8.16	121.9	55.9	8.256	121.2	54.87
LYS	266	I Helix	7.16	116.6	56.2	7.2	116.47	57.65
SER	267	I Helix	7.014	116.49	62.9	7.05	116	63.4
HIS	270	J Helix						59.62
ARG	271	J Helix				7.59	118.17	61.6
GLN	272	J Helix				8.35	117.1	59.2
GLU	273	J Helix	7.53	119.4	59.9	7.48	119	60.5
LEU	274	J Helix	7.03	115.8	56.2	7.04	115.98	56.7
GLU	279	JK Loop	10.05	122.67	58.4	10.01	122.35	58.8
ARG	280	K Helix	8.55	121.1	57.2	8.54	121.18	57
ASP	304	β3β4 Loop	8.4	121.94	55.7	8.43	121.94	55.7
TYR	305	β4 Sheet	8.4	126.6	55.7	8.45	126.7	55.4
GLU	306	β4 Sheet	7.52	130.49	56.7	7.58	130.44	57.3
PHE	307	β4 Sheet	8.46	130.4	54.5	8.44	130.07	55.2
HIS	308	β4 Sheet	8.8	124.4	57.8	8.85	124	57.8
GLY	309	β4 Sheet	8.23	102.74	45	8.25	102.67	45
VAL	310	β4 Sheet	7.48	124	61.96	7.4	124	61.93
GLN	311	β4 Sheet	8.29	126.3	57.13	8.34	126.3	56.78
LEU	312	β4 Sheet	8.57	128.75	54.2	8.57	128.7	54.2
LYS	313	β3β4 Loop	8.47	124.62	54.3	8.47	125.1	54.4
LYS	314	β3β4 Loop	8.66	122	59.3	8.7	121.6	59.3
GLY	315	β3 Sheet	8.677	117.5	44	8.56	117.63	45.9
ASP	316	β3 Sheet	8.86	123.36	56.3	8.82	123.11	56.2

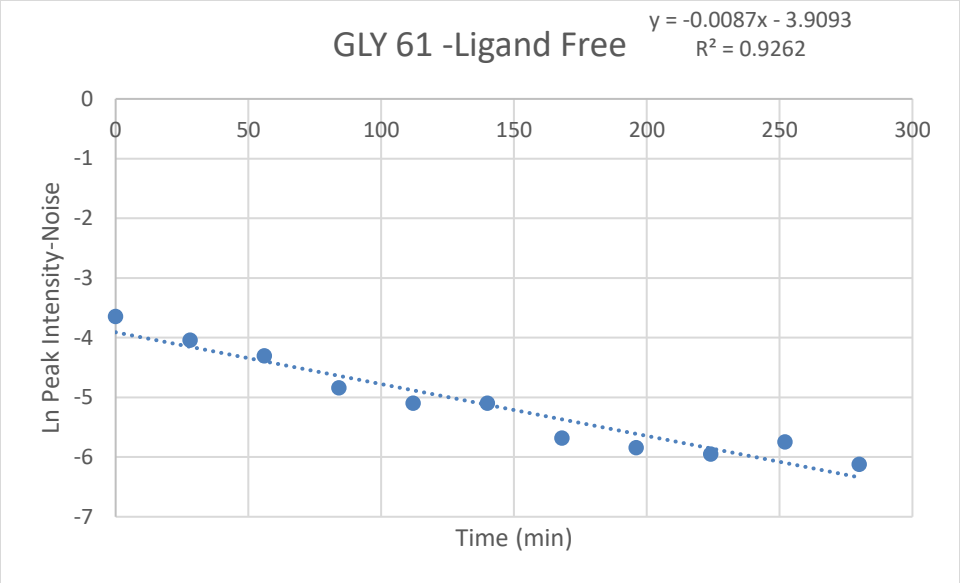
Table A.3 continued

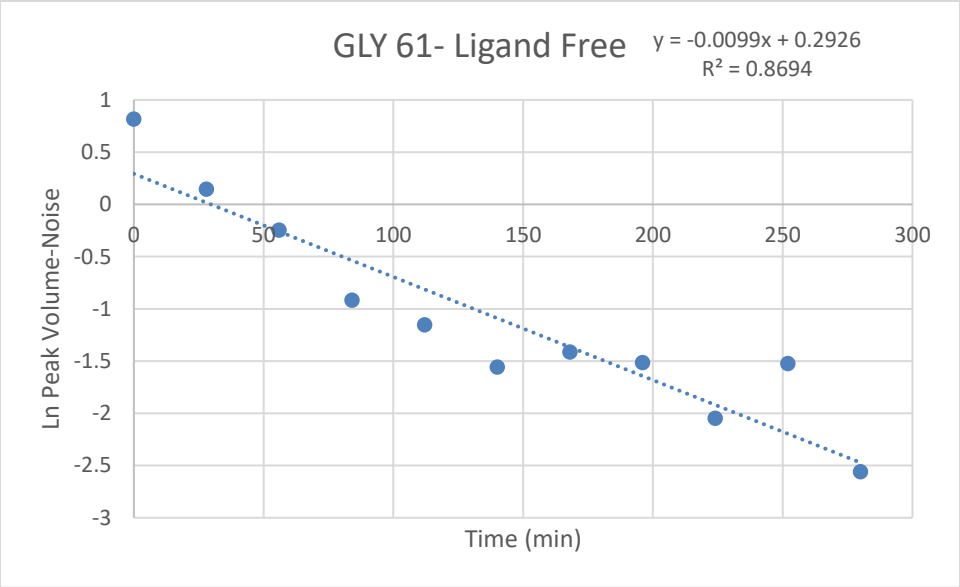
GLN	317	β3 Sheet	9.41	119.11	56.3	9.45	119.44	56.2
LEU	319	β3 Sheet	8.9	131.5	54.7	9.02	131.4	55
LEU	320	β3 Sheet	8.87	132.3	52.4	8.59	132.24	52.9
ALA	333	β3L loop	8.41	124.75	53.17	8.4	124.68	53.1
MET	336	β3L loop	8.68	116.78	58.5	8.71	116.5	58.5
HIS	337	β3L loop	8.05	123.77	56.2	8.05	124.2	56.1
VAL	338	β3L loop	8.04	126.5	61.5	8.04	126.95	62.3
ASP	339	β3L loop	10.09	131.56	52.5	10.08	131.66	52.3
PHE	340	β3L loop	9.4	124.85	52.5	9.33	125.1	52.7
SER	341	β3L loop	8.73	116.23	58.4	8.73	116.3	58.4
ARG	342	β3L loop	7.17	124.45	57.2	7.19	124.4	57.64
VAL	369	L Helix	7.15	109.9	60.5	7.3	113.1	61.5
LYS	372	L Helix	8.8	120.6		8.8	121	59
ASP	380	Lβ5 Loop	8.1	116.83	54	8.11	116.83	54
PHE	381	Lβ5 Loop	6.59	119.23	55.98	6.59	119.23	55.98
SER	382	β5 Sheet	9.02	111.67	57.5	9.01	111.85	57.56
ILE	383	β5 Sheet	8.41	122.8	61.38	8.49	123	61.45
ALA	384	β5 Sheet	8.16	131.99	51.97	8.09	131.96	51.7
GLY	386	β5 Sheet				7.32	109.762	44.5
ALA	387	β5 Sheet	7.62	124.2	52.8	7.63	123.866	53.3
GLN	388	β5 Sheet	8.42	123.13		8.38	122.89	55
ILE	389	β5 Sheet	8.58	127.3	58.9	8.57	127.08	59.7
GLN	390	β5 Sheet	9.5	130	54.6	9.48	129.78	55.1
HIS	391	β5 Sheet	9.01	125.7	53.3	8.82	125.43	54.1
LYS	392	β5 Sheet	8.97	120	56.2	8.86	119.56	55.3
SER	393	β5 Sheet	8.85	117.9	55.7	8.65	117.65	56
GLY	394	β5 Sheet	7.07	109.9	46.18	7.02	109.21	
VAL	396	β5 Sheet	6.13	115.2	62.2			62.47
GLY	398	β5 Sheet	8.68	103.5	44.5	8.59	103.09	46.4
VAL	399	β5 Sheet	8.03	119.6	62.6			63.3
GLN	400	β5 Sheet	8.9	126.5	53.3	9.0225	126.47	53.5
ALA	401	β5 Sheet	7.37	118.53	54.1	7.41	118.14	54
LEU	402	β5 Sheet	8.65	119.8	56.6	8.74	120.511	55.87
VAL	405	β5 Sheet	9.22	114.6	59.8	9.2	114.67	60.2
ALA	409	C term Loop	8.56	121.1	54	8.65	121.93	55
THR	410	C term Loop	8.09	108.32	62.4	8.01	108.32	62.5
LYS	412	C term Loop	8.8	121.3	55.3	8.92	121.22	55.87
ALA	413	C term Loop	7.97	125.8	51.7	8	125.6	52.1
VAL	414	C term Loop	7.88	127	64	7.83	126.86	64.6
	Total		149			139		

