Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the Degree of
Doctor of Natural Sciences

presented by

Xiaofeng Yu
born in: Longkou, Shandong, China
Oral examination date: 03.07.2015

Multiscale Simulations of Cytochrome P450 Systems

Referees:

Prof. Dr. Rebecca Wade

Prof. Dr. Robert Russell

Acknowledgement

I gratefully acknowledge the financial support of the Klaus Tschira foundation and the support of the Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology (HBIGS).

I thank the Molecular and Cellular Modeling (MCM) group at the Heidelberg Institute for Theoretical Studies (HITS) for scientific help during my doctoral project and especially Prof. Rebecca Wade for her excellent supervision. I thank the kind help of members of the MCM group: Dr. Jonathan Fuller, Dr. Neil Bruce, Dr. Michael Martinez, Dr. Stefan Richter, Ghulam Mustafa, Prajwal Nandekar, Antonia Stank and Ina Pöhner. Also I thank the HBIGS office for the kind support from Dr. Rolf Lutz.

I thank our collaborators, namely Dr. Vlad Cojocaru at the Max Planck Institute for Molcular Biomedicine in Münster, Prof. Galina Lepesheva at the Vanderbilt University in Nashville, Tennessee and Dr. Outi Salo-Ahen at the Åbo Akademi University in Turku, Finland.

A special thanks to my family, for their love which supports me to finish this thesis.

Abstract

Cytochrome P450 enzymes (CYP, P450) play a crucial role in drug metabolism and sterol biosynthesis. In humans, P450s catalyze more than 80% of drugs. P450s such as CYP51 are also involved in steroid biosynthesis. CYP51 catalyzes the 14α-demethylation of sterol-like molecules. CYP51 has been found to exist in all kingdoms. It has been used for antifungal drug design. It has also been shown to be a good target for antiparasitic drug design. Understanding the mechanism of P450s not only helps to invesigate drug metabolism but also to design parasite specific inhibitors that do not inhibit human orthologs.

P450s are heme-containing enzymes. The active site is buried in the protein and thus, substrates need to enter the active site and products need to exit from the active site via ligand tunnels. P450s are membrane-bound proteins. They anchor in the membrane with a single transmembrane helix. For the catalytic reaction of P450s, two electrons need to be transferred from their redox partners. In eukaryotes, the most often used redox partner of P450s is cytochrome P450 reductase (CPR). CPR is also a membrane-bound protein. The association between P450 and CPR is driven by electrostatics. Two electrons are transferred from the flavin mononucleotide (FMN) cofactor of CPR to the heme center of P450s.

Current experimental evidence does not provide a full understanding of dynamics of P450s, differences between P450s, of how P450s embed in the membrane, of how ligands can enter and exit from the active site and of how P450s interaction with CPR. A multiscale computational approach, including all-atom molecular dynamics simulations, coarse-grain molecular dynamics simulations and Brownian dynamics simulations was performed to tackle these problems.

A model of membrane-bound *T. brucei* CYP51 was built and the model was shown to be consistent with existing experimental data. The dynamics of the model of soluble *T. brucei* CYP51 were compared with those of models of soluble

human CYP51, human CYP2C9 and human CYP2E1. The dynamics of the model of membrane-bound *T. brucei* CYP51 was also compared with those of models of membrane-bound CYP2C9. The binding site residues of both *T. brucei* CYP51 and human CYP51 are more rigid than those of CYP2C9 and CYP2E1. The results suggest that differences between the active site residues of *T. brucei* and human CYP51s may be key for designing *T. brucei* specific inhibitors. The ligand tunnels in both *T. brucei* and human CYP51 were also studied. Tunnel 2f serves as the predominant ligand tunnel in both proteins, but *T. brucei* CYP51 often uses the solvent (S) tunnel, tunnel 1 and the water (W) tunnel as the second predominant tunnel whereas human CYP51 uses tunnel 1. The difference in the use of ligand tunnels may be important for ligand specificity of the two proteins. Analysis of tunnel entrance residues shows differences between these residues in the two proteins and this information can be helpful for designing *T. brucei* CYP51-specific inhibitors which do not inhibit human CYP51.

Interactions of P450 and cytochrome P450 reductase were studied using Brownian dynamics simulations and molecular dynamics simulations. Complexes of soluble forms of different P450s, including CYP51, CYP2B4, CYP1A2, CYP2A6, CYP2C9, CYP2D6 and CYP2E1 were investigated. The P450s bind to the reductase using a similar interface, the positively charged proximal side. The interface residues on the proximal side of P450s are normally arginine and lysine residues and those on CPR are aspartic acid and glutamic acid residues. P450s bind to CPR with different affinities and these affinities can be inferred from the computed binding energy of the complexes. A model of the membrane-bound complex of *T. brucei* CYP51 and human CPR was built based on the membrane-bound model of *T. brucei* CYP51 and the encounter complexes of P450s and CPR generated by the Simulation of Diffusional Association (SDA 7) software, which was used for the Brownian dynamics simulation. A 210 ns molecular dynamics simulation was performed for the system. In the simulation, *T. brucei* CYP51 changed its orientation slightly. The interface residues were maintained in the simulation and the distance between the heme center of the P450 and the FMN cofactor of CPR shortened to a distance suitable for electron transfer.

SDA 7 is a useful software package for performing Brownian dynamics simulations of macromolecules. However, it has many inter-dependent input parameters and the analysis of the results may be difficult for new users. A webserver for SDA 7 (webSDA) was built to improve the user-friendliness of SDA 7 and automate the

procedure for preparing and running SDA jobs. The webserver will not only help new users to to become familiar with the SDA software but also experienced users to set up their simulations easily.

Zusammenfassung

Cytochrom P450 Enzyme (CYP, P450) spielen eine wichtige Rolle im Stoffwechsel von Medikamenten und der Sterolbiosynthese. Humane P450s katalysieren mehr als 80% aller Medikamente. P450s, z.B. CYP51, nehmen in dem Prozess der Steroidbiosynthese teil. CYP51 katalysiert die 14-Demethylierung von Sterinmolekülen. CYP51 existiert in allen Organismen. Es wurde als Drug Target für antimykotische Medikamente verwendet. CYP51 könnte auch als Drug Target für antiparasitische Medikamente verwendet werden. Das Verständnis über den Wirkmechanismus von P450 Enzymen erlaubt nicht nur den Metabolismus von Medikamenten zu verstehen, sondern auch die Entwicklung von Parasiten-spezifischen Wirkstoffen, welche nicht bei menschlichen orthologen Enzymen wirken.

P450 ist ein Hämprotein. Das aktive Zentrum ist im inneren des Protein, deshalb müssen Substrate und Produkte über Tunnel zum aktive Zentrum ein bzw. aus geschleust werden. P450 ist ein Membranprotein mit einer Transmembranhelix. Für die katalytische Reaktion der P450, sollen zwei Elektronen von ihren Redoxpartner übertragen werden. In Eukaryoten ist meist Cytochrom P450-Reduktase (CPR) als Redoxpartner eingesetzt. CPR ist ebenfalls ein Membranprotein. Die Verbindung zwischen P450 und CPR wird durch elektrostatische Wechselwirkung vermittelt. Zwei Elektronen werden vom Flavinmononucleotid (FMN), Cofaktor von CPR, zur Hämgruppe des P450 übertragen.

Aktuelle experimentelle Ergebnisse liefern kein vollständiges Verständnis für die Dynamik der P450 Enzyme, der Unterschiede zwischen den P450, wie das Enzym in der Membran integriert ist, wie Liganden zum aktiven Zentrum hin und wieder weg transportiert werden und wie P450 mit CPR interagiert. Um diese Aspekte weiter zu untersuchen wurde ein Mehrskalenansatz durchgeführt, welcher „all-atom“-Moleküldynamiksimulationen, „coarse-grained“-Moleküldynamiksimulationen sowie Brownsche Dynamiksimulationen verwendet.

Ein Modell des membrangebundenen CYP51 von *T. brucei* wurde erstellt. Dieses ist in Übereinstimmung mit bestehenden experimentellen Daten. Die Dynamik des Modells von löslichem CYP51 von *T. Brucei* wurden mit den Modellen von löslichem humanem CYP51, humanem CYP2C9 sowie humanem CYP2E1 verglichen. Die Dynamik des Modells der membrangebundenen CYP51 von *T. brucei* wurde auch mit Modellen der membrangebundenen CYP2C9 verglichen. Die Aminosäurenreste an der Bindungsstelle von beiden CYP51 von *T. brucei* und humaner CYP51 sind fester als die von CYP2C9 und CYP2E1. Die Ergebnisse zeigen, dass Unterschiede zwischen den Resten des aktiven Zentrums von *T. brucei* und humanem CYP51s der Schlüssel für die Suche nach *T. brucei* spezifischen Inhibitoren sein kann. Die Ligandtunnel in beiden *T. brucei* und menschlichem CYP51 wurden untersucht. Tunnel 2f dient als vorherrschender Ligandtunnel in beiden Proteinen, aber *T. brucei* CYP51 verwendet oft den Solvens(S) Tunnel, Tunnel 1 und den Wasser (W) Tunnel als den zweit vorherrschenden Tunnel, wohin gegen menschliches CYP51 Tunnel 1 verwendet. Der Unterschied bei der Verwendung der Ligandentunnel kann wichtig für die Ligandenspezifität von beiden Proteinen sein. Die Analyse der Aminosäurereste am Tunneleingang zeigt Unterschiede zwischen den Aminosäureresten der beiden Proteine. Diese Information kann für das Design von *T. brucei* CYP51-spezifischen Inhibitoren wichtig sein.

Interaktionen von P450 und Cytochrom P450-Reduktase wurden mit Hilfe Brownscher Dynamiksimulationen und Moleküldynamiksimülationen untersucht. Komplexe der löslichen Formen der verschiedenen P450, inkl. CYP51, CYP2B4, CYP1A2, CYP2A6, CYP2C9, CYP2D6 und CYP2E1 wurden untersucht. Die P450 binden an die Reduktase unter Verwendung einer ähnlichen Grenzfläche, die positiv proximale Seite. Die Aminosäurereste der Grenzfläche auf der proximalen Seite des P450 sind normalerweise Arginin- und Lysinreste. Die von CPR sind Asparaginsäure- und Glutaminsäureresten. P450s bindet an CPR mit unterschiedlichen Affinitäten. Diese Affinitäten können von der berechneten Bindungsenergien der Komplexe abgeleitet werden. Ein Modell des membrangebundenen Komplexes von *T. brucei* CYP51 und humanem CPR wurde aufgrund des membrangebundenen Modells von CYP51 (*T. Brucei*) und dem „encounter"-Komplex von P450 und CPR erzeugt. Diese wurde durch Brownschen Dynamik-Simulation (Diffusional Association Software, SDA 7) erzeugt. Eine 210 ns Moleküldynamiksimulation wurde für das System durchgeführt. In der Simulation verändert CYP51 von *T. brucei* seine Richtung in der Membran leicht. Die Aminosäurereste der Grenzfläche haben sich

in der Simulation nicht verändert und hat die Distanz zwischen dem Häm des P450 und dem FMN Cofaktor des CPR haben sich verkürzt. Der Elektrontransfer ist mit der verkürzten Distanz möglich.

SDA 7 ist ein Software-Paket um Brownsche Dynamiksimulationen von Makromolekülen durchzuführen. Aber es hat viele verschiedene voneinander abhängige Eingabeparameter und die Analyse der Ergebnisse kann für neue Nutzer schwierig sein. Ein Webserver für SDA 7 (webSDA) wurde programmiert um die Benutzerfreundlichkeit von SDA 7, durch Verbesserung und Automatisierung des Verfahrens zur Vorbereitung und Durchführung SDA Jobs, zu erhöhen. Der Webserver ist hilfreich für neue Benutzer, welche die SDA-Software benutzen, aber auch für erfahrene Benutzer um ihre Simulationen zu erstellen.

Publications

1. **Xiaofeng Yu**, Vlad Cojocaru, Ghulam Mustafa, Outi Salo-Ahen, Galina Lepesheva and Rebecca Wade, Dynamics of CYP51: implications for function and inhibitor design. *Journal of Molecular Recognition*, **28**: 59-73, 2015.

   Chapter 3 is based on the above publication. Simulations of both soluble and membrane-bound models of CYP2C9 were provide by Vlad Cojocaru and those of CYP2E1 by Outi Salo-Ahen.

2. **Xiaofeng Yu**, Michael Martinez, Annika Gable, Jonathan Fuller, Neil Bruce, Stefan Richeter and Rebecca Wade, webSDA: a web server to simulate macromolecular diffusional association. *Nucleic Acids Research*, 2015, doi: 10.1093-/nar/gkv335.

   Chapter 7 is based on the above publication. For webSDA, the python script and addatoms module were developed by Michael Martinez, some front-end tools and result analysis by Annika Gable, the tool to use PDB ID and upload PDB files on the front page was implemented by Jonathan Fuller and back-end job launching tools by Stefan Richter.

3. **Xiaofeng Yu**, Vlad Cojocaru and Rebecca Wade, Conformational diversity and ligand tunnels of mammalian cytochrome P450s. *Biotechnology and Applied Biochemistry*, **60**: 134-45, 2013.

   Part of Chapter 1 is based on the above publication.

4. Ghulam Mustafa, **Xiaofeng Yu** and Rebecca Wade. Structure and dynamics of human drug-metabolizing cytochrome P450 enzymes. in Kirchmair Jonhannes, Editor, *Drug Metabolism and Prediction*, Chapter 4, 77-102, Wiley-VCH, Weinheim, 2014

Contents

1

Introduction

1.1 Protein Structure

Proteins are fundamental building blocks of life. In humans, muscle cells mainly use two proteins, actin and myosin, to produce contractions. In brain, another type of proteins, signaling receptors on the membrane of nerve cells can be used to transduce signals to response to stimuli. There are many other biological processes modulated by proteins, such as DNA transcription, RNA translation and the cell cycle.

Basic building blocks of proteins are amino acids. An amino acid has three functional groups: the amine group (-NH_2), the carboxylic group (-COOH) and the side chain (-R). The amine group and the carboxylic group are connected by a carbon atom (Cα) between them. The side chain of an amino acid is the organic substituent group (-R). Figure 1.1A shows the generic structure of amino acids. There are 20 common amino acids. At pH 7, they can be divided into three groups: charged amino acids, polar but not charged amino acids and hydrophobic amino acids. Charged amino acids are sub-divided into positively and negatively charged amino acids. Positively charged amino acids are arginine (ARG, R) and lysine (LYS, K). Histidine (HIS, H) can also be positively charged depending on its protonation state. Negatively charged amino acids are aspartic acid (ASP, D) and glutamic acid (GLU, E). Polar amino acids are serine (SER, S), threonine (THR, T), asparagine (ASN, N) and glutamine (GLN, Q). Less polar amino acids are glycine (GLY, G), alanine (ALA, A), proline (PRO, P) and tyrosine (TYR, Y). Hydrophobic amino acids are valine (VAL, V), isoleucine (ILe, I), leucine (LEU, L), methionine (MET, M), phenylalanine (PHE, F), tryptophan (TRP, W) and cysteine (CYS, C).

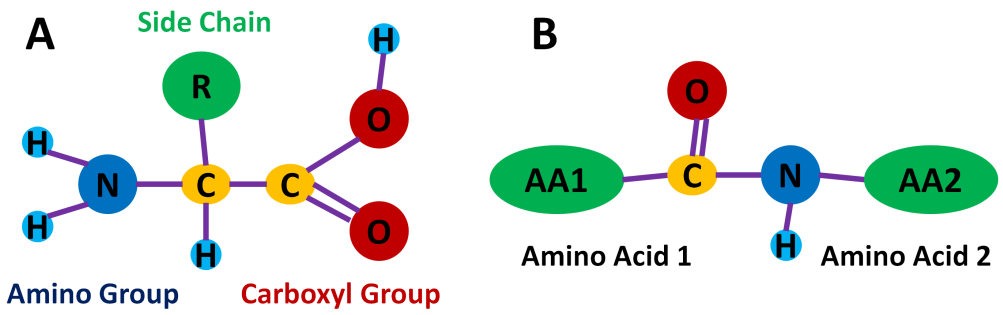

Figure 1.1: General structure of amino acids (A) and dipeptides (B).

Amino acids are connected by peptide bonds as shown in Figure 1.1B. The peptide bond is a covalent bond formed by condensation of two amino acids. There are two conformations for peptide bonds. The one shown schematically in Figure 1.1B is the trans peptide bond, which is the predominant conformation of peptide bonds. However, the cis conformation for peptide bonds may form in proteins as well. The cis peptide bonds have both -NH and -C=O groups on the same side of the peptide bond.

The primary structure of a protein is the amino acid sequence. The primary structure is held together by peptide bonds. The amino acid sequence is translated from RNA by ribosomes and RNA is transcribed from DNA by RNA polymerases. Mutations in the primary structure of proteins may result in disorders of protein function and thus cause human diseases.

The secondary structure is the general three-dimensional form of a protein. There are three major secondary structure elements, the alpha helix, beta sheet and loop. The secondary structure is characterized by hydrogen bond patterns.

The tertiary structure is the arrangement of secondary structures of proteins, or the fold of protein. The folding is driven by burying hydrophobic residues inside the protein and exposing the hydrophilic residues on the surface of the protein. The tertiary structure of a protein is crucial for its function. Misfolded proteins may lead to disrupted function of proteins and thus result in disease.

The quaternary structure is how subunits of proteins arrange in complexes. The subunits can be of the same amino acid sequence (homo-oligomer) or of different amino acid sequences (hetero-oligomer).

1.2 Cytochrome P450 Enzymes

Cytochrome P450 enzymes (CYP, P450) constitute a protein superfamily which is found in all the kingdoms of life existing in archaea, bacteria and eukaryota [1, 2]. The name is derived from the fact that, when the protein is complexed with carbon monoxide (CO) in the reduced state, the protein has its absorption maximum at 450 nm [3]. In humans, there are 57 genes encoding P450s [4, 5]. These genes are divided into 18 families and 44 subfamilies. Important families are CYP family 1 (CYP1), CYP family 2 (CYP2), CYP family 3 (CYP3) and CYP family 51 (CYP51) [4]. The CYP1, CYP2 and CYP3 families are involved in drug metabolism and steroid metabolism, and the CYP51 family is important for steroid biosynthesis [4, 6]. CYP51 is also used as a drug target for anti-fungal drugs and anti-parasitic drugs [7]. Important human P450s include CYP1A2, CYP2A6, CYP2C9, CYP2D6, CYP2E1 and CYP3A4 which metabolize 70-80% of all drugs [1, 5]. The following sections “Structure”, “Ligand Tunnels” and “Membrane-bound Cytochrome P450s” are based on the review by Yu et al. [2].

1.3 Structure

As of November 2014, there are 645 structures of P450s in the Protein Data Bank (PDB, `http://www.pdb.org/pdb/home/home.do`). 135 out of 645 structures are human P450s. The sequence similarity between different P450 families is lower than 40% [8]. However, the protein fold and secondary structure are conserved. There are 12 α-helices, named A-L, and four β-sheets, named 1-4 [9, 10]. The short helices between helices A-L are named with prime or double prime, such as B’, F’ and G’ helices [8].There are two important blocks in P450s, the region between B and C helices, the B-C loop region, and the region between F and G helices, the F-G loop region.

These crystal structures have been solved for 23 different families of P450s and are listed in Table 1.1. Most of the structures have ligands in the active site and are in a closed form. A few CYP structures are in an open form and some of them are partially open with ligands near the protein surface rather than in the active site. Table 1.1 gives the length of the secondary structure elements in the B-C and F-G blocks which are the regions of the protein that generally undergo the greatest conformational changes between open and closed structures. We define the B-C

block to extend from the beginning of the B-helix to the end of the C-helix, and the F-G block to extend from the beginning of the F-helix to the end of the G-helix. Although the overall fold of the CYPs is well conserved, the length and secondary structure of the B-C and F-G blocks vary considerably. While the flexibility and degree of conformational change of the B-C and F-G blocks of CYPs cannot be deduced just from the secondary structure content alone, the secondary structure does provide an indicator of the flexibility of the B-C and F-G blocks. In general, going from loop to helix to sheet, the flexibility decreases whereas the rigidity increases. The longer the loop region the block has, the more flexible this block should be. For some crystal structures, parts of the B-C and F-G blocks are missing because they are disordered in the crystal. In addition, the length of the B-C and F-G blocks and the amino acid composition of these regions can also influence their flexibility. But how exactly the flexibility relates to the conformational changes or the transformation between open and closed states, is still unknown [11].

CYP	PDBID	Ligand state	Conf.	B-C block			F-G block		
				αB	B-C loop	αC	αF	F-G loop	αG
CYP1A1	4I8V			*_9_2_4___7_____5_14_17*			*8_1_7_4__3___2_5__3_25*		
CYP1A2	2HI4	+L	C	*_9_2_4___7_____5_14_17*			*8_1_7_4__3___2_5__3_25*		
CYP2A6	1Z10, 1Z11, 2FDU, 2FDV, 2FDW, 2FDY, 2PG5, 3EBS	+L	C	*_9_2_4_**3**_3_____7__9_18*			*___17_2__7____5__4_25*		
	2PG6, 2PG7	-L	C	*_9_2_4_**3**_3_____7__9_18*			*___17_2__7____5__4_25*		
	3T3Q, 4EJJ	+L	C	*_9_8___**2**_3_____7__9_17*			*___16_3_13_____5__25*		
	3T3R	+L	C	*_9_8___**2**_3_____7__9_17*			*___17_2_13_____5__26*		
CYP2A13	2P85	+L	C	*_9_2_4_**3**_3_____7__9_18*			*___17_2__7____5__4_25*		
	3T3S	+L, -L	C	*_9_____13_____7__9_18*			*___16_3____13____5_25*		
	4EJG, 4EJH	+L, -L	C	*_9_5_3_**2**_3_____7__9_17*			*___16_3____13____5_24*		
	4EJI	+L	C	*_8__8__**2**_3_____7__9_17*			*___16_3____13____5_25*		

CYP3A4	1W0G, 2V0M, 3NXU	+L	C	*9 2 5 9 4 8 15*	*6 9 8 2 9 5 18*
	1TQN, 1W0E	-L	C	*9 6 5* **3** *7 3 8 15*	*6 9 8 2 9 5 18*
	1W0F	-L, $+L_S$	C	*9 6 5* **3** *7 3 8 15*	*6 9 8 2 9 5 18*
	2J0D	+L, -L	O^2_{MR}	*9 2 5 9 4 8 15*	*9* **2** *7 2 9 5 18*
	3TJS	+L	O^2_{MR}	*8* 17 *5 6 15*	*8 2 7 6 2 3 2 4 7 19*
	3UA1	+L	C	*8* 17 *5 6 15*	*6* 10 *8 1 9 7 15*
CYP7A1	3DAX	-L	C	*8 4* **4** *13* 14 *10*	*21 6 3 3 1 13*
	3SN5	+L	C	*8 5* **2** *2 11 6 3 6 9*	*16 2 6 3 3 1 13*
CYP8A1	3B6H	+L, $+L_S$	I^1	*8 6* **2** *2 9* 10 *9*	*20 4 17 2 4 7 12*
CYP11A1*	3N9Y, 3N9Z, 3NA1, 3NA0	+L	C	*9* 11 *9 6 3 2 15*	*20 4 7 1 32*
	3MZS	+L	C	*9* 11 *10* 11 *14*	*19 4 7 1 30*
CYP17A1	3RUK	+L	C	*8 3 4 6 6* 13 *15*	*16* 10 *4 4 21*
	3SWZ	+L	C	*8* 13 *6* 13 *13*	*16* 10 *3 5 23*
CYP19A1	3EQM	+L	C	*9 1 3 6 7* 11 *15*	*19 3 5 1 3 1 26*
	3S79	+L	C	*13 6 7* 12 *14*	*18 4 5 1 3 1 26*
	3S7S	+L	C	*9 1 3 6 7 12 14*	*18 4 5 1 3 1 26*
CYP21A1	3QZ1	+L, $+L_S$	I^1	*4* 34 *15*	*12 8 8 1 3 5 26*
CYP24A1*	3K9V, 3K9Y	+L, $+L_S$	O^2	*10* 10 *10* 10 *16*	*13 1 3 5 7 2 28*
CYP46A1	2Q9F	+L	C	*9 4* **2** *1 7* **3** *1* **3** *11 13*	*17* 11 *32*
	3MDM	+L	C	*9 4* **2** *1 7* **3** *1* **3** *11 13*	*18 3 3 2 34*
	3MDR, 3MDT	+L	C_{MR}	*9 4* **2** *1 7* **3** *1* **3** *11 13*	*18 2* ***7*** *32*
	2Q9G	-L	C_{MR}	*9 4* **2** *1 7* **3** *1* **3** *11 13*	*17 4* ***10*** *27*
	3MDV	+L	C_{MR}	*8 8 8* **2** *2* **2** 11 *15*	*16 2* **7** *1 30*
CYP1B1	3PM0	+L	C	*8* 13 *9 9 16*	*6 3 8 9 5 3 30*
CYP2B4	1SUO, 2Q6N	+L	C	*9 2 4* **3** 3 *7 9 19*	*18 1 9 5 4 24*
	3KW4	+L, $+L_S$	C	*9 2 4* **3** 3 *7 9 19*	*18 1 9 5 4 24*
	3ME6	-L	C	*9 2 4* **3** 3 *7 9 19*	*18 1 9 5 4 24*
	3MVR	$+L_S$	C	*9 2 4* **3** 3 *7 9 19*	*18 1 9 5 4 24*

	2BDM	+L, $+L_S$	O^{12}	*_9_2_4___*6_**3**_5_**3**__5_*18*	*___11_8___9_____*9_*25*
	3G5N, 3G93	+L, $+L_S$	O^{12}	*11*_1_*3___*4_*4*_2_*4*__3_*21*	*___11*_**8**_*10*_10__*5*__1_*18*
	1PO5	-L	O^1	*_9_2_3_____*21_*16*	*___13*_6___*9___6*__3_*25*
	3R1A	+L	C	*_8_8_***2**__6___*4_____*10_*17*	*___18*_1_____*13___*5_*24*
	3R1B	+L, $+L_S$	O^{12}	*_9_8_***2**___7_____*9*__3_*17*	*___11*__1__**2**__5_*10*_*9_23*
	3TK3	+L	C	*11*_6_**2**___6_____*4*__*9_17*	*___16*__3_____*16*_6_*25*
CYP2B6	3IBD	+L	C	*_8*_5_*3*_**2**__6_____*4*__*9_17*	*___17*__2_____*13*_5_*25*
	3QOA	+L	C	*_8*_3_*4*_**3**__6_*4*_3_*4*__2_*14*	*___18*__1_____*13*_5_*26*
	3QU8	+L, -L	C	*_8*_8_**2**__6_____*4*__*9_17*	*___18*__1_____*13*_5_*25*
CYP2C5	1DT6	-L	I^1_{MR}	*_9_2_4_____*7___*4*_14_*13*	*___16*_4_____***11***___4_*25*
	1N6B, 1NR6	+L	C	*_9_2_4_***3**_3_____*7*__*9_15*	*___17*_4___*5*__2__*6*_1_*27*
CYP2C8	1PQ2	-L, $+L_S$	C	*_9_2_3___*7_____*7*__*9_15*	*___17*_3___*6*__4__*4*_1_*27*
	2NNH, 2NNI, 2NNJ, 2VN0	+L, $+L_S$	C	*_9_2_4_***3**_3_____*7*__*9_15*	*___18*_1___*7*__1_*6*__1_*27*
CYP2C9	1OG2	-L	C	*_9_2_4_***3**_3_____*7*_10_*15*	*___18*_1___*8*__1_*6*__1_*28*
	1OG5	+L	C	*_9_2_4_***3**_3_____*7*_10_*15*	*___18*_1___*8*__1__*6*_1_*28*
	1R9O	+L	I^1_{MR}	*_9_2_4_***3**_____18_*15*	*___18*_4_____***7***___8_*26*
CYP2D6	2F9Q	-L	C	*_9_2_3___*7_____*6*_15_*19*	*___25___*13___*4___20*
	3QM4	+L	I^1	*_9_3_3___*7_____*6*_15_*16*	*___15*_4__*6*___*9__6*_2_*19*
CYP2E1	3E4E, 3E6I, 3GPH, 3KOH, 3LC4	+L	C	*10*_2_*4*_**3**_3_____*6*_11_*17*	*___17*__2__*9*_____*4*_6_*25*
	3T3Z	+L	C	*_9_7_***2**_3_____*3*_15_*16*	*___16*__3__*8*_1___*4*_6_*24*
CYP2R1	3C6G, 3CZH, 3DL9	+L	C	*_9_2_4___*7_____*7*_11_*17*	*___17*__2__*8*__1_*5*__4_*23*
CYP51	3JUS, 3LD6	+L	C	*_8*_6__**2**__2__*10*__4_*3*_2_*12*	*___30*__*7*__*6*__10_____*31*
	3JUV	-L	C	*_8*_6__**2**__2__*10*__4_*3*_2_*12*	*___30*__*7*__*6*__10_____*31*

Table 1.1: Summary of conformations observed in the crystal structures of mammalian CYPs. The structures with ligand(s) bound in the active site are denoted as "+L", the apo structures by "-L". When bound and apo forms of the enzyme are present in the same structure (different chains) with identical conformations, both labels "+L" and "-L" were used for the same structure. "$+L_S$" represents ligand(s) bound at the protein surface. The conformation type is labeled as: (i) "O"-open; (ii) "I"-partially open; (iii) "C"-closed. The open conformations were classified into 2 sub-types: (i) "O^1,", "I^1,"-structures with the open cleft between the B-C and F-G blocks (with the 2a, 2ac and 2c channels open); (ii) "O^2"-structures with open cleft between the F-G block and the β4 region (with the solvent and the 2f channels open). Incomplete structures are labeled with "MR" and the number of missing residues is shown as "x" in the conformations of the B-C and F-G blocks (columns 4, 5). The conformational signatures of the two blocks are defined as a sequence of secondary structure elements (α-helix in italic text, β-strand in bold text, coil in normal text, missing crystal residues in italic and bold text), each of which is formed by the indicated number of residues. The helices αB, αC, αF, and αG are flanking helices of the 2 variable structure blocks. (*) Crystal structures have been determined for 2 mitochondrial CYPs. Mitochondrial CYPs do not have an N-terminal transmembrane helix.

Examples of closed, open and partially open crystal structures are shown in Figure 1.2. CYP2B4 has been found to adopt a range of closed conformations (Figure 1A) as well as different open conformations (Figure 1.2, B and C). One of the open structures of CYP2B4 (PDB id: 3R1B) displays a very large conformational difference in the F-G block compared to the closed structure (Figure 1.2A vs. B). The F and G helices are twisted to adapt to the movement of the combined F'/G' helices which make contacts to another protein subunit in the crystal. This movement of the F-G block opens up a tunnel to the active site making it accessible to ligands. We previously developed a nomenclature for the different ligand pathways through tunnels in CYPs which are distinguished by the secondary structure elements lining them and their spatial location [12, 13]. In this open structure (Figure 1.2B), the channel encompasses and merges pathways 2a, 2ac and 2b, which all run between the F-G block and the B-C loop. Pathway 2a leaves the protein between the F-G loop, B' helix/B-B' loop/B-C loop and the β1 sheet, pathway 2c between the G and I helices and the B' helix/B-C loop, and pathway 2ac between pathways 2a and 2c. Pathway 2b leaves between the B-B' loop and the β1 and β3 sheets, whereas pathway 2d leaves between the N-terminus and helices a/A' and A, pathway 2e through the B-C loop, and pathway 2f between the F' helix/F-G loop and the β5 sheet [12]. In the open structures of other mammalian CYPs (CYP3A4 and CYP24A1) that are currently available, the ligand tunnel is located between the F-G block and the β4 region and corresponds to the solvent

channel (S) and pathway 2f. The solvent channel, so named because it has been proposed as a route for water molecules [14], passes between the E, F and I helices and the β5 sheet [12]. The two open structures of CYP3A4 show some disordered missing residues near the end of the F helix.