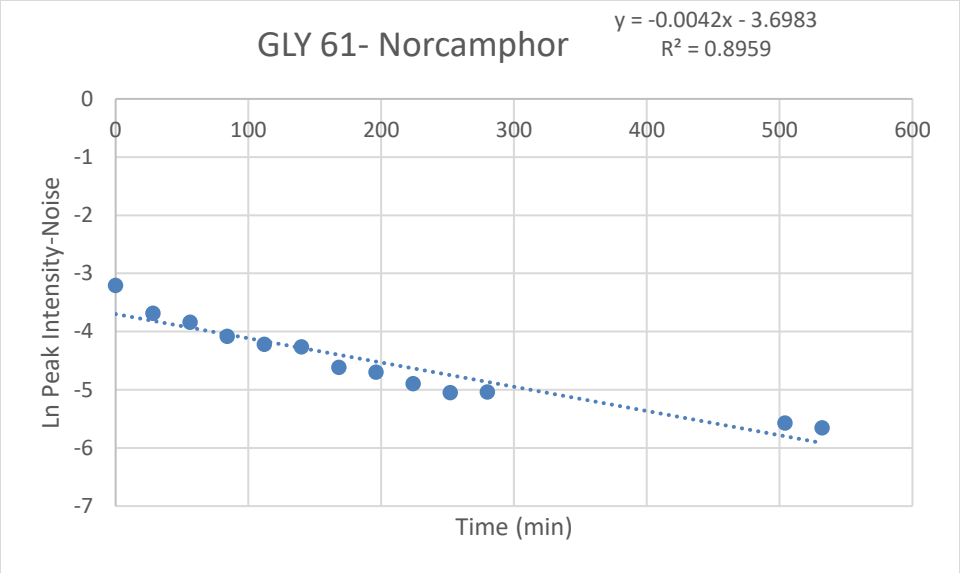
Table A.4 Example plots of determination of HDX exchange rates in various CYP101 forms using peak Intensities (top) and peak volumes (bottom) as a function of time for individual N-H amide peaks corresponding to residues in Cyp101

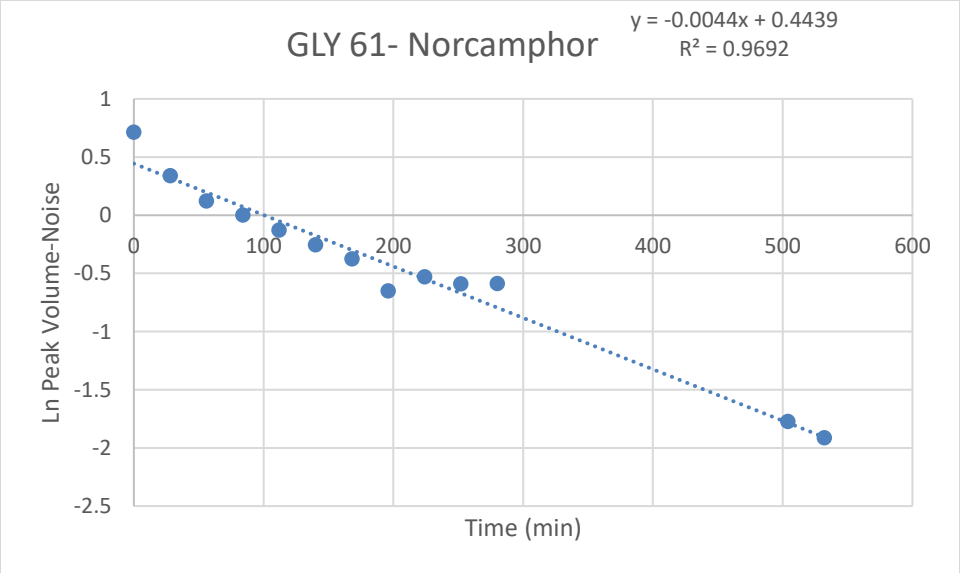


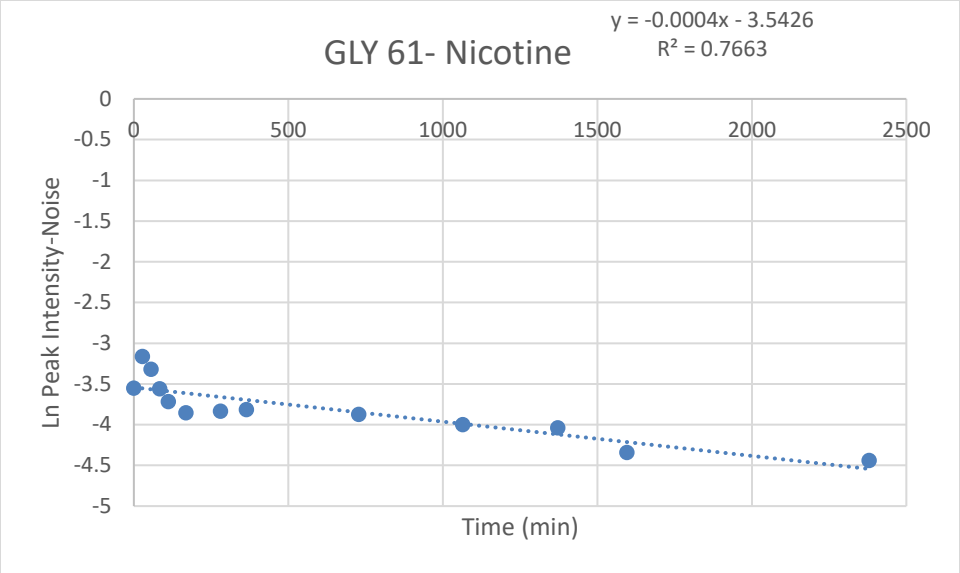


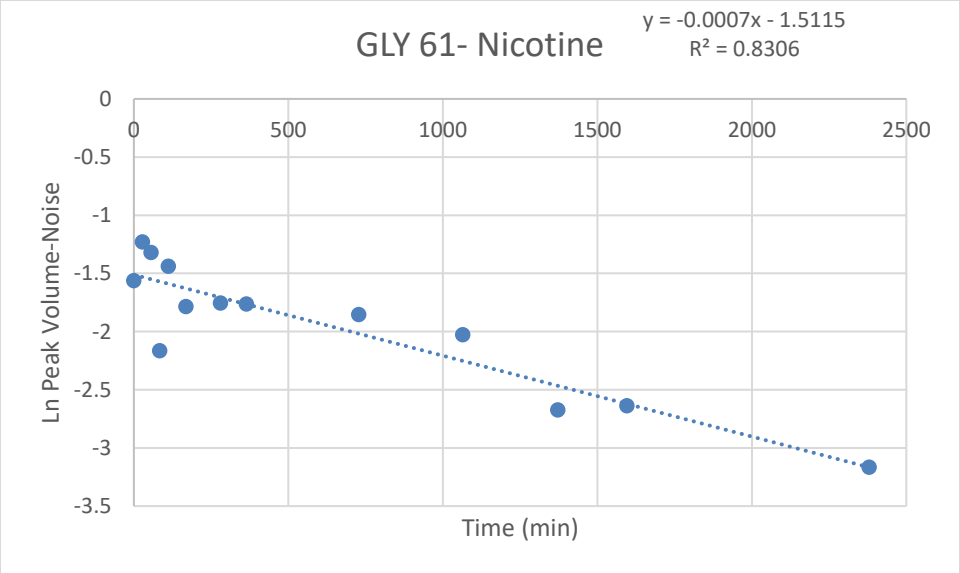

GLY 61 -Ligand Free
y = -0.0087x - 3.9093
R² = 0.9262
Ln Peak Intensity-Noise
Time (min)
0
-1
-2
-3
-4
-5
-6
-7
0
50
100
150
200
250
300

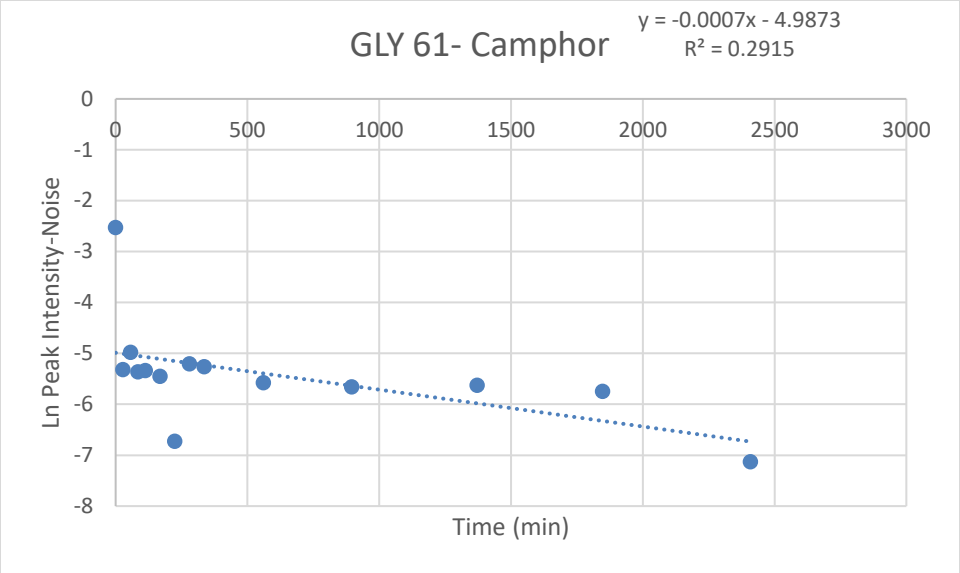

GLY 61- Ligand Free
y = -0.0099x + 0.2926
R² = 0.8694
Ln Peak Volume-Noise
Time (min)
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
0
50
100
150
200
250
300


GLY 61- Norcamphor
y = -0.0042x - 3.6983
R² = 0.8959
Ln Peak Intensity-Noise
Time (min)
0
-1
-2
-3
-4
-5
-6
-7
0
100
200
300
400
500
600


GLY 61- Norcamphor
y = -0.0044x + 0.4439
R² = 0.9692
Ln Peak Volume-Noise
Time (min)
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
100
200
300
400
500
600


GLY 61- Nicotine
y = -0.0004x - 3.5426
R² = 0.7663
Ln Peak Intensity-Noise
Time (min)
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
-5
0
500
1000
1500
2000
2500


GLY 61- Nicotine
y = -0.0007x - 1.5115
R² = 0.8306
Ln Peak Volume-Noise
Time (min)
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
0
500
1000
1500
2000
2500


GLY 61- Camphor
y = -0.0007x - 4.9873
R² = 0.2915
Ln Peak Intensity-Noise
Time (min)
0
-1
-2
-3
-4
-5
-6
-7
-8
0
500
1000
1500
2000
2500
3000

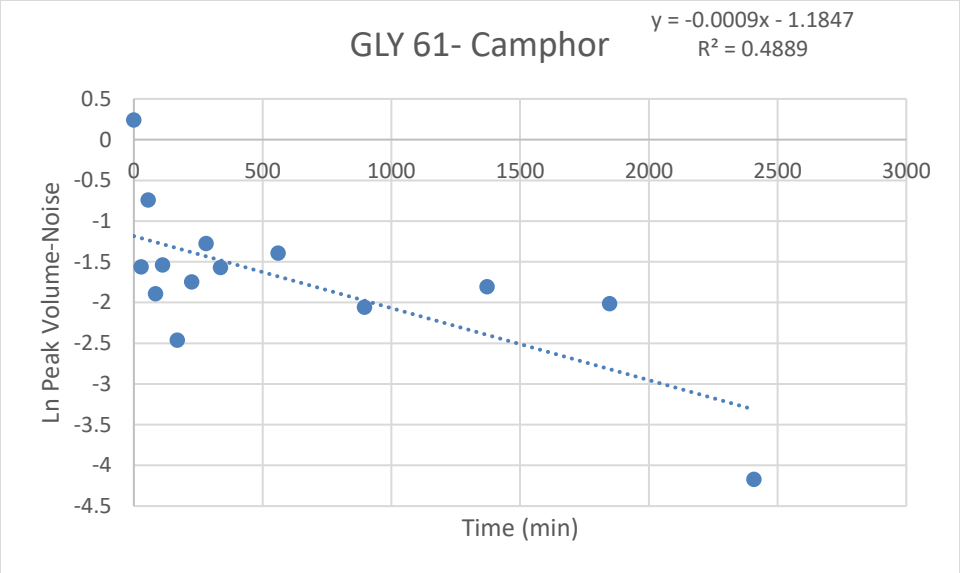

GLY 61- Camphor
y = -0.0009x - 1.1847
R² = 0.4889
Ln Peak Volume-Noise
Time (min)
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
0
500
1000
1500
2000
2500
3000

Table A.4 table of amide exchange rates for camphor and norcamphor

A (--) indicates no rates were measured for that residue. A (***) indicates rates could not be measured due to back exchange. Error values were calculated from the fitting error to the logarithmic plots