In the partially open structure of human CYP3A4 (Figure 1.2D), the F-G loop moves towards the B-C loop in the direction parallel to the heme plane leaving the tunnels S and 2f open. The structure of the bovine P450C1 (CYP21A1) has also been solved in a partially open conformation with a ligand ((9beta)-17-hydroxypregn-4-ene-3,20-dione) trapped in the 2b ligand tunnel (Figure 1.2, E and F). Here, it is the B-C loop rather than the F-G block that shows conformational diversity, its B-C loop lacks a B' helix. In some of the other partially open crystal structures listed in Table 1.1 though, the F-G block shows conformational change. In two of these cases (for CYP2C5 and CYP2C9), part of the F-G loop is disordered and missing from the structure.

1.4 Ligand Tunnels

1.4.1 Molecular Dynamics Techniques

MD simulations have been applied to CYPs to study many aspects of their function including ligand tunnel analysis, prediction of metabolic products and design of inhibitors or drugs. Here, we focus on their application to understand how the ligands egress from the buried active site of CYPs. Table 1.2 shows the simulations that have been done to study the ligand tunnels in eukaryotic CYPs. These employ classical MD simulation, in some cases, together with enhanced sampling procedures such as RAMD (random acceleration molecular dynamics), umbrella sampling and SMD (steered molecular dynamics).

In classical MD simulations, the typical simulation lengths are currently ca. 10-100 nanoseconds. This time is much shorter than the timescale for enzymatic turnover of CYPs and ligand dissociation [15]. Thus, while standard MD simulations can reveal many motions relevant to properties such as ligand tunnel opening, they cannot currently give a full picture of how the CYPs function.

To study phenomena that take place on longer timescales than the duration of a typical MD simulation, an enhanced sampling technique needs to be used. These methods increase the sampling of the event of interest during the MD simulations.

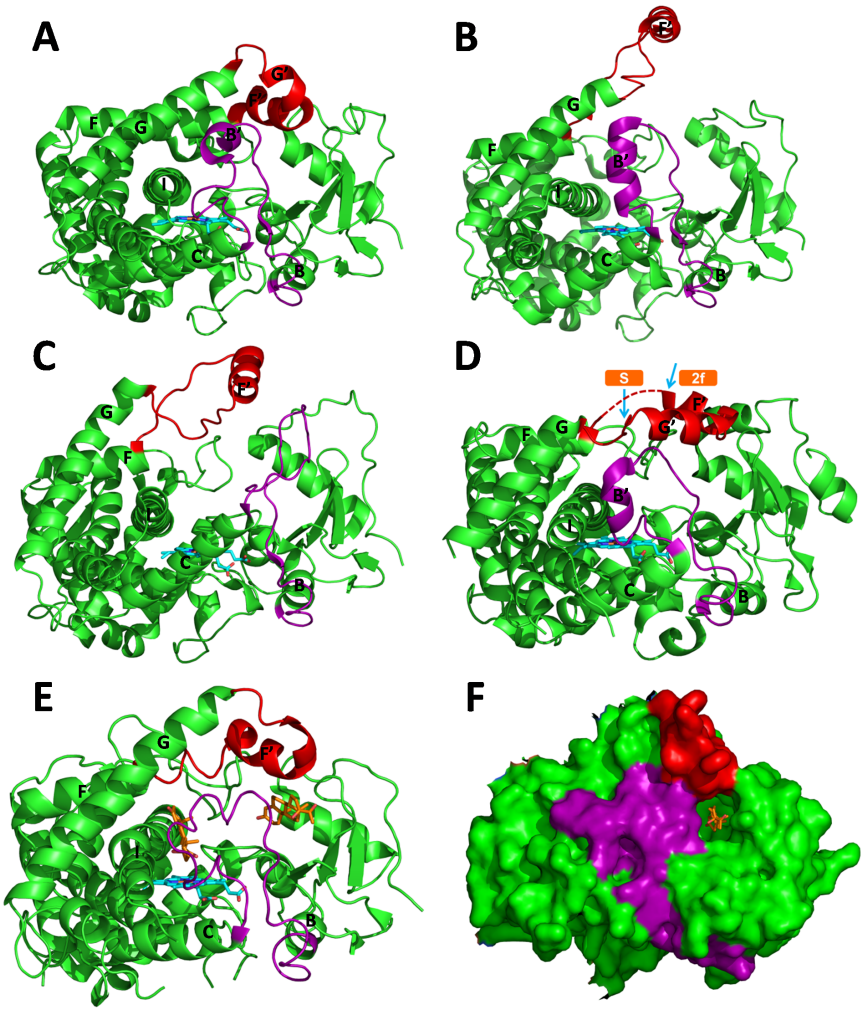

Figure 1.2: Comparison of closed, open, and partially open crystal structures of mammalian CYPs with key helices labeled. (A) Closed structure of human CYP2B4 (PDB id: 3R1A). (B) Open (PDB id: 3R1B) structure of rabbit CYP2B4. In this open structure, the F-G loop extends out to make crystal contacts with another protein subunit in the crystal. (C) Open structure with another conformation of rabbit CYP2B4 (PDB id: 3G5N) showing the possibility to adopt different open conformations. (D) Structure of human CYP3A4 (PDB id: 2J0D) with the 2f and S tunnels open. The partially open structure of bovine CYP21A1 (P450c21, PDB id: 3QZ1) is shown in cartoon (E) and surface (F) representations. Both the ligand in the active site and the ligand trapped in the 2b tunnel, holding it open, are shown in stick representation. The proteins are shown in cartoon representation with the heme cofactors in stick representation. The F-G loop region is shown in red and the B-C loop/B' helix in purple.

Random acceleration molecular dynamics (RAMD) was designed to study ligand egress from the buried active sites of enzymes such as CYPs and was first applied to study the bacterial cytochromes P450cam and P450-BM3 [16, 17]. In RAMD, a user-defined force with a randomly chosen direction is applied to the center of mass of the ligand in the active site of the protein. The motion of the sys-

tem is then propagated by classical MD simulation. After a pre-defined number of timesteps, the distance that the ligand has traveled is checked and, if it is shorter than a pre-defined minimum distance, the direction of the force will be changed. If it is longer, the same force with the same direction will continue to be applied to the ligand. The simulation is continued with the random force direction continually adjusted using the same criterion until the ligand has exited from the protein. RAMD thus allows ligand egress to be observed in short simulations (in published studies ranging from several ps to a couple of ns). Many RAMD trajectories can be run for one ligand-receptor complex to explore the range of possible ligand egress routes from a protein. RAMD provides a way to probe weak spots in the active site cavity by which a ligand can leave. It provides mechanistic insights into how ligand egress occurs but it should be born in mind that the timescale of RAMD simulations does not correspond to the real ligand egress timescale. Furthermore, RAMD will sample fast and/or ligand-induced conformational changes upon ligand egress better than low-frequency motions of the protein such as inter-domain breathing.

The ligand egress routes identified by RAMD can be studied in more detail to compute force and free energy profiles by several methods. SMD [33] has been used in a number of studies of ligand egress from CYPs. As in RAMD, a force is again applied to the center of mass of the ligand. In SMD, though, the force is usually applied in the same direction throughout the simulation and thus the route to be taken by the ligand must be known before the simulation is performed. Analogously to an atomic force microscopy (AFM) experiment, the ligand is pulled slowly out of the protein and the corresponding force profile is computed. Another procedure to compute the free energy profile or potential of mean force (PMF) for ligand egress from a protein is Umbrella Sampling. This is a method that allows free energy barriers between two states to be computed by adding an additional 'umbrella term' to the force field to enhance the sampling of energetically unfavorable configurations in the vicinity of the free energy barriers. SMD and umbrella sampling, as well as other methods for computing free energy profiles, can require long calculations in order to obtain converged and accurate free energy profiles. The difficulty in obtaining accurate free energy profiles for CYPs comes from the difficulty of fully sampling the lower frequency, large scale degrees of conformational freedom. Thus, free energy barriers are generally overestimated but these

CYP	PDB id[i]	Resolution (Å)	Organism	Solubility[ii]	Ligand[iii]	Preferred Exit Tunnels[iv]	Method[v]	Ref.	Year
CYP1A2	2HI4	1.95	human	SOL	– – – – –[vi]	– – – – –[viii]	MD	[18]	2011
CYP2A6	1Z10	1.90	human	SOL	coumarin	2c	MD, RAMD, SMD	[19]	2011
	1Z10	1.90	human	SOL	– – – – –[vi]	(2b), (2e)	MD	[20]	2012
CYP3A4	1W0G	2.73	human	SOL	metyrapone	2e	MD, SMD	[21]	2007
	1TQN	2.05	human	SOL	temazepam, testosterone-6βOH	2e, 2e, 3	MD, SMD	[22]	2009
	2V0M	2.80	human	SOL	ketoconazole	2b, 2e	MD	[23]	2010
	1TQN	2.05	human	MEM	– – – – –[vi]	2b	MD	[24]	2012
	1TQN	2.05	human	SOL	– – – – –[vi]	2e, S, (2a), (2b)	MD	[20]	2012
CYP4A22	Model	—	human	SOL	– – – – –[vi]	– – – – –[viii]	MD	[25]	2010
CYP2B1	Model	—	rat	SOL	testosterone	2c, 2e	MD, SMD	[26]	2005
CYP2B4	1PO5	1.60	rabbit	SOL	– – – – –[vi]	2a	MD	[20]	2012
CYP2C5	1N6B	2.30	rabbit	SOL	progesterone, 21-hydroxy-progesterone, DMZ[vii]	2c	MD, RAMD	[27]	2005
CYP2C9	1R9O	2.00	human	SOL	Warfarin[vii], flurbiprofen[vii]	2a, 2ac, 2c, 2e	MD, RAMD	[28]	2012
	1R9O	2.00	human	MEM SOL	– – – – –[vi]	2a, 2c, S	MD	[29]	2011
	1OG2	2.60	human	MEM	– – – – –[vi]	S	MD	[30]	2011
	1OG2	2.60	human	SOL	– – – – –[vi]	2b, S, (2a), (2e)	MD	[20]	2012
CYP2D6	2F9Q	3.00	human	SOL	– – – – –[vi]	S	MD	[20]	2012
CYP2E1	3E6I	2.20	human	SOL	indazole	2a, 2c	MD, RAMD, SMD	[31]	2012
	3E6I	2.20	human	SOL	– – – – –[vi]	(2ac), (2e)	MD	[20]	2012

Table 1.2: Summary of the ligand tunnel simulations of mammalian cytochrome P450s. i. PDB identifier of the crystal structure used for the simulations. If the structure used is a modeled structure, this is denoted by "Model"; ii. Solubility, "MEM" indicates simulation of the CYP embedded in a lipid bilayer; "SOL" indicates simulation of a water soluble form of the CYP without a lipid bilayer and without the N-terminal transmembrane region of the protein; iii. Ligands present in the simulations; iv. Tunnels are denoted by the nomenclature described in ref. [12] (see Figure 1.3), the tunnels that are partially open are shown in brackets; v. MD: molecular dynamics (these studies give information on the dynamics of ligand tunnels, but do not involve simulation of the process of ligand egress), RAMD: random acceleration molecular dynamics, SMD: steered molecular dynamics; vi. Plain MD simulation together with analysis of the ligand tunnels without a ligand being present or considered explicitly using CAVER [32] or a similar approach; vii. DMZ:4-methyl-N-methyl-N-(2-phenyl-2H-pyrazol-3-yl)benzenesulfonamide; "warfarin" includes S-warfarin and 7-OH-S-warfarin in charged and neutral forms; "flurbiprofen" includes flurbiprofen and 4-OH-flurbiprofen; viii. These studies were mainly focused on the tunnel dynamics and their gating residues, not on which tunnel is preferred as a ligand exit tunnel.

approaches can be useful to estimate the relative free energy barriers, and thus relative probabilities, of ligand egress by different routes.

1.4.2 Application of Molecular Dynamics Techniques to Study Ligand Egress Tunnels

The ligand egress tunnels were firstly studied using prokaryotic P450s, such as P450cam [34], P450BM-3 [16, 17, 34] and P450eryF [34]. Ligand tunnel 2a is often used by these P450s for ligand egress. The first study of ligand tunnels in a mammalian CYP was done using RAMD by Schleinkofer et al. [27]. Tunnel 2c was found to be the predominant egress tunnel in the simulations of CYP2C5 in water. Two mechanisms of ligand entrance and exit were proposed from these studies based on the first crystal structure of a mammalian CYP, that of rabbit CYP2C5: a one-pathway mechanism with the same entrance and exit route for substrates and products (pathway 2a towards the membrane), and a two-pathway mechanism with a different entrance route for substrates (pathway 2a) from the exit route for the more polar products (pathway 2c to the cytoplasm).

SMD simulations starting with a human CYP2B1 structure derived by homology modeling with testosterone docked in the active site showed the2c and 2e tunnels to be the preferred tunnels [26]. Pathway 2e threads through the B-C loop. A similar procedure was applied to the crystal structure of CYP3A4 with metyrapone. Tunnel 2ewas found to have a higher probability to serve as the ligand exit tunnel [21]. The preferred ligand exit tunnels for ketoconazole from the active site of human CYP3A4 were pathway 2b and 2e using MD simulation and geometric analysis with MOLAxis [23].

The protocol of using a conventional MD simulation followed by a combination of RAMD and SMD has been employed in several studies [16, 17, 19]. First, tunnels are identified and the frequency of ligand egress through the different tunnels is computed with RAMD. Then, the most frequent egress routes are chosen for computation of the force profile for ligand exit using SMD. Lastly, statistical analysis of the PMF with methods such as Jarzynski's equality [35] gives a quantitative energetic comparison of the ligand tunnels. This protocol has been applied to several mammalian CYPs, including human CYP2A6 [19] and human CYP2E1 [31]. Tunnels 2c was proposed to be the preferred egress tunnel for CYP2A6, whereas tunnels 2a and 2c were preferred for CYP2E1. A recent RAMD study of human

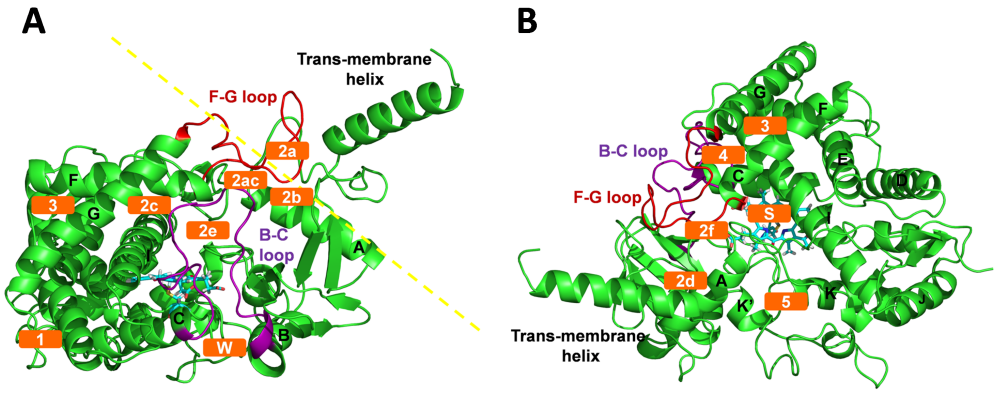

Figure 1.3: Ligand tunnels and secondary structure elements of a model of membrane-bound CYP2C9. Not every CYP has all these ligand tunnels. (A) The dashed line shows the approximate position of the membrane in the simulation of the membrane-bound CYP2C9 [30]. The F-G loop is shown in red and the B-C loop in purple. The ligand tunnels are labeled following the nomenclature in Ref. [12]. The identifier numbers of the simulations (see Table 1.2) in which the tunnel is found to be a preferred ligand egress tunnel it shown next to the name of the tunnel. (B) A view from the other side of the protein showing the other ligand tunnels. The same coloring scheme is used as in (A).