Type	Number	Secondary Structure	Camphor K (min^{-1})	Error +/-	Norcamphor K (min^{-1})	Error +/-
LEU	11	N term Loop	Fast		0.0011	0.000401
ALA	12	N term Loop	0.0009	0.000323	Fast	
HIS	17	N term Loop	Fast		Fast	
VAL	18	N term Loop	.0431	0.006694	0.0566	0.001175
LEU	22	N term Loop	0.0045	0.001365	--	
VAL	23	N term Loop	0.00003	0.000242	0.0024	0.0000916
PHE	24	N term Loop	0.0004	0.0000475	0.0013	0.000395
PHE	26	N term Loop	Fast		0.072	0.008915
ASP	27	N term Loop	0.085	0.000426	Fast	
TYR	29	N term Loop	0.0003	0.002252	0.0012	0.0003
ASN	30	N term Loop	.0021	0.00086	0.0026	0.00066
ALA	36	N term Loop	Fast		Fast	
GLY	37	A Helix	0.0001	0.000079	Fast	
ASN	49	Aβ1 Loop	Fast		0.0005	0.00026
VAL	50	Aβ1 Loop	0.0105	0.002722	Fast	
VAL	54	β1 Sheet	0.0001	0.000086	0.0011	0.000109
TRP	55	β1 Sheet	0.00007	0.000062	0.0009	0.000146
THR	56	β1 Sheet	0.0004	0.00022	0.0011	0.00033
CYS	58	**β1 Sheet**	0.0013	0.000309	Fast	
ASN	59	**β1 Sheet**	0.0438	0.006694	Fast	
GLY	60	**β1 Sheet**	Fast		Fast	
GLY	61	**β1 Sheet**	0.0007	0.000322	0.0015	0.000315
HIS	62	β1 Sheet	0.0002	0.000176	0.0008	0.000336
ILE	64	β1 Sheet	0.00006	0.000055	0.0011	0.000192
ALA	65	β1 Sheet	0.0001	0.000087	0.0012	0.000141
THR	66	β1 Sheet	0.00007	0.000054	0.0009	0.00038
ARG	67	B Helix	0.00007	0.000064	0.0009	0.000214
GLY	68	B Helix	Fast		Fast	
GLN	69	B Helix	0.001	0.000383	Fast	
LEU	70	B Helix	Fast		Fast	
CYS	85	BB' Loop	0.0086	0.003085	0.0317	0.006901
PHE	87	BB' Loop	0.001	0.000598	Fast	
ILE	88	BB' Loop	Fast		--	
ARG	90	**B' Helix**	.0055	0.004686	0.00005	0.000212
GLU	91	**B' Helix**	Fast		Fast	
ALA	92	**B' Helix**	Fast		0.0448	0.004001
GLY	93	**B' Helix**	Fast		0.0724	0.000003
GLU	94	**B' Helix**	0.0002	0.00017	0.00005	0.000226
ALA	95	**B' Helix**	Fast		0.0188	0.000163
TYR	96	**B' Helix**	0.1156	0.00078	Fast	
PHE	98	**B'C Loop**	Fast		--	
VAL	123	C Helix	0.0002	0.000164	0.001	0.000071
VAL	124	C Helix	Fast		0.0011	0.000591
ASP	125	C Helix	Fast		0.0005	0.000105
LYS	126	C Helix	0.0001	0.000087	0.0016	0.000282
SER	141	D Helix	0.054	0.010337	Fast	
GLN	145	D Helix	.016	0.010276	0.0099	0.001609
GLY	146	DE Loop	.069	0.003446	0.0176	0.001115
CYS	148	DE Loop	0.0008	0.00021	0.0007	0.000015
ASN	149	E Helix	Fast		0.0282	0.002004

Table A.4 Continued

PHE	150	E Helix	0.082	0.003755	0.0018	0.000135
THR	151	E Helix	Fast		0.0224	0.00040721
GLU	152	E Helix	Fast		--	
ASP	153	E Helix	0.0001	0.0000804	0.0003	0.0000342
LEU	165	E Helix	0.0007	0.00038	--	
LEU	166	E Helix	0.0002	0.000159	--	
ALA	167	E Helix	0.0001	0.0000848	0.0007	0.000355
GLY	168	E Helix	0.055	0.003549	Fast	
GLU	171	EF Loop	Fast		Fast	
GLU	172	EF Loop	Fast		Fast	
LYS	178	**F Helix**	0.0006	0.000283	0.0121	0.000163
TYR	179	**F Helix**	0.0114	0.000848	0.042	0.000158
THR	185	**F Helix**	0.0002	0.000158	Fast	
ASP	188	**FG Loop**	Fast		Fast	
GLY	189	**FG Loop**	0.1532	0.000768	Fast	
SER	190	**FG Loop**	Fast		0.0082	0.001163
MET	191	**FG Loop**	Fast		0.0014	0.000076
THR	192	**G Helix**	0.0016	0.000415	0.0018	0.000166
PHE	193	**G Helix**	Fast		Fast	
ALA	194	**G Helix**	0.153	0.001269	0.0001	0.0000338
LYS	197	**G Helix**	0.0007	0.000393	--	
GLU	198	**G Helix**	0.0021	0.001039	Fast	
ALA	199	**G Helix**	0.0008	0.000297	Fast	
LEU	200	**G Helix**	0.0001	0.000088	0.0012	0.000598
ILE	207	G Helix	Fast		0.0401	0.007972
ILE	208	G Helix	0.0005	0.000253	0.0003	0.0000637
LYS	214	G Helix	0.0004	0.000256	0.0006	0.000346
GLY	216	GH Loop	Fast		Fast	
ALA	219	H Helix	0.0001	0.000089	--	
ILE	220	H Helix	0.0002	0.000147	Fast	
VAL	223	H Helix	0.0006	0.000292	0.0005	0.0000136
ALA	224	H Helix	0.0003	0.000177	0.0024	0.0000461
ASN	225	H Helix	0.0065	0.00078	0.0115	0.000427
GLY	226	β2 Sheet	0.0557	0.004589	0.0643	0.000322
GLN	227	β2 Sheet	0.0002	0.00017	0.00007	0.000355
VAL	228	β2 Sheet	0.2052	0.100309	0.0012	0.000589
ASN	229	β2 Sheet	Fast		--	
GLY	230	β2 Sheet	Fast		Fast	
ARG	231	β2 Sheet	0.00002	0.000183	0.0011	0.000498
ILE	233	β2 Sheet	0.0016	0.000596	0.0057	0.000514
THR	234	I Helix	0.0215	0.002647	Fast	
ALA	265	I Helix	0.0005	0.000254	0.0007	0.000326
LYS	266	I Helix	Fast		0.0013	0.000521
SER	267	I Helix	.0375	0.00215	0.001	0.000608
ARG	271	J Helix	0.0015	0.000235	0.0009	0.000245
GLN	272	J Helix	0.0506	0.00261	0.0121	0.003301
GLU	273	J Helix	Fast		0.0012	0.00008
LEU	274	J Helix	.0017	0.000006	0.0012	0.000266
GLU	279	JK Loop	Fast		Fast	
ARG	280	K Helix	0.0101	0.001928	0.083	0.000269
ASP	304	β3β4 Loop	.0098	0.001587	0.0006	0.00027
TYR	305	β4 Sheet	0.0003	0.000198	0.0013	0.0000669
GLU	306	β4 Sheet	Fast		Fast	
PHE	307	β4 Sheet	0.0011	0.000402	0.001	0.000301
HIS	308	β4 Sheet	Fast		0.0007	0.000316

Table A.4 Continued

GLY	309	β4 Sheet	Fast		Fast	
VAL	310	β4 Sheet	0.0016	0.000141	0.0013	0.000239
GLN	311	β4 Sheet	0.0006	0.000332	0.0015	0.000442
LEU	312	β4 Sheet	0.00006	0.000056	0.001	0.000394
LYS	313	β3β4 Loop	0.00001	0.000087	0.0016	0.000266
LYS	314	β3β4 Loop	0.0004	0.000278	0.0008	0.000417
GLY	315	β3 Sheet	Fast		0.0555	0.000278
ASP	316	β3 Sheet	0.0002	0.000158	0.0018	0.000034
GLN	317	β3 Sheet	0.1553	0.071454	0.0295	0.0000383
LEU	319	β3 Sheet	0.00009	0.0000804	0.0019	0.001062
LEU	320	β3 Sheet	0.0003	0.000161	0.0015	0.000172
ALA	333	β3L loop	Fast		0.0018	0.00015
MET	336	β3L loop	Fast		Fast	
HIS	337	β3L loop	0.0004	0.000225	Fast	
VAL	338	β3L loop	Fast		Fast	
ASP	339	β3L loop	.0021	0.000568	0.0083	0.001849
PHE	340	β3L loop	Fast		Fast	
SER	341	β3L loop	Fast		0.0015	0.000528
ARG	342	β3L loop	Fast		0.1057	0.000539
VAL	369	L Helix	Fast		Fast	
LYS	372	L Helix	0.0008	0.000437	Fast	
ASP	380	Lβ5 Loop	0.0162	0.005133	0.0402	0.000459
PHE	381	Lβ5 Loop	0.0385	0.003347	0.0075	0.003713
SER	382	β5 Sheet	0.0002	0.000157	0.0003	0.000027
ILE	383	β5 Sheet	0.0017	0.0002	0.0049	0.000692
ALA	384	β5 Sheet	0.037	0.003004	0.0177	0.000303
GLY	386	β5 Sheet	Fast		Fast	
ALA	387	β5 Sheet	Fast		Fast	
GLN	388	β5 Sheet	0.0403	0.007308	Fast	
ILE	389	β5 Sheet	Fast		Fast	
GLN	390	β5 Sheet	Fast		Fast	
HIS	391	β5 Sheet	Fast		Fast	
LYS	392	β5 Sheet	Fast		0.0013	0.000289
SER	393	β5 Sheet	0.0002	0.000161	0.0882	0.000851
GLY	394	β5 Sheet	0.0006	0.000426	0.0143	0.008857
VAL	396	β5 Sheet	0.0598	0.0000445	--	
GLY	398	β5 Sheet	0.0002	0.000168	0.0045	0.000944
VAL	399	β5 Sheet	0.0015	0.00051	--	
GLN	400	β5 Sheet	0.00009	0.00008	0.0003	0.000064
ALA	401	β5 Sheet	0.0293	0.004793	Fast	
LEU	402	β5 Sheet	0.0006	0.000307	0.0013	0.000384
VAL	405	β5 Sheet	0.00003	.000002	0.00009	0.000358
ALA	409	C term Loop	0.0004	0.000247	0.0003	0.000118
THR	410	C term Loop	Fast		0.0798	0.0004
LYS	412	C term Loop	Fast		0.0011	0.00044
ALA	413	C term Loop	Fast		Fast	
VAL	414	C term Loop	Fast		Fast	
	Total		152		140	
	Fast		52		51	
	Intermediate		40		31	
	Slow		60		59	

Table A.5 table of amide exchange rates for Nicotine, Ligand free and Ketoconazole

A (--) indicates no rates were measured for that residue. A (***) indicates rates could not be measured due to back exchange. Error values were calculated from the fitting error to the logarithmic plots

Type	Number	Secondary Structure	Nicotine K (min^{-1})	Error +/-	Ligand Free K (min^{-1})	Error +/-	Keto K (min^{-1})	Error +/-
LEU	11	N term Loop	0.0004	0.000005	.0008	0.00059	0.0509	0.005379
ALA	12	N term Loop	Fast		Fast		Fast	
HIS	17	N term Loop	Fast		Fast		Fast	
VAL	18	N term Loop	.0455	0.002005	0.0644	0.001304	0.0494	0.001518
LEU	22	N term Loop	Fast		0.0064	0.000541	Fast	
VAL	23	N term Loop	0.0005	0.000041	0.0024	0.000033	0.0013	0.000180
PHE	24	N term Loop	.00007	0.000053	0.0024	0.00012	0.0011	0.000087
PHE	26	N term Loop	0.0003	0.000118	Fast		Fast	
ASP	27	N term Loop	Fast		Fast		Fast	
TYR	29	N term Loop	0.0001	0.000039	Fast		0.0001	0.000067
ASN	30	N term Loop	.0504	0.000234	0.0099	0.000455	0.0194	0.000155
ALA	36	N term Loop	Fast		Fast		0.0175	0.004320
GLY	37	A Helix	Fast		Fast		0.0915	0.001086
ASN	49	Aβ1 Loop	.000006	0.000005	0.0007	0.000226	0.0003	0.000091
VAL	50	Aβ1 Loop	0.0002	0.000176	0.46	0.002722	0.0311	0.002515
VAL	54	β1 Sheet	0.00005	0.000008	0.0024	0.00011	0.0009	0.000032
TRP	55	β1 Sheet	0.00003	0.000023	0.0495	0.002632	0.001	0.000121
THR	56	β1 Sheet	0.00004	0.000005	0.019	0.002578	0.0044	0.001085
CYS	58	**β1 Sheet**	Fast		0.0013	0.000274	0.0283	0.001085
ASN	59	**β1 Sheet**	Fast		Fast		Fast	
GLY	60	**β1 Sheet**	Fast		Fast		Fast	
GLY	61	**β1 Sheet**	0.00004	0.000005	0.0087	0.000327	0.0107	0.000200
HIS	62	β1 Sheet	0.00005	0.000034	0.0152	0.001798	0.0013	0.000200
ILE	64	β1 Sheet	***		0.0032	0.000168	0.0008	0.000163
ALA	65	β1 Sheet	0.000008	0.000007	0.0042	0.00018	0.0016	0.000587
THR	66	β1 Sheet	0.00001	0.000009	0.0025	0.000156	0.0009	0.000108
ARG	67	B Helix	0.00001	0.000034	0.0049	0.000038	0.0017	0.000108
GLY	68	B Helix	Fast		Fast		Fast	
GLN	69	B Helix	0.0014	0.000139	0.0012	0.000621	0.04888	0.000299
LEU	70	B Helix	Fast		Fast		Fast	
CYS	85	BB' Loop	0.00005	0.000044	0.0004	0.000288	Fast	
PHE	87	BB' Loop	0.00001	0.000007	--		--	
ILE	88	BB' Loop	0.0002	0.000153	--		--	
ARG	90	**B' Helix**	Fast		Fast		Fast	
GLU	91	**B' Helix**	Fast		Fast		--	
ALA	92	**B' Helix**	0.0442	0.002878	--		--	
GLY	93	**B' Helix**	Fast		Fast		Fast	
GLU	94	**B' Helix**	0.00001	0.000009	0.0004	0.000073	0.00005	0.000029
ALA	95	**B' Helix**	Fast		0.0055	0.002215	Fast	
TYR	96	**B' Helix**	Fast		Fast		Fast	
PHE	98	**B'C Loop**	0.000005	0.000003	--		--	
VAL	123	C Helix	0.0006	0.000392	0.0173	0.000172	Fast	
VAL	124	C Helix	0.0007	0.000459	Fast		0.0224	0.000553
ASP	125	C Helix	0.0002	0.000007	--		0.0019	0.000387
LYS	126	C Helix	Fast		0.0036	0.000885	0.0052	0.000736
SER	141	D Helix	0.0015	0.000011	Fast		Fast	
GLN	145	D Helix	Fast		0.088	0.000441	0.0004	0.000218