CYP2C9 suggests multiple ligand exit tunnels depending on the ligand properties (Figure 1.3). The main routes identified were tunnels 2a, 2ac, 2c and 2e. The substrate flurbiprofen tended to exit via tunnel 2c or 2e whereas the flurbiprofen product tended to exit via tunnel 2c. The charged form of the S-warfarin product tended to exit by tunnels 2ac and 2e whereas the neutral form of the S-warfarin product was found to exit preferentially through tunnel 2a from the soluble form of CYP2C9. This latter route was shown to point into the lipid bilayer in the models of membrane-bound CYP2C9 [29, 28].

Hendrychova et al. have systematically simulated the six major drug-metabolizing CYPs (2A6, 2B4, 2C9, 2D6, 2E1, 3A4). Each CYP was simulated for 100 ns or longer in aqueous solution [20]. A transformation from an open towards a closed conformation of CYP2B4 was observed in these simulations. The systematic analysis of six major mammalian CYPs demonstrates that the variability in flexibility of the active site and the F-G loop contributes the largest movements in the 6CYPs.During the simulation, the ligand tunnels remained open in CYP2C9 and CYP2B4, temporarily closed in CYP3A4 and CYP2D6, and were always closed in CYP2A6 and CYP2E1.

1.4.3 Aromatic Gating of Ligand Tunnels

The ligand tunnels are often gated by aromatic residues. Despite the conservation of the CYP fold, there is a high diversity in the number and arrangements of the aromatic gates in CYPs. From analysis of crystal structures, it has been suggested that aromatic residues can form a network of gates that regulate the opening and closing of different ligand tunnels cooperatively [36].

The gating of a ligand tunnel was first observed in a CYP in simulations of bacterial P450cam in which ligand egress by tunnel 2a was gated by rotation of phenylalanine and/or tyrosine residues [16, 17].In human CYP2C9 (PDB id: 1R9O, 1OG5), there are three phenylalanine residues (F100, F114, F476) which act as aromatic gates of tunnel 2a (Figure 1.4, A and B). In Figure 1.4A, the three aromatic residues are in a closed state where they cap the ligand in the bound state, whereas in Figure 1.4B, the aromatic residues are in an open state which allows for ligands to enter or exit. In a simulation of the membrane-bound CYP2C9, opening motions of the aromatic gates were observed and the lipid bilayer also stabilized the open aromatic gates [29]. In a study of the complex of CYP3A4 with the ligand temazepam (PDB id: 1TQN), different gating residues were observed for different ligand tunnels (Figure 1.4C) [22]. The pairs of aromatic residues gating ligand tunnels 2a, 2b, 2c and 3 are, respectively, F57-F215, F108-F220, F108-F241 and F213-F241.

Not all ligand tunnels in mammalian CYPs have gating aromatic residues. For example, CYP2C5 (PDB id: 1N6B) is shown in Figure 1.4D. K241 and V106, which are not aromatic residues, gate ligand tunnel 2c [27]. The residues at the same position as the aromatic residues gating tunnel 2a in CYP2C9 are also shown in the figure. There are two aromatic residues (F114 and F473) and one non-aromatic residue (V100), suggesting that tunnel 2a to the membrane might be gated by aromatic residues. Further potential aromatic gates have been identified in simulations of CYP2B1, CYP2E1 and CYP2A6. In the SMD simulation of CYP2B1, the gating residues of tunnel 2e were F108 and F297 whereas F115 was the gating residue of tunnel 2b [26]. The ligand tunnel 2c of CYP2E1 has a ceiling of 5 phenylalanines over the active site [31]. F298 was suggested to be the gate keeper due to its association with the opening and closing of the ligand tunnel. Mutagenesis of another of these phenylalanines, F478 to valine, has shown the importance of this residue and its key position implicates it as the gate keeper for ligand tunnel 2a. In a similar study, rotation of the side chain of F111 was observed to occur

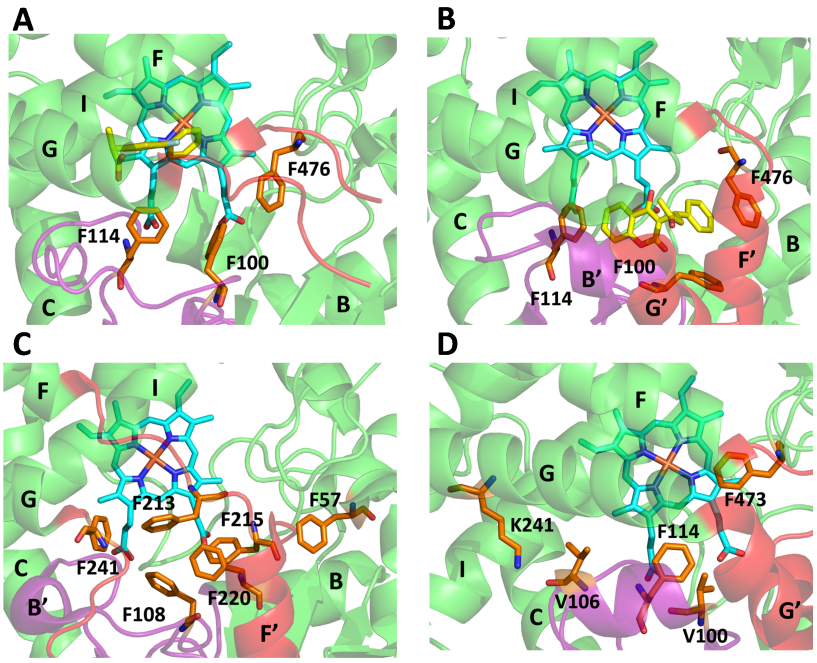

Figure 1.4: Aromatic residues gating ligand tunnels in different CYPs. (A) Gating residues F100, F114, and F476 (tunnel 2a) in a closed conformation of human CYP2C9 (PDB id: R9O). (B) Gating residues F100, F114, and F476 in an open conformation in human CYP2C9 (PDB id: 1OG5). (C) Gating residues F57-F215 (tunnel 2a), F108-F220 (tunnel 2b), F108-F241 (tunnel 2c), and F213-F241 (tunnel 3) of human CYP3A4 (PDB id: 1TQN). (D) Gating residues of rabbit CYP2C5, K241-V106 (tunnel 2c) and residues at the same position as the CYP2C9 aromatic gates (V100, F114, and F473) indicate the diversity of the gating residues which are not limited to aromatic gates.

during ligand egress through tunnel 2c of CYP2A6 in RAMD and SMD simulations [19].

1.5 Membrane-bound Cytochrome P450s

The simulations discussed so far were performed for each CYP in an aqueous environment. The eukaryotic CYPs are membrane bound but the inclusion of a lipid bilayer in the simulation system is challenging because there is no experimentally determined structure of a CYP in a lipid bilayer. It is important to model CYPs in their membrane-bound form because the association with the membrane can influence the protein dynamics, the ligand binding kinetics and the binding selectivity of CYPs.

While a number of models of how CYPs position and orient with respect to a lipid bilayer have been constructed [30], these have been made by hand using molecular graphics and fitting to available experimental data and they give a

range of possible orientations. Recently, Cojocaru et al. built a membrane-bound model of a mammalian CYP in a1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer from first principles using a combination of coarse-grain and atomic-detail MD simulation techniques [29] (Figure 1.3). The model was made for human CYP2C9. The height of the protein above the membrane is consistent with the experimental measurements, as are most of the buried and exposed residues. This membrane-bound model gives indications of how the membrane can influence the dynamics of CYPs and their ligand tunnels. In the membrane-bound form of human CYP2C9, the F-G loop is inserted into the membrane so that the 2a and 2f ligand tunnels lead into the membrane. The other tunnels, including 2ac, 2b, 2c, 2d, 2e and S lead to the aqueous solution. This orientation of CYP2C9 above the membrane, such that some of the potential ligand tunnels point into membrane while others point towards the aqueous solvent, suggests that the preferred tunnel for a given ligand depends on its hydrophilicity. Hydrophobic ligands would favor the tunnels to/from the membrane (2a and 2f). Hydrophilic ligands would favor the tunnels to/from the aqueous solution.

Since the flexible F-G block is inserted into the lipid bilayer and the B-C block is partly interacting with the bilayer, the membrane can affect the flexibility of the B-C and F-G blocks. In the simulations, the mobility of these regions of the protein is reduced in the membrane-bound form compared to the soluble form [30, 29, 24]. This suggests that the membrane could constrain the conformational changes in these regions of the protein and hinder the conversion between open and closed states. However, the simulations performed for the membrane-bound CYP2C9 are relatively short (10 ns) and this time is not long enough to allow the conformational changes between open and closed states to occur.

Other studies of membrane-bound CYPs give insights into the influence of the membrane on the dynamics of mammalian CYPs. In simulations of soluble and membrane-bound forms of CYP3A4 [24], tunnel 2b remained open in both the soluble and the membrane-bound forms whereas tunnel 2e stayed open in the soluble form and closed in the membrane-bound simulation. The movement of the F and G helices resulted in the closing of tunnel 2e and was observed only in the simulation of the membrane-bound form [24]. Berka et al. simulated CYP2C9 in a phospholipid bilayer and found the solvent tunnel to be most open and suitable for product egress [30].

1.6 Focus of this thesis

Although there have been extensive computational studies of both prokaryotic and eukaryotic P450s, the CYP51 family has not been studied a lot. Homology models have been built and studied for *Saccharomyces cerevisiae* [37], human [38] and *Penicillium digitatum* (plant) CYP51s [39]. Computational study for parasitic CYP51s, which are important drug target for antiparastic drug design [7], such as *Trypanosoma brucei*, have not been performed.

This thesis will focus on computational studies of *T. brucei* CYP51. It compares dynamics of *T. brucei* CYP51 with that of human CYP51 and human drug-metabolizing CYP51s in **Chapter 3** and this chapter is based on the published research paper by Yu et al. [40]. It investigates ligand tunnels in both *T. brucei* and human CYP51s in **Chapter 4**. **Chapter 5** is a study of interactions of different P450s with their redox partner, cytochrome P450 reductase (CPR) and **Chapter 6** is a study of the simulation of the complex of *T. brucei* CYP51 and human CPR in the membrane. In the last chapter, **Chapter 7**, a webserver built for the Simulation of Diffusional Association (SDA) software tool is described.

2

Methods

2.1 Introduction

Computational methods offer a way to study biological systems and have been developing very fast in recent years. Methods of computational biology use mathematical models to describe biological systems [41]. Computational methods give a complementary way to experiments to study biological systems on the microscopic scale.

The 2013 Nobel Prize in Chemistry was awarded to Martin Karplus, Michael Levitt and Arieh Warshel "for the development of multiscale models for complex chemical systems" (http://www.nobelprize.org/nobel_prizes/chemistry/laureates-/2013/). This award not only shows that computational studies are important for current molecular research, but also promotes the field of computational molecular studies.

Using computational methods can allow researchers to predict biomolecular interactions, such as protein-ligand binding, protein-protein binding, and to study the dynamics of biological systems, such as in this thesis, cytochrome P450 systems. There are many related theories and methods in molecular modeling and in this chapter, it is impossible to cover all aspects. Here, only theories and methods that were used in this PhD study are introduced and discussed.

2.2 Molecular Mechanics

Molecular mechanics (MM) is developed to model molecular systems using classical mechanics. In MM, interatomic energy terms are computed using force fields.

Force fields define the mathematical equations to compute forces on each atom. MM can be used to study interactions of macromolecules, investigate dynamics of biological molecules and design inhibitors for macromolecules. Using MM can reproduce experimentally determined properties, such as dynamics of proteins, protein-protein association, as described in this thesis in Chapters 3, 4, 5 and 6.

2.3 Force Fields

Different force fields are used depending on the properties of the system of study. All-atom (AA) force fields are used when every atom of a system is considered. This section will introduce AA force fields. In united-atom (UA) force fields, only non-hydrogen and polar hydrogen atoms are treated as particles. The effects of the non-polar hydrogen atoms are subsumed into the parameters of the non-hydrogen atoms that they are bonded to. UA force fields are beyond the scope of this thesis and will not be introduced in detail. Various force fields can be applied and will be introduced in the next section.

2.3.1 Components of a Force Field

A simple force field consists of bonded and non-bonded potential energies:

$$E_{total} = E_{bonded} + E_{nonbonded} \tag{2.1}$$

The total energy of the system E_{total} is the sum of the energies of the bonded terms E_{bonded} and the non-bonded terms $E_{nonbonded}$. The bonded interactions E_{bonded} include the bond stretching (bond) E_{bond}, the bond-angle bending (angle) E_{angle} and the dihedral terms $E_{dihedral}$:

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} \tag{2.2}$$

The non-bonded interactions $E_{nonbonded}$ include the electrostatic $E_{electrostatics}$ and the van der Waals energies $E_{van\ der\ Waals}$:

$$E_{nonbonded} = E_{electrostatics} + E_{van\ der\ Waals} \tag{2.3}$$

2.3.2 Bonded Interactions

Bonded interactions describe the forces between atoms connected by chemical bonds. These interactions are short-range interactions and stronger than non-bonded interactions.

2.3.2.1 Bond stretching

The bond stretching term describes the energy to stretch a chemical bond between two atoms. One common approach is to use the Hooke's law formula, which is a harmonic potential:

$$E_{bond} = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 \tag{2.4}$$

E_{bond} is the total bond stretching energy. k_i is the stiffness of the bond spring. l_i is the bond length and $l_{i,0}$ is the equilibrium bond length. There are other forms for the bond stretching energy, such as the Morse potential. Bond stretching energy is normally the strongest of interactions and depends on the type of atoms.

2.3.2.2 Angle bending

The angle bending term describes the energy to change the angle between three atoms connected by chemical bonds. Similarly to the bond stretching energy, it can be described using Hooke's law:

$$E_{angle} = \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 \tag{2.5}$$

E_{angle} is the total angle bending energy. k_i is the stiffness of the angle spring. θ_i is the bond angle and $\theta_{i,0}$ is the equilibrium bond angle.

2.3.2.3 Torsional terms

The bond rotation describes the energy to rotate an atom around a dihedral bond angle between four atoms connected by chemical bonds.

$$E_{dihedral} = \sum_{torsions} \frac{V_n}{2}(1 + cos(n\omega - \gamma)) \tag{2.6}$$

$E_{dihedral}$ is the total bond rotation energy. n is the periodicity of the curve of the potential function. γ is a position on the ω axis of the potential curve. V_n is the "barrier" height. There are other expressions for the torsional potential and this is one term that often varies among force fields. For some bonds, there can be more than one term from the cosine series expansion.

Improper torsions are introduced to force fields as an additional term which uses a harmonic or periodic function to keep planar carbons and the three atoms bonded to it in the same plane.

2.3.3 Non-bonded Interactions

The non-bonded interaction terms describe the interactions between atoms that are not connected by chemical bonds. Interactions of atoms that are further away than a certain distance are normally very weak and ignoring these interactions will not change properties of a system. Thus, for the computation of non-bonded interactions, a "cutoff" is introduced. The cutoff defines the range for the force calculation and it can be defined and may vary between different systems. Non-bonded interactions that are out of the range of the cutoff will not be computed. The non-bonded interactions include electrostatic interactions and van der Waals interactions. The electrostatic interactions are described by the Coulomb potential and the van der Waals interactions are often described by the Lennard-Jones potential. The sum of both non-bonded interactions is:

$$E_{nonbonded} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}} \right) \tag{2.7}$$

The meaning of the symbols is explained in the next two sections.

2.3.3.1 Electrostatic Interactions

Electrostatic interactions are the attractive interactions between electropositive elements and electronegative elements and the repulsive interactions between elements with the same sign property. In classical force fields, partial charges of atoms are used to compute electrostatic interactions. Coulomb's law can be used

to calculate electrostatic interactions between two molecules or between different parts of the same molecule [41]:

$$E_{elec} = \sum_{i=1}^{N_A} \sum_{j=i+1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}} \tag{2.8}$$

N_A is the number of partial charges on the first molecule and N_B is that on the second molecule. The formula sums pair-wise interactions between each pair of atoms of the two molecules. q_i and q_j are the partial charges on the atom of the first and second molecule, respectively. $\epsilon_r\epsilon_0$ is the permittivity of the environment and r_{ij} is the distance between the two atoms.

2.3.3.2 Van der Waals Interactions

Van der Waals interactions are the forces not only between neutral atoms, but also between charged atoms. These forces include the force between two permanent dipoles (Keesom force), the force between a permanent dipole and a corresponding induced dipole (Debye force) and the force between two induced dipoles (London dispersion force). The Lennard-Jones potential is used to approximate van der Waals interactions:

$$E_{VdW} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right) \tag{2.9}$$

σ_{ij} is the collision diameter and ϵ_{ij} is the depth of the potential well. r_{ij} is the distance between two atoms. The Lennard-Jones potential is widely used in force fields to calculate van der Waals interactions of molecules.

2.3.4 Different Types of Force Field

Different force fields were developed to serve different purposes, but nowadays most force fields can be used to solve similar problems. Some of the most popular force fields are: the Assisted Model Building with Energy Refinement (Amber) force field [42], the Chemistry at HARvard Macromolecular Mechanics (CHARMM) force field [43] and the GROningen MOlecular Simulation (GROMOS) force field [44]. There are many papers comparing the differences between force fields to be applied to different systems but it is difficult to determine which force field is better than the others [45]. Choosing which force field to use depends on properties

of the system of study and the questions to answer. Here, the Amber force field, which is used in all-atom simulations of this thesis, will be introduced.

2.3.5 Amber Force Field

There have been many different versions of the Amber force field, such as ff94 [42], ff96 [46], ff99 [47], ff03 [48], ff10 [49], ff99SB [50], ff99SB-ILDN [51] and ff99-PHI [52]. These force fields can be used for amino acids and nucleic acids. The Generalized Amber Force Field (GAFF) is developed for small molecules such as organic compounds and drug-like molecules [53]. Amber force fields are implemented in simulation software packages such as AMBER [54] and NAMD [55].

The functional form of the Amber force field [42] is the sum of the previously mentioned energy terms:

$$\begin{aligned} E_{total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \\ \sum_{dihedrals} \sum_{n} \frac{V_n}{2}[1 + cos(n\phi - \gamma)] + \\ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \right\} \end{aligned} \tag{2.10}$$

In the Amber force field, experimental data and calculated quantum mechanical data are used to refine the parameters to reproduce the experimental observations.

AMBER force field uses the restrained ESP-fit (RESP) charge model [42]. RESP is from the electrostatic potential (ESP) fit atomic centered charges derived from quantum mechanical calculations using a 6-31G* basis set [56]. The major problem of the ESP fit is that it was derived to reproduce intermolecular properties but not intramolecular properties [57]. The RESP charge model was then developed by Bayly et al. [57] to address problems of ESP fitting. This avoids large variations of atomic charges for different conformations of molecules and improves derived atomic charges for buried atoms [42]. RESP is a two-stage approach: weak constraints on methyl groups without forced symmetry are applied in the first stage

and the second stage is to refit the methyl groups using forced symmetry and strong constraints [57].

The ff96 [46] and ff99 [47] force fields are improved by fitting parameters describing the ψ and ϕ potentials to tetrapeptide and dipeptide quantum conformational energies [58]. The ff99SB force field improves the ϕ/ψ dihedral term in the ff99 energy function [50]. This solves the problem of having unreasonable conformations of glycine in ff99 [50].

2.4 Energy Minimization

2.4.1 Introduction

As previously described, the energy or potential energy of a system is the sum of all particle/atom interaction terms of the system. Energy changes when the coordinates of atoms of the system change. The dependence of energy on the coordinates of atoms is referred to as the "potential energy surface". It is important in molecular modeling to find minimum points on the energy surface and how coordinates change from one minimum point to another. The algorithm to find minima on the energy surface is referred to as a "minimization algorithm".

Minimization is the process of finding with which set of coordinates, the energy of the system has its local minimum. This can be described as: given a system with the energy of E, when E has its minimum, the first derivative of E is zero and second derivatives are all positive:

$$\frac{\partial E}{\partial x_i} = 0; \frac{\partial^2 E}{\partial x_i^2} > 0 \tag{2.11}$$

Solving the equation completely is normally impossible. Thus, minimization in molecular modeling is finding a set of Cartesian coordinates which gives the local minimum of E. Most minimization algorithms search for the nearest minimum, rather than the global minimum. The minimum is indicated by the direction of the first derivative. The steepness of the local minimum is indicated by the magnitude of the gradient.

There are many minimization algorithms that are used in molecular modeling. Here in this chapter, the steepest descents method and the conjugate gradient method will be introduced. These two algorithms are first order minimization methods. In first order minimization methods, the current configuration of the

system (k) is the starting point for the next iteration (k + 1). The initial starting configuration is the user input, which is often obtained from the PDB structure of the protein.

2.4.2 The Steepest Descents Method

The steepest descents method walks down the energy "hill" in the direction of the net force on the system. On a two-dimensional energy surface, a line can be drawn from the starting point along the direction of the net force. On the line, the local minimum will be passed through and then the local minimum can be chosen for the starting point of the next step.

2.4.3 The Conjugate Gradient Method

One problem of the steepest descents method is the oscillatory behavior in narrow valleys. Instead of having both the gradients and the direction of successive steps orthogonal, in the conjugate gradient method, the directions are conjugate but the gradients at each point are orthogonal [41]. For conjugate directions, if a quadratic function has M variables, the minimum can be reached in M steps.

2.5 Molecular Dynamics Simulation

2.5.1 Introduction

Molecular dynamics simulation is a way to explore the dynamic properties of biological systems. Following energy minimization, MD simulation is used to investigate kinetic and thermodynamic properties of systems. MD simulations use Newton's equations of motion to compute the forces on atoms of the system in a defined time interval, typically a timestep of 1-2 fs. The timestep should be defined to be shorter than the fastest motion of atoms. A collection of microscopic states of the system is called an *ensemble*. The average properties of an ensemble can be compared with macroscopic experimental data to reproduce or predict experimental data.

Certain state variables are kept constant in simulations and thus different statistical ensembles can be produced. Popular ensembles include: NPT which has

constant number of particles (N), pressure (P) and temperature (T) in the simulation of the system and NVT which keeps number of particles (N), volume (V) and temperature (T) constant during the simulation.

2.5.2 Periodic Boundary Conditions

Avoiding boundary effects is important for simulations to reproduce macroscopic data. A common boundary condition used in simulations of biological systems is the *periodic boundary condition*. The periodic boundary condition mimics a "real" environment by positioning molecules in a box with a defined size. The box can be cubic, octahedral or have other shapes. The cubic box is most widely used. It treats the periodic condition identically in all three dimensions. In a simulation, when defining the minimum necessary size, the distance between the surface of molecules and the edge of the box should be greater than the cutoff. In a periodic box, when an atom diffuses out of a boundary of the box, it is considered to enter the opposite side of the box. In this way, the system simulated, for example a protein, keeps as whole in the simulation regardless of diffusion.

2.5.3 Non-bonded Cutoffs

In an MD simulation, the most computationally expensive part is to compute the non-bonded forces. The bonded terms, which are bond-stretching, angle-bending and torsional terms are proportional to the number of atoms/residues of a system. However, the non-bonded terms should be calculated between each pair of atoms and the algorithm to calculate of the non-bonded terms is N^2 where N is the number of atoms. To simplify the computation, cutoffs are used to truncate the non-bonded interactions to only atom pairs within a certain distance. In this way, the non-bonded interactions between atoms that are "far away" from each other are set to zero. Using improper cutoffs might lead to wrong simulations.

However, even using cutoffs, it is still computationally expensive to compute non-bonded interactions. This is because distances between each pair of atoms need to be computed which is a similar magnitude effort to computing all the energies. The "non-bonded neighbor" list can be used to further simplify the computation. The non-bonded neighbor list records neighboring atoms of a given atom. Because the neighbors of the atom are recorded, only interactions between the atom and its neighbors need to be computed, rather than the pair-wise inter-

actions of all atoms of the system. The cutoff for neighbor lists should be larger than the cutoff of non-bonded interactions to make sure that all atoms within the interaction range are in the list.

However, cutoffs make the potential energy and forces discontinuous. To eliminate this, a shifted potential is used by subtracting a constant term from the potential. Another approach, which is used in most simulation software packages, is to use a "switching function". The switching function is the multiplication of the potential energy by a polynomial function.

2.5.4 Electrostatics

Electrostatic forces are long range interactions. The charge-charge interactions have to be properly modeled to reproduce the macroscopic properties of systems. There are various methods to compute long range forces.

One such method is the Ewald summation method [59]. The Ewald summation is computationally very expensive and many tradeoffs to accelerate the computation have been introduced. One solution is to use fast Fourier transforms (FFT) to compute the reciprocal space summation and a grid-based approach such as the particle-mesh Ewald (PME) method [60]. The PME method maps charges onto grid points.

2.5.5 Constraint Dynamics

It is possible to constrain the motion of atoms, for example, the covalent bond between two atoms. Constraints are requirements that the system has to satisfy. In molecular dynamics, the most commonly used method to constrain systems is the SHAKE algorithm [61]. The SHAKE algorithm is used to perform MD simulations with fixed bond lengths. In constraint dynamics, the mathematical equations are solved by taking the constraints into account.

2.5.6 Constant Temperature and Pressure

Molecular dynamics simulation naturally samples NVE ensemble in which the system has constant volume and energy. But two other ensembles are often used to compare simulation results with experiments. These are the NVT ensemble and the NPT ensemble as introduced previously.

2.5.6.1 Constant Temperature Dynamics

Constant temperature is often applied to molecular dynamics simulations. The temperature is kept constant to investigate the dynamics of systems, such as proteins. Other approaches may use different temperatures. For an unconstrained system, the temperature and the time average of kinetic energy are related as:

$$\langle E_k \rangle_{NVT} = \frac{3}{2} N k_B T \tag{2.12}$$

E_k is the kinetic energy of the system. k_B is the Boltzmann constant and T is the temperature.

One way to control the temperature introduced by Berendsen [62], is to couple the system to an external heat bath. The temperature of the heat bath, T_{bath} is then the desired temperature. The velocities are rescaled at each timestep of simulations:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_{bath} - T(t)) \tag{2.13}$$

τ is the coupling factor.

2.5.6.2 Constant Pressure Dynamics

Constant pressure condition is another often used condition in molecular dynamics simulations. Most experiments are done under constant pressure and thus, to compare with these experiments, it is necessary to keep the pressure constant. When the pressure is constant, the volume of the simulation cell changes. The volume fluctuation is determined by the isothermal compressibility, κ:

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{2.14}$$

In an isobaric simulation, the volume change can be facilitated by changing the volume in all three directions or in one direction. Changing three directions is often used for simulations of soluble systems, such as protein in water. Whereas changing one direction is often used for simulations of membrane-bound systems, which the volume change is facilitated by just changing the volume in the direction perpendicular to the membrane plane, which is often z direction.

Many pressure control methods are similar to those of temperature control. One commonly used method is the Berendsen pressure control. Instead of having

a heat bath of constant temperature, a "pressure bath" of pressure, P_{bath} is coupled to the system. The rate of change of pressure is:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p}(P_{bath} - P(t)) \tag{2.15}$$

τ_p is the pressure coupling factor.

The "extended system" method of Anderson [63] can also be applied to pressure control. In this method, an extra degree of freedom is introduced to the system, which is similar to a piston acting on the system.

2.6 Solvent

Solvent is an important factor for simulating biological systems. There are two major classes of solvent models: *implicit solvent* model and *explicit solvent* model. Implicit solvent models the solvent environment without having "real" solvent molecules. One of the most used models is the Generalized Born model. Implicit solvent is computationally more efficient because of having no solvent particles but less "realistic". Explicit solvent allows systems to have explicit solvent molecules, such as water molecules. This approach is closer to the "real" environment than the implicit solvent. However, introducing solvent molecules causes the computation to be more expensive.

Molecular dynamics simulations are more often performed with explicit solvent molecules, mostly water molecules. Using aqueous solution is a condition that is similar to biological environments. It is important to simulate proteins in solvent to study the behavior of proteins. Different water models have been developed for molecular dynamics simulations. The simplest models are the three-site transferable intermolecular potential (TIP3P) water, and the simple point charge (SPC) water. Other water models include the four-site transferable intermolecular potential (TIP4P) water and the five-site transferable intermolecular potential (TIP5P) water.

2.6.1 TIP3P Water Model

TIP3P water model is a three-site model [64]. Each water molecule has three "atoms" representing the one oxygen and two hydrogen atoms of a water molecule. All the three sites have partial charges and the oxygen site also has parameters for

Lennard-Jones interactions. TIP3P water is widely used in modern molecular dynamics simulations because of its simplicity and computational efficiency.

2.6.2 SPC Water Model

The simple point charge (SPC) model is also a three-site water model [65]. Similar to the TIP3P water model, it is a rigid model without having internal degrees of freedom [65]. Experimental values are used to define many of the parameters of this water model.

2.7 Coarse-grained Simulation

2.7.1 Introduction

All-atom (AA) molecular dynamics simulation offers a great tool to study biological systems computationally, yields a great amount of information about biological systems and allows researchers to predict the output of experiments. However, one challenging problem of all-atom simulations is that for large systems, such as those containing millions of atoms, the computation can be too expensive. This limits the use of simulations for large systems, such as a system of proteins on a membrane within a water box. Depending on the question to answer, another limitation may be that the simulation time is too short to investigate biological processes. The microscopic time scale is normally a few nano-seconds, whereas the macroscopic time scale is often a few micro-seconds or even longer. Researchers tend to simulate a system "long" enough to make their observations closer to the reality. However, because of the limitation of computational power, currently a simulation can rarely be performed for a time scale of more than a few micro-seconds.

Coarse-grained (CG) methods are thus developed to counter these problems. Coarse-grain methods simplify systems by introducing *beads* to represent groups of atoms of a molecule. For example, for a protein, instead of representing all atoms of amino acids as individual particles, a coarse-grained model might use one bead to represent all backbone atoms and one to three beads to represent all side chain atoms, depending on the size of the side chain. In this way, the number of particles in the system is reduced five to ten times in comparison with the all-atom system. Coarse-grained simulations also allow much longer timesteps as

high frequency motions are removed. As a result, with the same amount of computational power, it is possible to run much longer simulations and simulations of much bigger systems. For example, for a system of membrane-bound CYP51, it is possible to simulate 3 μs in 48 hours using a coarse-grained model in comparison to 8 ns using all-atom model with the same computational resources, as discussed in Chapter 3.

2.7.2 Martini Coarse-grained Model

Various coarse-grained models have been developed. One of the most commonly used methods is the Martini coarse-grained model for biomolecular systems. Martini was initially developed to simulate a large number of lipids [66] and later extended to protein systems [67]. Martini has been proven to perform well for lipid systems [66] and can also be used to study membrane-bound proteins. For example, a membrane-bound model of cytochrome P450s can be simulated to determine the preferred orientation of a P450 protein in a lipid bilayer [29, 40].