Table A.5 Continued

GLY	146	DE Loop	0.0504	0.006732	0.0145	0.00005	0.0217	0.000004
GLN	147	DE Loop	0.0012	0.000043	0.0012	0.000009	0.0011	0.000027
CYS	148	DE Loop	0.0011	0.000159	0.0006	0.000024	0.0005	0.000025
ASN	149	E Helix	Fast		0.0196	0.013719	0.0076	0.000004
PHE	150	E Helix	0.0045	0.000466	0.0025	0.00006	0.0557	0.001953
THR	151	E Helix	Fast		0.0015	0.000103	0.0085	0.000126
GLU	152	E Helix	0.041	0.001837	0.0014	0.000814	--	
ASP	153	E Helix	***		0.0007	0.000026	0.0004	0.000215
LEU	165	E Helix	--		0.004	0.000263	0.0005	0.000123
LEU	166	E Helix	Fast		--		Fast	
ALA	167	E Helix	0.00007	0.000054	0.0013	0.000031	0.0112	0.002296
GLY	168	E Helix	0.0538	0.000003	0.0068	0.000003	0.04275	0.000009
GLU	171	EF Loop	Fast		Fast		0.1093	0.001226
GLU	172	EF Loop	Fast		Fast		0.0009	0.000553
LYS	178	**F Helix**	Fast		0.004	0.000394	0.0672	0.000003
TYR	179	**F Helix**	0.008	0.000486	--		--	
THR	185	**F Helix**	0.0008	0.000368	0.026	0.003589	Fast	
ASP	188	**FG Loop**	0.0057	0.005700	0.0131	0.000063	0.0005	0.000181
GLY	189	**FG Loop**	Fast		0.0494	0.000188	Fast	
SER	190	**FG Loop**	0.0009	0.000577	0.0007	0.000443	0.038	0.001360
MET	191	**FG Loop**	0.0018	0.000371	Fast		0.0001	0.000066
THR	192	**G Helix**	0.0018	0.000040	Fast		0.041	0.002786
PHE	193	**G Helix**	Fast		--		--	
ALA	194	**G Helix**	0.00003	0.000023	0.014	0.001610	Fast	
LYS	197	**G Helix**	0.0001	0.000041	--		Fast	
GLU	198	**G Helix**	0.00006	0.000048	0.0233	0.000172	0.0012	0.000088
ALA	199	**G Helix**	0.0017	0.000586	Fast		Fast	
LEU	200	**G Helix**	0.0002	0.000174	0.0158	0.002579	0.0025	0.000685
ILE	207	G Helix	0.0013	0.000559	0.0007	0.000669	0.0049	0.000958
ILE	208	G Helix	0.00006	0.000051	0.0011	0.00006	0.0004	0.000230
LYS	214	G Helix	0.00003	0.000025	Fast		0.0054	0.001433
GLY	216	GH Loop	Fast		Fast		Fast	
ALA	219	H Helix	0.0001	0.000074	--		--	
ILE	220	H Helix	0.0002	0.000066	Fast		Fast	
VAL	223	H Helix	0.00003	0.000026	0.0109	0.000942	0.0027	0.000089
ALA	224	H Helix	0.00005	0.000039	0.0671	0.000003	0.0162	0.000740
ASN	225	H Helix	--		--		--	
GLY	226	β2 Sheet	0.0982	0.000005	0.0025	0.000122	Fast	
GLN	227	β2 Sheet	0.00007	0.000043	0.0184	0.00043	0.0002	0.000103
VAL	228	β2 Sheet	0.0327	0.000002	Fast		Fast	
ASN	229	β2 Sheet	0.00003	0.000030	--		--	
GLY	230	β2 Sheet	Fast		0.0038	0.000425	Fast	
ARG	231	β2 Sheet	0.00007	0.000059	0.0016	0.00042	0.0003	0.000037
ILE	233	β2 Sheet	0.0013	0.000139	0.0889	0.000045	Fast	
THR	234	I Helix	0.0002	0.000087	0.0146	0.000836	Fast	
ALA	265	I Helix	Fast		0.0004	0.000006	0.014	0.000353
LYS	266	I Helix	0.00002	0.000014	0.0025	0.000122	0.0003	0.000196
SER	267	I Helix	0.0012	0.000216	0.0017	0.001212	0.0011	0.000061
ARG	271	J Helix	Fast		0.0016	0.000198	0.0017	0.000111
GLN	272	J Helix	0.047	0.000366	0.0008	0.000062	0.0003	0.000199
GLU	273	J Helix	0.0002	0.000046	Fast		Fast	
LEU	274	J Helix	0.0013	0.000072	0.0639	0.004083	0.0005	0.000154
GLU	279	JK Loop	Fast		Fast		Fast	
ARG	280	K Helix	0.1245	0.000730	0.002	0.000146	0.001	0.000397
ASP	304	β3β4 Loop	Fast		0.0012	0.000068	0.0012	0.000462

Table A.5 Continued

TYR	305	β4 Sheet	0.0001	0.000027	0.0023	0.000131	0.0008	0.000209
GLU	306	β4 Sheet	Fast		Fast		Fast	
PHE	307	β4 Sheet	0.0013	0.000068	0.0032	0.000040	0.0021	0.000042
HIS	308	β4 Sheet	Fast		Fast		Fast	
GLY	309	β4 Sheet	Fast		Fast		Fast	
VAL	310	β4 Sheet	0.0021	0.000047	0.0038	0.000045	0.0027	0.000024
GLN	311	β4 Sheet	0.0263	0.002607	0.0125	0.003048	Fast	
LEU	312	β4 Sheet	***		0.0022	0.000104	0.0014	0.000104
LYS	313	β3β4 Loop	0.00003	0.000018	0.0594	0.003687	0.0013	0.000046
LYS	314	β3β4 Loop	0.00005	0.000040	0.00119	0.000040	0.0003	0.000105
GLY	315	β3 Sheet	Fast		0.0197	0.003799	Fast	
ASP	316	β3 Sheet	***		0.019	0.000095	0.0017	0.000536
GLN	317	β3 Sheet	0.0251	0.000013	Fast		Fast	
LEU	319	β3 Sheet	0.00006	0.000059	Fast		0.0011	0.000705
LEU	320	β3 Sheet	0.0001	0.000080	0.0044	0.000615	0.0008	0.000105
ALA	333	β3L loop	Fast		0.0003	0.000244	Fast	
MET	336	β3L loop	Fast		0.0017	0.00016	0.0064	0.001137
HIS	337	β3L loop	Fast		Fast		Fast	
VAL	338	β3L loop	Fast		0.023	0.004431	Fast	
ASP	339	β3L loop	0.0014	0.000241	0.0323	0.001871	0.0754	0.000004
PHE	340	β3L loop	Fast		Fast		Fast	
SER	341	β3L loop	Fast		0.0023	0.000187	0.002	0.000074
ARG	342	β3L loop	0.0503	0.008999	0.089	0.000005	Fast	
VAL	369	L Helix	0.0501	0.002421	Fast		Fast	
LYS	372	L Helix	0.0001	0.000049	0.0017	0.000224	Fast	
ASP	380	Lβ5 Loop	0.0648	0.002731	0.0013	0.000159	0.0003	0.000112
PHE	381	Lβ5 Loop	Fast		0.0064	0.000281	0.0048	0.001009
SER	382	β5 Sheet	0.000004	0.000004	0.0002	0.000169	0.0008	0.000182
ILE	383	β5 Sheet	0.0018	0.000431	0.0015	0.000416	0.0005	0.000175
ALA	384	β5 Sheet	0.0263	0.000001	0.016	0.000175	0.0271	0.000889
GLY	386	β5 Sheet	Fast		Fast		Fast	
ALA	387	β5 Sheet	0.1314	0.000007	0.1035	0.000005	Fast	
GLN	388	β5 Sheet	0.0281	0.000863	Fast		0.0005	0.000161
ILE	389	β5 Sheet	Fast		0.0132	0.002031	Fast	
GLN	390	β5 Sheet	Fast		Fast		Fast	
HIS	391	β5 Sheet	Fast		Fast		Fast	
LYS	392	β5 Sheet	0.0002	0.000061	0.0195	0.005366	0.0012	0.000136
SER	393	β5 Sheet	0.00007	0.000040	0.014	0.000246	0.0001	0.000045
GLY	394	β5 Sheet	0.00002	0.000005	0.026	0.003589	0.0074	0.000177
VAL	396	β5 Sheet	0.0664	0.000003	--		--	
GLY	398	β5 Sheet	0.00004	0.000015	0.0134	0.001630	0.0408	0.003089
VAL	399	β5 Sheet	Fast		--		--	
GLN	400	β5 Sheet	***		0.0006	0.000034	0.0001	0.000056
ALA	401	β5 Sheet	Fast		--		0.0011	0.000507
LEU	402	β5 Sheet	0.0002	0.000071	0.0008	0.000123	0.0007	0.000068
VAL	405	β5 Sheet	0.00001	0.000008	0.0002	0.000033	0.0001	0.000037
ALA	409	C term Loop	Fast		0.0035	0.00004	0.0003	0.000050
THR	410	C term Loop	Fast		Fast		Fast	
LYS	412	C term Loop	0.00004	0.000017	0.0004	0.000291	Fast	
ALA	413	C term Loop	0.000004	0.000004	0.0002	0.000186	0.0306	0.011744
VAL	414	C term Loop	Fast		0.0007	0.000435	0.0567	0.000473
Total			149		138		139	
Fast			49		40		49	
Intermediate			27		72		50	
Slow			73		26		40	