The Martini coarse-grained model maps multiple heavy atoms to one Martini bead [66]. Four types of interaction sites are introduced, which are polar (P), non-polar (N), apolar (C) and charged (Q). Subtypes are derived from these four basic types [66].

In the Martini force field, the non-bonded and bonded interactions are treated similarly to the all-atom force field introduced previously [66]: Solvent is treated similarly to heavy atoms. One solvent bead represents four water molecules. Antifreeze particles have to be introduced to avoid freezing of water molecules [66]. The time scale in coarse-grained simulations is different from that in all-atom simulations. The time scale is normally 2- to 10-fold longer than all-atom simulations [66].

Martini coarse-grained method was extended for amino acids after it was developed for lipids. Mapping for amino acids is shown in Figure 2.1. For interactions of coarse-grained amino acids, the non-bonded terms and bonded terms are implemented similarly to those of lipids.

2.7.3 Elastic Network Model

An elastic network model is a structure-based method for coarse-grained simulations [68]. It can be used to describe the structural properties of a system, such

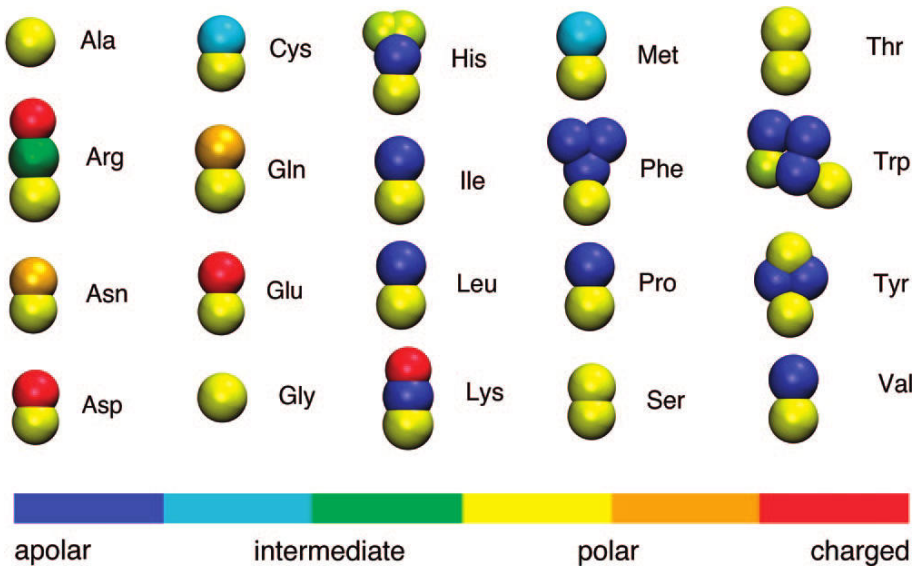

Figure 2.1: Mapping of all-atom amino acids to coarse-grained amino acids. Figure taken from [67].

as a protein. For proteins, when the elastic network model is not used, results of coarse-grained simulations are not in good agreement with all-atom simulations [68]. However, with elastic network, the results of global and local deformation of proteins in coarse-grained simulations are in agreement with all-atom simulations [68].

Interaction terms of the normal Martini force field for this elastic network model are slightly modified, because the backbone bead, in this model, is positioned at Cα atoms [68]. The elastic network uses a simple spring model with only two parameters, R_c and K_{spring} [68]. R_c and K_{spring} have to be tuned for different systems to be comparable to all-atom simulations. In the initial study, using R_c ranging from 0.8 to 1.0 nm and K_{spring} from 500 to 1000 kJ mol^{-1} nm^{-2} gives consistent results in comparison with all-atom simulations of proteins.

2.7.4 Reverse Conversion

Coarse-grained simulations allow longer simulations to be performed. However, in many cases, particularly in the study of proteins, the lack of atom level information might lead to unrealistic conclusions. It is necessary, in many cases, to start with coarse-grained simulations to get information from long simulations that all-atom simulations can not offer, for example, how a P450 protein orients on a lipid bilayer. But all-atom simulations are required to study more detailed infor-

mation of the system. Thus, it is important to allow the reverse conversion from coarse-grained model to all-atom model. After that, all-atom simulations can be performed using models obtained from a coarse-grained simulation as a starting point.

Many tools have been developed to convert a Martini coarse-grained model back to an all-atom model using the mapping information in topology files. There are generally two steps of the reverse conversion. The first step is to backmap/-superimpose atoms, that have been coarse-grained, onto coarse-grained beads. The following step is to refine the superimposed all-atom structure. A common approach is to use simulated annealing (SA). As previously discussed, simulated annealing heats the system to a high temperature rapidly and then cools it down slowly to allow the structure to explore its conformational space to find the most stable conformation. Recently, a new tool named "Backward", which is able to transform both coarse-grained proteins and lipids back to atomistic structures, has been developed [69].

2.8 Brownian Dynamics Simulation

2.8.1 Introduction

All-atom and coarse-grained simulations are two approaches to simulate microscopic properties of biological systems and thus connect experiments with computational studies. Brownian dynamics (BD) is a mesoscopic (between macroscopic and microscopic) approach to simulate biological systems. It allows simulations of larger time scales in comparison with all-atom simulations. Thus, Brownian dynamics can be used to study biological processes, such as protein-protein association.

2.8.2 Brownian Dynamics

Brownian motion is the motion of particles in random, small steps in fluids. Brownian motion can be investigated by techniques such as light microscopy and was observed by Robert Brown [70]. In Brownian dynamics simulations, particles

move by systematic forces and stochastic collisions with solvent molecules [71]. In Brownian motion, the mean displacement Δr of a particle in time Δt is:

$$\langle \Delta r^2 \rangle = 6D\Delta t \tag{2.16}$$

D is the translational coefficient. D can be calculated for spherical particles using Boltzmann constant k_b, absolute temperature T, solvent viscosity η and radius of the particle, a:

$$D = k_b T / 6\pi\eta a \tag{2.17}$$

The method developed by Ermak and McCammon [72] is often used to simulate protein-protein association [71]. With this method, both translational and rotational motions are considered.

The translational displacement is calculated as [71]:

$$\Delta \mathbf{r} = \frac{D\Delta t}{k_b T}\mathbf{F} + \mathbf{R} \tag{2.18}$$

D is the translational diffusional coefficient. **F** is the net force on the particle and vector **R** satisfies $\langle \mathbf{R} \rangle = 0$ and $\langle \mathbf{R}^2 \rangle = 6D\Delta t$. The rotational displacement is calculated as [71]:

$$\Delta \mathbf{w} = \frac{D_R \Delta t}{k_b T}\mathbf{T} + \mathbf{W} \tag{2.19}$$

Similar to the equation of translational displacement, D_R is the rotational diffusional coefficient. **T** is the torque on the particle and **W** satisfies $\langle \mathbf{W} \rangle = 0$ and $\langle \mathbf{W}^2 \rangle = 6D_R\Delta t$.

Brownian motion can be studied computationally using Brownian dynamics simulations. Brownian dynamics simulations can be used to simulate systems in which solutes are notably larger than solvent molecules. For a system of two solutes, such as two proteins, Brownian dynamics simulation can be performed to investigate the encounter complexes of the two solutes and calculate the association rates of the two solutes. Brownian dynamics can also be performed to simulate systems of hundreds of solutes, which gives insights into the behavior of solutes in an environment surrounded by many others of the same kind or different kinds of solutes.

2.8.3 The Simulation of Diffusional Association (SDA) Software

To simulate the diffusional association of proteins, the Simulation of Diffusional Association (SDA) software package [73, 71] was developed to perform Brownian dynamics simulation [73, 71]. SDA allows simulation of protein-protein association which is crucial for many biological processes. The force field of SDA considers interaction forces, such as electrostatic interactions, electrostatic desolvation interactions, and hydrophobic desolvation (non-polar, soft-core) interactions. Rigid body simulations can be performed with SDA. Rigid body simulations neglect the internal flexibility of solutes. However, using different conformations of solutes to perform rigid body docking, flexibility can be introduced into BD simulations and this has been implemented in the newest version of SDA, SDA version 7 (SDA7, `http://mcm.h-its.org/sda7/`).

2.8.3.1 Interactions

As previously discussed, electrostatic, electrostatic desolvation and non-polar desolvation interactions can be used in SDA. The interaction energy of two solutes is calculated as follows [74]:

$$\begin{aligned}\Delta G = &\frac{1}{2}\sum_{i2}\Phi_{el1}q_{i2} + \frac{1}{2}\sum_{j1}\Phi_{el2}q_{j1} \\ &+ \sum_{i2}\Phi_{edesolv1}q_{i2}^2 + \sum_{j1}\Phi_{edesolv2}q_{j1}^2 \\ &+ \sum_{m2}\Phi_{npdesolv1}\text{SASA}_{m2} + \sum_{n1}\Phi_{npdesolv2}\text{SASA}_{n1}\end{aligned} \tag{2.20}$$

The first two terms are electrostatic interactions and the second two terms are the electrostatic desolvation interactions. The last two terms are the non-polar interactions. Φ_{el1} and Φ_{el2} are the electrostatic potential of the two solutes. $\Phi_{edesolv1}$ and $\Phi_{edesolv2}$ are the electrostatic desolvation potential of the two solutes. q_{i2} and q_{j1} are the charges on the two solutes. $\Phi_{npdesolv1}$ and $\Phi_{npdesolv2}$ are the "burial potentials" and, SASA_{m2} and SASA_{n1} are the solvent accessible surfaces of atoms $m2$ and $n1$.

2.8.3.2 Effective Charges

The effective charge method (ECM) was developed to decrease computational expenses of solving the Poisson-Boltzmann equation with partial atomic charges on all atoms [75]. Effective charges for charged residues are assigned to certain atoms. For example, for charged amino acids, charges are assigned to the "NZ" atom of lysine, the "NH1" and "NH2" atoms of arginine, the "OD1" and "OD2" atoms of aspartic acid and, the "OE1" and "OE2" atoms of glutamic acid. The effective charges are computed to reproduce the electrostatic potential with an inhomogeneous dielectric in a shell outside of the solute with a homogeneous dielectric.

2.8.3.3 Electrostatics

The electrostatic energy of SDA is computed as [73, 71]:

$$\Delta G_{el} = \frac{1}{2}\sum_i q_{i1}^{eff}(r)\phi_2^{el}(r) + \frac{1}{2}\sum_j q_{j2}^{eff}(r)\phi_1^{el}(r) \tag{2.21}$$

$q_{i1}^{eff}(r)$ and $q_{j2}^{eff}(r)$ are effective charges of the two solutes, that were fitted to reproduce the electrostatic potentials $\phi_1^{el}(r)$ and $\phi_2^{el}(r)$, respectively. The electrostatic potential of two solutes are pre-computed and assigned to grids centered on the solutes. A Poisson-Boltzmann electrostatic potential value $\phi_{1/2}^{el}$ is assigned to each grid point. Two common software packages/tools that can be used to generate electrostatic grids are the Adaptive Poisson-Boltzmann Solver (APBS) [76] and the University of Houston Brownian Dynamics (UHBD) [77].

2.8.3.4 Electrostatic Desolvation

The electrostatic desolvation energy of SDA is computed as [78]:

$$\begin{aligned}\Delta G^{edesolv} &= \sum_i \phi_{edesolv}(i) q_i^{eff}(r) \\ &= \alpha \frac{\epsilon_s - \epsilon_p}{\epsilon_s(2\epsilon_s - \epsilon_p)} \sum_i \sum_j a_j^3 \frac{1 + kr_{ij}}{r_{ij}^4} \exp(-2kr_{ij})\end{aligned} \tag{2.22}$$

ϵ_s is the relative dielectric permittivity of the solvent and ϵ_p is that of the protein. k is the Debye-Hückel constant. The electrostatic desolvation is the penalty

of interactions of charged atoms with cavities of the other protein. a_j^3 is the radius to define solute atoms for the penalty. r_{ij} is the distance between the i^{th} charged atom of the first protein and the j^{th} cavity of the second protein.

2.8.3.5 Hydrophobic Desolvation

The non-polar energy of SDA is computed as [74]:

$$\phi^{npdesolv}(r) = \beta \mathrm{c} \begin{cases} 1 & \mathrm{r < a} \\ (b-r)/(b-a) & \mathrm{a < r < b} \\ 0 & \mathrm{r > b} \end{cases} \tag{2.23}$$

β is a factor to convert the calculated buried area to non-polar desolvation energy. c is a factor. The non-polar desolvation energy is assigned to each grid point. The maximum distance that an interacting solute atom can lie from the molecular surface is defined by a and b.

2.8.3.6 SDA Docking

The SDA software package has a number of functions. Three of the most popular functions used are: *SDA docking*, *SDA association* and *SDA multiple molecules*.

SDA docking is used to generate structures of complexes that are driven by diffusional motions for two input solutes. The two input solutes can be two proteins, protein and DNA, protein and ligand, and so on. SDA employs Brownian dynamics to perform the sampling of the positions of the solutes. Interactions via electrostatic and non-polar intermolecular forces can be accounted for. SDA docking is particularly helpful when there are biochemical data about structures of encounter complexes. The biochemical data can then be used to introduce distance constraints to SDA docking to predict complexes of proteins as shown in [79, 80, 81, 82]. The docked encounter complexes can then be refined by molecular dynamics simulations considering the flexibility of the solutes [83]. Using molecular dynamics simulation to refine encounter complexes can improve the reproduction of experimentally determined structures of protein complexes [83].

2.8.3.7 SDA Association

The SDA association method can be used to compute the bimolecular diffusional association rate constant for two input solutes. The relative diffusional motion of

the two solutes is simulated by Brownian dynamics subject to interaction forces [72].

2.8.3.8 SDA Multiple Molecules

SDA docking and SDA association use Brownian dynamics to simulate two solutes. However, in the cellular environment, macromolecules are surrounded by many other macromolecules, of the same kind or different kinds. The SDA multiple molecules (SDAMM) method simulates the diffusive motion of several hundred macromolecules, subject to electrostatic, non-polar and steric intermolecular forces. SDAMM allows simulations with several types of molecule, and different conformations of each type of molecule. These simulations can be performed to investigate the behavior of a molecule in a crowded environment, similar to the cellular environment, to study physical and biological properties of molecule [84, 85, 86, 87, 88].

2.9 Web Interfaces for Scientific Software

Scientific software is designed to serve certain scientific purposes. Running scientific software normally needs many inter-dependent parameters and produces multiple output files. Because of this nature, it is often not easy for beginners to use a new scientific software tool. However, developing web interfaces for scientific software provides a way to make scientific software more user-friendly, provide automatic generation of input parameters and easy visualization for results. There are various frameworks to build web servers. In this section, the Play framework and programming languages and tools used for webserver development will be introduced.

2.9.1 Play Framework

The Play framework (`https://www.playframework.com/`) is a modern model-view-controller (MVC) framework which has been used by many well-known websites such as LinkedIn (`https://www.linkedin.com/`) and Klout (`https://klout.com/home`). Two programming languages can be used in Play: Java (`https://www.java.com/en/`) and Scala (`http://www.scala-lang.org/`). Java is one of the most popular programming languages and there are plenty of packages writ-

ten in Java to support software development. Java is also a static typing language which requires strict typing of variables.

Play can be used to build **RE**presentational **S**tate **T**ransfer (REST) webservers. A RESTful webserver is stateless, cacheable and has a layered system. By following the workflow of Play, the webserver built is a RESTful webserver. Play uses the object relational mapping (ORM) technique to convert Java objects to entries in relational databases and vice versa. Older versions of Play use the Ebean ORM and newer version of Play use Java Persistence API (JPA) (`https://www.playframework.com/`). The ORM of Play is developed for persisting objects to relational databases and retrieving data from relational databases to construct objects. It is easy to use but the performance of persisting or retrieving big data is not very good because of the difficulty of converting two intrinsically different things.