LIST OF REFERENCES

1. Wei, L., C.W. Locuson, and T.S. Tracy, *Polymorphic variants of CYP2C9: mechanisms involved in reduced catalytic activity.* Molecular pharmacology, 2007. **72**(5): p. 1280-1288.
2. Glembo, T.J., et al., *Collective dynamics differentiates functional divergence in protein evolution.* PLoS computational biology, 2012. **8**(3).
3. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**(7172): p. 964-972.
4. Petit, C., *Fast Side-Chain Dynamics: Measurement, Analysis, and Their Role in Protein Function.* Biochem Physiol S, 2013. **2**: p. 2.
5. Kovermann, M., P. Rogne, and M. Wolf-Watz, *Protein dynamics and function from solution state NMR spectroscopy.* Quarterly reviews of biophysics, 2016. **49**.
6. Piston, D.W. and G.-J. Kremers, *Fluorescent protein FRET: the good, the bad and the ugly.* Trends in biochemical sciences, 2007. **32**(9): p. 407-414.
7. Lai, W.-J.C. and D.N. Ermolenko, *Ensemble and single-molecule FRET studies of protein synthesis.* Methods, 2018. **137**: p. 37-48.
8. Hinterdorfer, P. and A. Van Oijen, *Handbook of single-molecule biophysics*. 2009: Springer Science & Business Media.
9. Van Den Bedem, H. and J.S. Fraser, *Integrative, dynamic structural biology at atomic resolution—it's about time.* Nature methods, 2015. **12**(4): p. 307.
10. Schotte, F., et al., *Picosecond time-resolved X-ray crystallography: probing protein function in real time.* Journal of structural biology, 2004. **147**(3): p. 235-246.
11. Mazur, A.K., *Hierarchy of fast motions in protein dynamics.* The Journal of Physical Chemistry B, 1998. **102**(2): p. 473-479.
12. Klug, C.S. and J.B. Feix, *Methods and applications of site-directed spin labeling EPR spectroscopy.* Methods in cell biology, 2008. **84**: p. 617-658.
13. Thomas, D.D., *Large-scale rotational motions of proteins detected by electron paramagnetic resonance and fluorescence.* Biophysical journal, 1978. **24**(2): p. 439-462.
14. Kleckner, I.R. and M.P. Foster, *An introduction to NMR-based approaches for measuring protein dynamics.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2011. **1814**(8): p. 942-968.
15. Garcia, R.A., D. Pantazatos, and F.J. Villarreal, *Hydrogen/deuterium exchange mass spectrometry for investigating protein-ligand interactions.* Assay and drug development technologies, 2004. **2**(1): p. 81-91.
16. Behroozmand, A.A., K. Keating, and E. Auken, *A review of the principles and applications of the NMR technique for near-surface characterization.* Surveys in geophysics, 2015. **36**(1): p. 27-85.
17. McDonald, P., et al., *Surface relaxation and chemical exchange in hydrating cement pastes: a two-dimensional NMR relaxation study.* Physical Review E, 2005. **72**(1): p. 011409.
18. Nelson, D.R., *Cytochrome P450 diversity in the tree of life.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2018. **1866**(1): p. 141-154.

19. Danielson, P.á., *The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans.* Current drug metabolism, 2002. **3**(6): p. 561-597.
20. Meunier, B., S.P. De Visser, and S. Shaik, *Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes.* Chemical reviews, 2004. **104**(9): p. 3947-3980.
21. Jun, X., X.-y. WANG, and W.-z. GUO, *The cytochrome P450 superfamily: Key players in plant development and defense.* Journal of Integrative Agriculture, 2015. **14**(9): p. 1673-1686.
22. Koprivova, A., et al., *Root-specific camalexin biosynthesis controls the plant growth-promoting effects of multiple bacterial strains.* Proceedings of the National Academy of Sciences, 2019. **116**(31): p. 15735-15744.
23. Cryle, M.J., J.E. Stok, and J.J. De Voss, *Reactions catalyzed by bacterial cytochromes P450.* Australian journal of chemistry, 2003. **56**(8): p. 749-762.
24. Guengerich, F.P., *Cytochrome P450s and other enzymes in drug metabolism and toxicity.* The AAPS journal, 2006. **8**(1): p. E101-E111.
25. Guengerich, F.P., *Cytochrome P450 enzymes in the generation of commercial products.* Nature Reviews Drug Discovery, 2002. **1**(5): p. 359-366.
26. Hannemann, F., et al., *Cytochrome P450 systems—biological variations of electron transport chains.* Biochimica et Biophysica Acta (BBA)-General Subjects, 2007. **1770**(3): p. 330-344.
27. Lewis, D.F., *57 varieties: the human cytochromes P450.* Pharmacogenomics, 2004. **5**(3): p. 305-318.
28. Williams, J.A., et al., *Comparative metabolic capabilities of CYP3A4, CYP3A5, and CYP3A7.* Drug Metabolism and Disposition, 2002. **30**(8): p. 883-891.
29. Keshava, C., E.C. McCanlies, and A. Weston, *CYP3A4 polymorphisms—potential risk factors for breast and prostate cancer: a HuGE review.* American journal of epidemiology, 2004. **160**(9): p. 825-841.
30. Hines, R.N. and D.G. McCarver, *The ontogeny of human drug-metabolizing enzymes: phase I oxidative enzymes.* Journal of Pharmacology and Experimental Therapeutics, 2002. **300**(2): p. 355-360.
31. Sim, S.C. and M. Ingelman-Sundberg, *Update on allele nomenclature for human cytochromes P450 and the Human Cytochrome P450 Allele (CYP-allele) Nomenclature Database*, in *Cytochrome P450 Protocols*. 2013, Springer. p. 251-259.
32. Padmanabhan, S., *Pharmacogenomics and stratified medicine.* 2014.
33. Poulos, T.L., *Cytochrome P450 flexibility.* Proceedings of the National Academy of Sciences, 2003. **100**(23): p. 13121-13122.
34. Poulos, T.L., B.C. Finzel, and A.J. Howard, *High-resolution crystal structure of cytochrome P450cam.* J Mol Biol, 1987. **195**(3): p. 687-700.
35. Zhang, W., et al., *Solution NMR structure of putidaredoxin–cytochrome P450cam complex via a combined residual dipolar coupling–spin labeling approach suggests a role for Trp106 of putidaredoxin in complex formation.* Journal of molecular biology, 2008. **384**(2): p. 349-363.
36. Nagano, S. and T.L. Poulos, *Crystallographic Study on the Dioxygen Complex of Wild-type and Mutant Cytochrome P450cam IMPLICATIONS FOR THE*

DIOXYGEN ACTIVATION MECHANISM. Journal of Biological Chemistry, 2005. **280**(36): p. 31659-31663.
37. Ahalawat, N. and J. Mondal, *Mapping the substrate recognition pathway in cytochrome P450.* Journal of the American Chemical Society, 2018. **140**(50): p. 17743-17752.
38. Hasemann, C.A., et al., *Structure and function of cytochromes P450: a comparative analysis of three crystal structures.* Structure, 1995. **3**(1): p. 41-62.
39. Lüdemann, S.K., V. Lounnas, and R.C. Wade, *How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms.* Journal of molecular biology, 2000. **303**(5): p. 797-811.
40. Lee, Y.-T., et al., *P450cam visits an open conformation in the absence of substrate.* Biochemistry, 2010. **49**(16): p. 3412-3419.
41. Thielges, M.C., J.K. Chung, and M.D. Fayer, *Protein dynamics in cytochrome P450 molecular recognition and substrate specificity using 2D IR vibrational echo spectroscopy.* Journal of the American Chemical Society, 2011. **133**(11): p. 3995-4004.
42. Seifert, A., et al., *Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: the molecular basis of substrate binding and regioselectivity toward warfarin.* Proteins: Structure, Function, and Bioinformatics, 2006. **64**(1): p. 147-155.
43. Hendrychová, T., et al., *Flexibility of human cytochrome P450 enzymes: molecular dynamics and spectroscopy reveal important function-related variations.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2011. **1814**(1): p. 58-68.
44. Asciutto, E.K., et al., *Structural and Dynamic Implications of an Effector-induced Backbone Amide cis–trans Isomerization in Cytochrome P450cam.* Journal of molecular biology, 2009. **388**(4): p. 801-814.
45. Hamuro, Y., et al., *Hydrogen–deuterium exchange mass spectrometry for investigation of backbone dynamics of oxidized and reduced cytochrome P450cam.* Journal of inorganic biochemistry, 2008. **102**(2): p. 364-370.
46. Miao, Y., et al., *Temperature-dependent dynamical transitions of different classes of amino acid residue in a globular protein.* Journal of the American Chemical Society, 2012. **134**(48): p. 19576-19579.
47. Hong, L., et al., *Determination of functional collective motions in a protein at atomic resolution using coherent neutron scattering.* Science advances, 2016. **2**(10): p. e1600886.
48. Schiffler, B. and R. Bernhardt, *Bacterial (CYP101) and mitochondrial P450 systems—how comparable are they?* Biochemical and biophysical research communications, 2003. **312**(1): p. 223-228.
49. Van Vleet, T.R., D.W. Bombick, and R.A. Coulombe Jr, *Inhibition of human cytochrome P450 2E1 by nicotine, cotinine, and aqueous cigarette tar extract in vitro.* Toxicological Sciences, 2001. **64**(2): p. 185-191.
50. Yamazaki, H., et al., *Roles of CYP2A6 and CYP2B6 in nicotine C-oxidation by human liver microsomes.* Archives of toxicology, 1999. **73**(2): p. 65-70.

51. Ekroos, M. and T. Sjögren, *Structural basis for ligand promiscuity in cytochrome P450 3A4.* Proceedings of the National Academy of Sciences, 2006. **103**(37): p. 13682-13687.
52. Cupp-Vickery, J.R., et al., *Ketoconazole-induced conformational changes in the active site of cytochrome P450eryF.* Journal of molecular biology, 2001. **311**(1): p. 101-110.
53. Strushkevich, N., S.A. Usanov, and H.-W. Park, *Structural basis of human CYP51 inhibition by antifungal azoles.* Journal of molecular biology, 2010. **397**(4): p. 1067-1078.
54. Gaudineau, C. and K. Auclair, *Inhibition of human P450 enzymes by nicotinic acid and nicotinamide.* Biochemical and biophysical research communications, 2004. **317**(3): p. 950-956.
55. Raag, R. and T.L. Poulos, *The structural basis for substrate-induced changes in redox potential and spin equilibrium in cytochrome P-450CAM.* Biochemistry, 1989. **28**(2): p. 917-922.
56. Jones, J.P., W.F. Trager, and T.J. Carlson, *The binding and regioselectivity of reaction of (R)-and (S)-nicotine with cytochrome P-450cam: parallel experimental and theoretical studies.* Journal of the American Chemical Society, 1993. **115**(2): p. 381-387.
57. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of biomolecular NMR, 1995. **6**(3): p. 277-293.
58. Johnson, B., *NMRViewJ, version 8.2. 36, One Moon Scientific.* Inc., Newark, NJ, 2010.
59. Skinner, S.P., et al., *Delicate conformational balance of the redox enzyme cytochrome P450cam.* Proceedings of the National Academy of Sciences, 2015. **112**(29): p. 9022-9027.
60. Pochapsky, S.S., T.C. Pochapsky, and J.W. Wei, *A model for effector activity in a highly specific biological electron transfer complex: The cytochrome P450cam−putidaredoxin couple.* Biochemistry, 2003. **42**(19): p. 5649-5656.
61. Hiruma, Y., et al., *The structure of the cytochrome P450cam–putidaredoxin complex determined by paramagnetic NMR spectroscopy and crystallography.* Journal of molecular biology, 2013. **425**(22): p. 4353-4365.
62. Dominguez, C., R. Boelens, and A.M. Bonvin, *HADDOCK: a protein− protein docking approach based on biochemical or biophysical information.* Journal of the American Chemical Society, 2003. **125**(7): p. 1731-1737.
63. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis.* Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.
64. Vermeer, L.M., et al., *Evaluation of ketoconazole and its alternative clinical CYP3A4/5 inhibitors as inhibitors of drug transporters: the in vitro effects of ketoconazole, ritonavir, clarithromycin, and itraconazole on 13 clinically-relevant drug transporters.* Drug Metabolism and Disposition, 2016. **44**(3): p. 453-459.
65. Chenge, J.T., et al., *Structural characterization and ligand/inhibitor identification provide functional insights into the Mycobacterium tuberculosis cytochrome P450 CYP126A1.* Journal of Biological Chemistry, 2017. **292**(4): p. 1310-1329.

66. Montemiglio, L.C., et al., *Azole drugs trap cytochrome P450 EryK in alternative conformational states.* Biochemistry, 2010. **49**(43): p. 9199-9206.
67. Bernal Gomez, A.V., *Conformational Dynamics of Cytochrome P450cam Upon Ligand Binding.* 2015.
68. Colthart, A.M., et al., *Detection of substrate-dependent conformational changes in the P450 fold by nuclear magnetic resonance.* Scientific reports, 2016. **6**(1): p. 1-11.
69. Follmer, A.H., et al., *Substrate-dependent allosteric regulation in cytochrome P450cam (CYP101A1).* Journal of the American Chemical Society, 2018. **140**(47): p. 16222-16228.
70. Scheuermann, T.H. and C.A. Brautigam, *High-precision, automated integration of multiple isothermal titration calorimetric thermograms: new features of NITPIC.* Methods, 2015. **76**: p. 87-98.
71. Zhao, H., G. Piszczek, and P. Schuck, *SEDPHAT–a platform for global ITC analysis and global multi-method analysis of molecular interactions.* Methods, 2015. **76**: p. 137-148.
72. Brautigam, C.A., et al., *Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions.* Nature protocols, 2016. **11**(5): p. 882.
73. Fox, J.M., et al., *The molecular origin of enthalpy/entropy compensation in biomolecular recognition.* Annual review of biophysics, 2018. **47**: p. 223-250.
74. Muralidhara, B., et al., *Conformational Flexibility of Mammalian Cytochrome P450 2B4 in Binding Imidazole Inhibitors with Different Ring Chemistry and Side Chains SOLUTION THERMODYNAMICS AND MOLECULAR MODELING.* Journal of Biological Chemistry, 2006. **281**(12): p. 8051-8061.
75. Fenley, A.T., H.S. Muddana, and M.K. Gilson, *Entropy–enthalpy transduction caused by conformational shifts can obscure the forces driving protein–ligand binding.* Proceedings of the National Academy of Sciences, 2012. **109**(49): p. 20006-20011.
76. Bhojane, P.P., et al., *Investigation of osmolyte effects on FolM: comparison with other dihydrofolate reductases.* Biochemistry, 2014. **53**(8): p. 1330-1341.

VITA

Rebecca Shrem was born in Michigan, but has spent most of her life living outside of Nashville, TN. She attended the University of Tennessee, Knoxville for her undergraduate studies, receiving a BS in Biology with a concentration in Biochemistry, Cellular, and Molecular Biology. She continued her journey at the University of Tennessee for her graduate studies with the completion of this Master's degree. Outside of the protein work, Rebecca has taken a large interest in discipline-based education research and hopes to continue working on biology education research in the future.