Play also supports the design of asynchronous webservers, to allow webservers to run jobs in parallel or asynchronously. Akka (http://akka.io/) can be used to manage concurrent jobs on webservers and is integrated in Play. Actors of Akka, which are messages to manage job execution, allow webservers to parallelize jobs non-blockingly and the actor system is a very lightweight event-driven processes which allows web servers to create and manage millions of actors.

2.9.2 Back-end Tools for webSDA

The Adaptive Poisson-Boltzmann Software (APBS, `http://www.poissonboltzmann.org/`) is a tool to solve Poisson-Boltzmann equation for electrostatics of biological macromolecules, such as proteins and nucleic acids, upon solvation. The Simulation of Diffusional Association (SDA), as discussed previously, is a software package to perform Brownian dynamics simulations to study biological systems. Both software packages are powerful scientific software and can be used by running batch/shell commands.

The Apache Commons Exec (`http://commons.apache.org/proper/commons-exec/`) is a powerful tool to execute external processes from Java. It provides a reliable way to run commands, check the status of commands and reports error if necessary. It can be used together with Akka to manage multiple concurrent jobs, which make web servers robust.

2.9.3 User Interface Tools

To design web interfaces, HTML5 and Cascading Style Sheets (CSS) are most commonly used. HTML5 is supported by all modern web browsers, such as Firefox, Internet Explorer and Google Chrome. Bootstrap (`http://getbootstrap.com/`) is a popular package for developing front-end web interfaces. It combines HTML, CSS and JavaScript and allows fast and flexible development of web interfaces.

JavasSript is a programming language that allows the client-side scripts to interact with users dynamically. JQuery (`http://jquery.com/`) and Ajax (`http://en.wikipedia.org/wiki/Ajax_(programming)`) are different extensions of JavaScript. jQuery is an open-source JavaScript library. Ajax supports asynchronous interactions between the front-end and the back-end of webservers. By using the above mentioned languages/techniques, dynamic, interactive and user-oriented web servers can be built.

3

Comparison of P450s

3.1 Introduction

Sterol 14-α demethylase (14DM or cytochrome P450 family 51 (CYP51)) enzymes form a cytochrome P450 (CYP, P450) family which is essential in sterol biosynthesis [7, 6]. The catalytic function of CYP51 is to remove the 14α-methyl group from cyclized sterol precursors such as lanosterol [6]. CYP51 is the most ancient CYP family and has been found in all the biological kingdoms [6, 89, 90, 91]. Drugs (azoles such as itraconazole) that inhibit CYP51 in fungi have been used to treat human fungal infection for more than 20 years [92, 93]. CYP51 is also considered a good target for antiprotozoan drugs [94]. The inhibitor N-[(1R)-1-(2,4-dichlorophenyl)-2-(1H-imidazol-1-yl)ethyl]-4-(5-phenyl-1,3,4-oxadiazol-2-yl)benzamide (VNI) has been shown to inhibit CYP51 in protozoan pathogens, such as *Trypanosoma brucei* and *Trypanosoma cruzi*, specifically but not to affect the human CYP51 [7]. VNI was tested in mouse models of the acute and chronic forms of Chagas disease and the VNI-treated mice were all able to survive without severe observable side effects [95]. Other inhibitors have also been found that target parasitic CYP51s such as 14α-methylenecyclopropyl-Δ7-24,25-dihydrolanosterol (MCP, LNP) [96] and (R)-4'-chloro-N-(1-(2,4-dichlorophenyl)-2-(1H-imid-azol-1-yl)ethyl)biphenyl-4-carboxamide (VNF) [97, 98, 99].

There are differences in many of the properties of CYP51 compared with other P450s. The very low root-mean-square deviation (RMSD) (0.75 Å) between the Cα atoms of the ligand-free and VNI-bound crystal structures of *T. brucei* CYP51 suggests high structural rigidity of the enzyme [7]. This notion has been supported by the analysis of crystal structures of other eukaryotic CYP51 orthologs, including

T. cruzi, *Leishmania infantum* and human, in ligand-free forms and co-crystallized in complexes with different heme-binding inhibitors and with the substrate analog MCP [96, 97, 100].

In crystal structures of eukaryotic CYP51 enzymes, there is only one wide open ligand access tunnel [7]. This corresponds to tunnel 2f in Wade's nomenclature [12], and is lined by helices A' and F" and the tip of the β4 hairpin [7, 96, 97, 100]. This tunnel is open (to a probe with a radius of 1.4 Å, corresponding to the size of a water molecule) in all crystal structures and often accommodates the longest portion of the ligand molecule that, in the case of VNI, approaches the tunnel's outer entrance (Protein Data Bank (PDB) ID: 3GW9), and in the case of posaconazole, extends from the active site to the protein surface (PDB ID: 3K1O) [97].

Little is known about the influence of the membrane on CYP51, the dynamics of CYP51, and the opening and closing of ligand tunnels in CYP51. Here, we address the questions: (i) In which orientation is the *T. brucei* CYP51 embedded in the membrane and how is the flexibility of the protein influenced when it is in the membrane? (ii) Is CYP51 more rigid compared with other membrane-bound CYPs as suggested by the crystal structures? (iii) Why do inhibitors of parasitic CYP51s not inhibit the human ortholog? Is this because of different flexibility or different interactions of the binding cavity residues? (iv) How do the ligand tunnels open and close? Which tunnels can be used for ligand (including substrate, product, inhibitor and water) access and egress?

CYP51 is anchored in the membrane by a single N-terminal transmembrane helix and has part of its globular domain embedded in the membrane. The positioning of CYPs in the membrane has been studied experimentally and by computer simulation [101, 102, 103, 104, 105, 106, 107, 29, 30, 24, 108, 109]. A few CYPs, including CYP1A2, CYP2A6, CYP2C9, CYP2D6, CYP2E1, CYP3A4 and CYP19, have been modeled and simulated in phospholipid bilayers using different approaches [29, 30, 24, 110, 111, 112, 113]. The protocol described in ref. [29], which involves a combination of coarse-grained and all-atom molecular dynamics (MD) simulations, does not use prior knowledge about the protein-membrane interactions and can be applied to predict the orientation of CYPs in the membrane using solely the crystal structure of the globular domain.

To study the influence of the membrane on *T. brucei* CYP51, we built a model of *T. brucei* CYP51 in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)

bilayer using a procedure similar to that in ref. [29]. The positioning of CYP51 in the membrane in this model is in good agreement with available experimental data [101, 102, 104, 107, 111]. We refer to this model as the "membrane-bound" *T. brucei* CYP51. We also performed MD simulations of CYPs immersed in aqueous solvent without a bilayer. We refer to these models as "soluble" CYPs. Here, we describe these models and the analysis of the protein dynamics during the simulations. Simulations of several hundreds of nanoseconds duration have been performed for some soluble CYPs, including P450cam [114, 115] and the major human-drug metabolizing CYPs [24, 112, 20]. Here, we instead carried out a number of shorter simulations of different CYPs in soluble and membrane-bound forms in order to perform a comparative analysis of the dynamics of the active site and ligand tunnel regions. We focused on comparing simulations of the soluble models of *T. brucei* CYP51 and human CYP51 with human CYP2C9 and human CYP2E1, and simulations of the membrane-bound model of *T. brucei* CYP51 with membrane-bound models of human CYP2C9. The analysis reveals features specific to CYP51 relating to the flexibility of the active site and the heme propionate groups, ligand access and egress tunnels, the opening of the water tunnel and hydrogen-bonding networks. These have implications for the recognition of substrates and inhibitors by CYP51.

3.2 Materials and Methods

3.2.1 Models of Soluble *T. brucei* and Human CYP51

The crystal structure of the ligand-free form of *T. brucei* CYP51 (PDB ID: 3G1Q solved at 1.89-Å resolution) was used to build models for the simulations [116]. The first three residues of the N-terminus were mutated from GKL to the wild-type PTD sequence with the mutation tool in PyMOL [117]. The PDB2PQR server [117] was used to assign the protonation states of the ligand-free CYP51 at pH 7. HIS420 was doubly protonated. For the ligand-bound model, MCP (14α-methylenecyclopropyl-Δ7-24,25-dihydrolanosterol) was positioned in the active site of the ligand-free model with Visual Molecular Dynamics (VMD) [118] by using the transformation matrix for superimposing the protonated ligand-free model and the MCP-bound crystal structure (PDB ID: 3P99). This procedure could be applied because the cavity lining residues were in almost identical positions in the

two crystal structures. This model was used rather than the crystal structure of the complex to permit direct comparison of the ligand-free and ligand-bound systems.

For the simulation of human ligand-free CYP51, the crystal structure of human CYP51 (PDB ID: 3LD6 solved at 2.80-Å resolution) with ketoconazole removed was used. This structure was chosen because it has the best resolution. The same procedure to that for *T. brucei* CYP51 was used to protonate the structure. HIS167 and HIS447 (close to the heme center) were doubly protonated.

3.2.2 MD Simulations of Soluble CYP51

The modeled protein structures were further processed with the tleap program of AmberTools 1.5 [49]. The Amber ff99SB force field was used for the protein and the generalized Amber force field (GAFF) for the POPC phospholipids [50, 53]. The parameters for the heme center were assigned as in references [29, 119]. The partial charges of MCP were derived using a two-stage Restrained Electrostatic Potential (RESP) fit method: RESP-A1A (HF/6-31G$*$) of the R.E.D server [120, 121, 57]. The parameter files of the two ligands were generated using the Antechamber program of the AmberTools 1.5 with the GAFF [49, 53, 122]. The systems were solvated in an octahedral box of three-site transferable intermolecular potential (TIP3P) water molecules that extended at least 12 Å from the protein and were neutralized with counter ions [123]. The Amber (version 11) program was used to minimize and equilibrate the systems. Periodic boundary conditions with the Particle Mesh Ewald (PME) model for electrostatics were used. The energy minimization was performed in 14 stages, each with a harmonic restraint (decreasing from 1000 to 0 kcal/mol·Å^2) to the crystallographic positions on the non-hydrogen atoms. For each restraint value, 1000 minimization cycles were performed consisting of 100 cycles of steepest descent and 900 cycles of conjugate gradient. Heating to 300 K was then performed for 25 ps under constant pressure, constant volume (NVT) conditions with a restraint of 50 kcal/mol·Å^2 on all non-hydrogen atoms. A timestep of 1 fs was used. Langevin dynamics with a collision frequency $\gamma = 1$/ps was performed. After heating, the system was further equilibrated in eight stages each with a decreasing restraint (50 to 0 kcal/mol·Å^2) on non-hydrogen atoms. For each restraint force constant, 25 ps of equilibration was performed. The temperature was maintained at 300 K by Langevin dynamics with a collision frequency $\gamma = 1$/ps, and the pressure was maintained at 1.0 bar by isotropic position scaling with a relaxation time of 1 ps. The same setup

was applied to a final equilibration under constant pressure, constant temperature (NPT) conditions with no positional restraints for 1 ns.

The production runs for the ligand-free *T. brucei* CYP51, MCP-bound *T. brucei* CYP51 and ligand-free human CYP51 models were performed with NAMD2.9 [55]. For the MCP-bound *T. brucei* CYP51 simulation, the same force field as the simulations of the ligand-free *T. brucei* CYP51 was used together with the same minimization and equilibration protocol. The switching function was set to "on", and the non-bonded cutoff was set to 10 Å. All the bonds to hydrogen atoms were constrained with the SHAKE algorithm. Temperature was controlled by Langevin dynamics with a damping coefficient of 0.5/ps at 300 K. Pressure was controlled by the Nosé-Hoover Langevin piston method with an oscillation period of 1 ps and a damping time scale of 1 ps at 1.011325 bar. The production runs were 30 ns for the two models of ligand-free and MCP-bound *T. brucei* CYP51 and 15 ns for the model of ligand-free human CYP51 with a timestep of 1.5 fs.

3.2.3 MD Simulations of the Models of Soluble CYP2C9 and CYP2E1

The MD simulation trajectories of the models of soluble CYP2C9 and CYP2E1 at 300 K from previous studies were analyzed (Table 3.1; and unpublished data by Outi M. H. Salo-Ahen). The simulations of human CYP2C9 were performed using two models built from the crystal structure (PDB ID: 1R9O) with different conformations of the F-G loop region (one model with F' and G' helices and the other model with an F-G loop only) [29]. The simulations of human CYP2E1 were performed using two CYP2E1 models based on two crystal structures (PDB ID: 3KOH, 3T3Z; and unpublished data by Outi M. H. Salo-Ahen). The trajectories were compared with those of the soluble CYP51s.

3.2.4 Coarse-grained Model of Membrane-bound *T. brucei* CYP51

A model of the membrane-bound ligand-free *T. brucei* CYP51 was built using an improved version of the protocol described in ref. [29]. MODELLER_9.10 [127] was used to model the missing residues. The missing residues 1 to 21 were predicted to form the transmembrane helix and modeled as a helix. Residues 22 to 34 were modeled as a loop (residues 22-28 were missing in the crystal structure; more residues were modeled for the superimposed of the globular do-

main). The modeled transmembrane helix and loop structure were then coarse-grained using MARTINI2.1 with an elastic network between the non-bonded beads [128, 67, 68, 66]. The elastic bond force constant was 1.19 kcal/mol·Å^2 (500 kJ/mol·nm^2), and the upper bond length cut off was 9 Å. The coarse-grained structure (modeled residues 1 to 34) was inserted into a pre-equilibrated bilayer of 602 coarse-grained POPC molecules [128]. The overlapping lipids within 3 Å of the transmembrane helix were removed. A 5 μs simulation was performed to find the preferred orientation of the transmembrane helix alone in the membrane. The

CYP	PDB	Environment[b]	Ligand[c]	Ensemble[d]	Time (ns)
CYP51 (*T. brucei*)	3G1Q	SOL	—-	NPT	30
	3G1Q	MEM	—-	NP$_\gamma$T	60
	3P99[e]	SOL	MCP	NPT	30
	3P99[e]	MEM	MCP	NP$_\gamma$T	15
CYP51 (human)	3LD6	SOL	—-	NPT	15
CYP2C9_M1 (human)[a]	1R9O	SOL	—-	NPT	28.28
	1R9O	MEM	—	NP$_\gamma$T	18.75
	1R9O	SOL	FLO	NPT	27.84
	1R9O	MEM	FLO	NP$_\gamma$T	11.25
CYP2C9_M2 (human)[a]	1R9O	SOL	—-	NPT	27.78
	1R9O	MEM	—-	NP$_\gamma$T	18.75
	1R9O	SOL	FLO	NPT	27.4
	1R9O	MEM	FLO	NP$_\gamma$T	11.25
CYP2E1 (human)	3KOH	SOL	—-	NPT	10
	3T3Z	SOL	—-	NPT	10

Table 3.1: Main features of the all-atom molecular dynamics production simulations of the different CYPs.
(a) The F-G loop region in the crystal structure of human CYP2C9 (PDB ID: 1R9O) is missing. Two models of the F-G loop were generated for the simulations [29]. Model 1 has F' and G' helices and model 2 has an unstructured F-G loop.
(b) SOL, the soluble CYPs (no membrane), MEM, the membrane-bound CYPs in a bilayer.
(c) The ligands used in the simulation. FLO, 4'-hydroxy flurbiprofen, MCP, 14α-methylenecyclopropyl-Δ7-24,25-dihydrolanosterol, —-, ligand-free structures.
(d) In the NP$_\gamma$T simulations of the membrane-bound CYPs, a surface tension, γ=60 dyn/cm was applied to the plane of the membrane together with the pressure control to maintain the bilayer dynamics [29]. A comparison of the area per lipid, the bilayer thickness and the order parameters of the membrane in the 60 ns simulation of the membrane-bound CYP51 with experimental data is plotted in Figures 3.1 & 3.2.
(e) Only the ligand MCP of 3P99 was transferred into the active site of the equilibrated ligand-free CYP51 structure (originally derived from the crystal structure 3G1Q). The simulations were then done starting with this structure (see Methods).

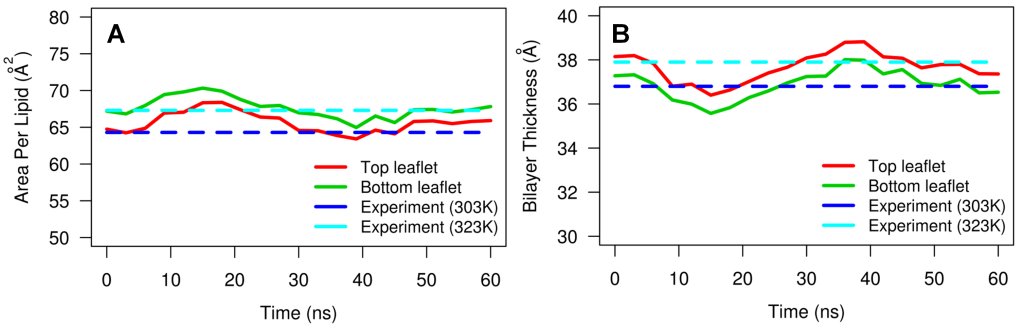

Figure 3.1: Comparison of the experimental and simulated area per lipid and thickness of the POPC membrane in the 60 ns simulation of membrane-bound *T. brucei* CYP51. The 60 ns trajectory was sampled at 3 ns intervals which resulted in 20 snapshots for the calculation. The comparison of (A) the computed the area per lipid and (B) the bilayer thickness with the experimental values at 303 K and 323 K (dotted lines) is shown.

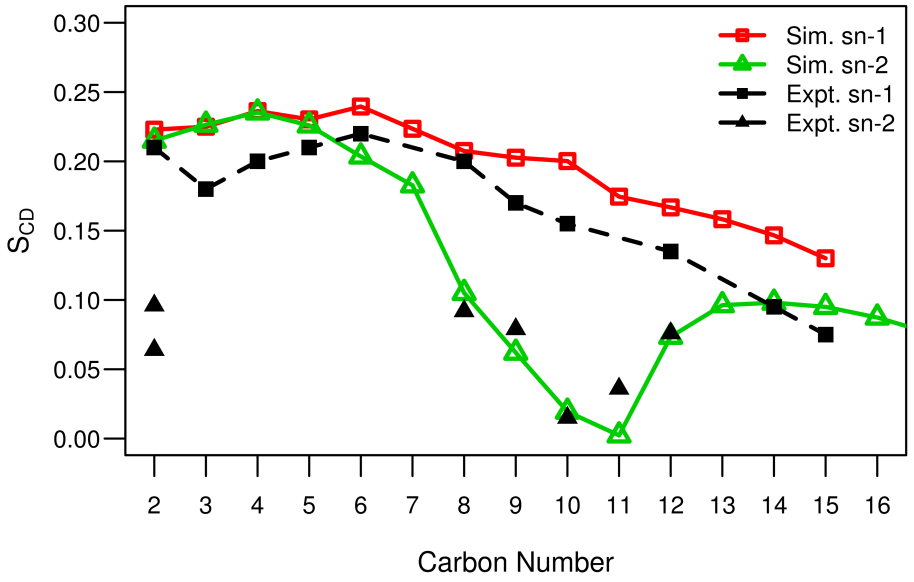

Figure 3.2: Comparison of the order parameters of the POPC membrane in the 60 ns simulation of membrane-bound *T. brucei* CYP51 with the experimental values. The black symbols show the experimental values and the red and green symbols the values from the simulations. "sn-1" is the saturated chain of POPC and "sn-2" is the unsaturated chain of POPC. The SCD values of a few carbon atoms of sn-2 were not observed in the experiments [124]. The SCD values of sn-1 are higher than the experimental values and this is consistent with reference [125]. As in [125], the C2 splitting was not reproduced in our simulations. The other SCD values of sn-1 are in agreement with experimental values except that the minimum was at C11 instead of C10 which has previously been seen for simulations with GAFF parameters [126].

protocol is the same as for the whole membrane-bound protein system (refer to succeeding text).

A coarse-grained model of the *T. brucei* CYP51 globular domain (PDB ID: 3G1Q) was also generated with MARTINI and combined with the preferred ori-

entation of the transmembrane helix in the 5 μs simulation. The elastic network of the linker region (residues 22-36) connecting the transmembrane helix and the globular domain was removed. Ten different initial orientations of the globular domain with different angles and height above the membrane were generated by changing the dihedral angles of this flexible loop region randomly as described in ref. [29].

Each complex was then solvated in a box of coarse-grained water molecules with the same x and y dimensions as the POPC bilayer taken from ref. [128] and extending 26 nm along the z axis. The system was minimized with the steepest descent method using GROMACS_4.5.5 [129, 130, 131, 132]. The system was minimized until the maximum force was less than 10 kJ/mol·nm (23.8 cal/mol·Å) for 10000 steps. Then the system was relaxed at 200 K for 50000 steps. After that, a 3 μs simulation with a timestep of 20 fs was performed. Periodic boundary conditions were applied. The electrostatic interactions and van der Waals interactions were calculated with the"Shif" algorithm. The lower and upper cutoffs for electrostatic interactions were 0 and 12 Å whereas those of the van der Waals interactions were 9 Å and 12 Å. Temperature coupling using velocity rescaling with a stochastic term was applied with a 1 ps time constant for coupling at 310 K. Semi-isotropic Berendsen pressure coupling was used with a 2 ps coupling time constant at 1 bar.

To check for possible artifacts from the definition of the elastic network, four more simulations of the coarse-grained *T. brucei* CYP51 in the POPC bilayer were performed after the removal of the elastic network for residues 22-23, 22-31, 22-33, 22-35. Despite the differences in the definition of the elastic network, three of these four simulations converged to the same orientation as the previous simulations except the simulation with the elastic network removed for residues 22-31, in which the protein did not embed in the membrane and the F-G loop region was out of the membrane. Although the extent to which the elastic network was different, the other three simulations converged to the same final orientation as the nine out of 10 previous simulations. To check whether the flexibility of the F-G loop influences the final orientation, one additional simulation was performed with the elastic network of the F-G loop (residues 22-36 and 207-224) removed. This simulation converged to the same final orientation as well.

3.2.5 Reverse Conversion from the Coarse-grained to the All-atom Model

The conversion from the coarse-grained model to the all-atom model was performed separately for the membrane, the transmembrane helix (residues 1-21), the linker region (residues 22-36) and the globular domain (residues 37-476). The transmembrane helix and the globular domain were converted by superimposing the modeled structure of the trans-membrane helix and the equilibrated all-atom models from the soluble ligand-free *T. brucei* CYP51 simulation, respectively, onto the coarse-grained models. The same protocol as in reference [29] was used to convert the lipids.

Because of the large conformational changes of the linker region, the MARTINI reverse conversion tool was used to convert the linker region [133]. Firstly, the MARTINI mapping information of the linker regions, which correlates the all-atom amino acids with the coarse-grained beads of the amino acids, was used to position the all-atom amino acids onto the coarse-grained beads after the coarse-grained simulations. Then a subsequent simulated annealing was performed for 60000 steps with a timestep of 2 fs. The linker was coupled at 1300 K at the beginning of the annealing; the temperature was decreased linearly to 300 K. The four parts (transmembrane helix, the linker, the globular domain, and the POPC bilayer) of the system were then combined, and the system was neutralized in a NaCl concentration of 150 mM and solvated in a box of TIP3P water molecules (same dimension as the membrane in the x and y directions and extending 16 nm in the z direction). For the MCP-bound model, MCP was superimposed onto the active site of the membrane-bound ligand-free *T. brucei* CYP51 by using the transformation matrix calculated by superimposing it with the crystal structure (PDB ID: 3P99).

3.2.6 All-atom MD Simulations of Membrane-bound *T. brucei* CYP51

The ff99SB force field was used for the protein and GAFF was used for the POPC bilayer [50, 53, 125]. Energy minimization of the all-atom system was performed in Amber (version 12) [54] with decreasing force constants from 1000 to 0 kcal/mol·Å^2. One hundred cycles of steepest descent and 900 cycles of conjugate gradient were used for each restraint value. The system was then equilibrated in NAMD2.9 as described previously [29]. For the production runs, periodic boundary conditions

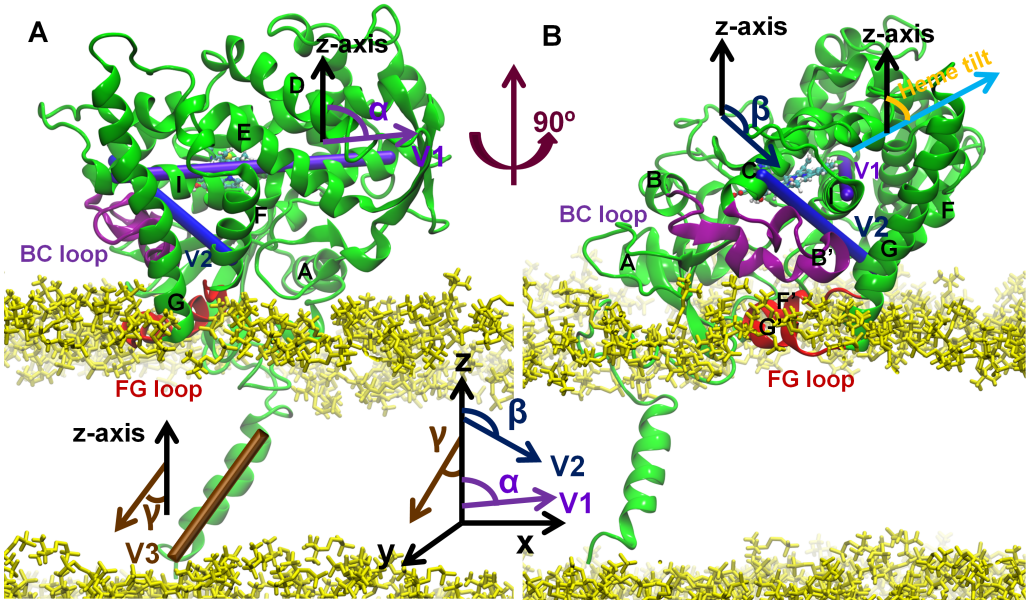

Figure 3.3: Two views of CYP51 in the bilayer at the last snapshot of the 60 ns simulation of ligand-free *T. brucei* CYP51. Important secondary structure elements are labeled. The B-C loop is shown in purple and the F-G loop in red. The head groups of the POPC lipids are shown in yellow and the membrane plane is the x-y plane. Two vectors (V1 and V2) that define the orientation of the protein with respect to the membrane are represented by cylinders. V1 follows the I helix. V2 points from the first helical turn of the C helix to the last helical turn of the F helix. Vector V3 points along the transmembrane helix. The "Heme tilt" (orange) is the angle between the heme plane (cyan arrow) and the normal to the membrane plane (black arrow). Detailed information of the three angles is shown next to the cylinder representation of the three vectors. For more details, refer to Materials and Methods: Data Analysis.

were used and the electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method. The NPT ensemble was applied with a target surface tension of 60 dyn/cm. All bonds to hydrogen atoms were constrained with the SHAKE algorithm. Temperature was controlled by Langevin dynamics with a damping coefficient of 0.5/ps at 310 K on non-hydrogen atoms. Pressure was controlled by the Nosé-Hoover Langevin piston method with an oscillation time of 1000 fs and a damping time of 1000 fs. A production run of 60 ns was performed for the ligand-free model without restraints with a timestep of 1.5 fs. A production run of 15 ns was performed for the MCP-bound model with the same setup.

3.2.7 Data Analysis

Three parameters [29] were used to define the position and orientation of the globular domain on the membrane (Figure 3.3): (i) d: the distance of CYP51 from the membrane was calculated by subtracting the z-coordinate of the center of mass

of the protein from the z-coordinate of the center of mass of the membrane. (ii) α: angle between the z-axis and vector V1; V1 is the vector defined by the center of mass of the first (residues 278-282) and last (residues 304-308) helical turns of the I helix. (iii) β: angle between the z-axis and vector V2; V2 is the vector defined by the center of mass of one helical turn of the C helix (residues 122-126) and one turn of the F helix (residues 202-206). For the coarse-grained simulations, the system was analyzed only after the system was equilibrated and the parameters fluctuated stably.

The angle γ between the z-axis and vector V3 was used to define the orientation of the transmembrane helix in the membrane. V3 is the vector defined by the center of mass of the first (residues 1-5) and last (residues 17-21) helical turns of the transmembrane helix. The heme tilt angle is the angle between the z-axis and the average heme plane defined by the "NA", "NB", "NC" and "ND" atoms of the heme. The "Height" of the proteins was calculated by taking the z-coordinate of the highest non-hydrogen atom of the protein above the membrane and subtracting the average z-coordinate of the phosphate groups (atoms: "P", "O1P", "O2P", "O3P" and "O4P") of the upper leaflet of the lipid bilayer.

All the previously mentioned parameters were calculated in VMD [118]. The RMSD and root-mean-square fluctuation (RMSF) calculations were performed using the cpptraj program of the AmberTools12 [54]. For the comparison of the computed B-factors, the four CYPs (human CYP51, *T. brucei* CYP51, human CYP2C9 and human CYP2E1) were aligned with the sequence and structure alignment tool PROMALS3D [134].

The area per lipid and the thickness of the membrane were analyzed with the grid-based tool GridMAT-MD [135]. Two hundred grids were assigned to the x-y plane and the x and y ratio was then modified according to the dimensions of the membrane of the CYP system. Lipids that were within 10 Å of the protein were removed for the calculation. The order parameters were calculated with the g_order tool of the GROMACS_4.6 package [136].

The opening and closing of the ligand tunnels was analyzed using CAVER3.0 [32]. PDB format coordinate files (100-200) were generated along each trajectory and were used as the input for the CAVER analysis. For example, in the membrane-bound ligand-free model of *T. brucei* CYP51, 200 PDB files were generated from the 60 ns simulation with a 0.3 ns sampling interval. The probe radius was assigned a value of 1.4 Å, corresponding to the size of a water molecule. A tunnel

was considered to be open when the radius of the bottleneck was larger than 1.4 Å.

3.3 Results and Discussion

3.3.1 Model of *T. brucei* CYP51 in a POPC Bilayer

3.3.1.1 One predominant orientation of the membrane-bound *T. brucei* CYP51

An improved version of the protocol in ref. [29] as described in the Methods section was used to build the membrane-bound ligand-free model of *T. brucei* CYP51. In the 5 μs coarse-grained simulation of the transmembrane helix alone in the POPC bilayer, the transmembrane helix adopted an orientation of 17 $\pm$ 8° to the normal of the membrane (angle γ in Figure 3.3). In the coarse-grained simulations of the whole protein in the POPC bilayer, nine out of the 10 simulations converged to one final orientation after the system was equilibrated (Figure 3.4). The three parameters defining the position of the protein after the equilibration of the nine converged coarse-grained simulations were: d = 41.3 $\pm$ 1.8 Å, α = 82.2 $\pm$ 6.3°, and β = 126.5 $\pm$ 8.4° (Figure 3.5). One snapshot from the coarse-grained simulations having values of all three parameters within 1% of the mean values of the parameters was taken as the starting structure for the all-atom simulations. In the all-atom simulations of the ligand-free and MCP-bound *T. brucei* CYP51, the globular domain stayed in about the same orientation above the membrane. The parameters were: d = 38.1 $\pm$ 0.9 Å, α = 82.8 $\pm$ 5.2°, and β = 122.9 $\pm$ 6.0° (Figure 3.5).

3.3.1.2 Comparison with experimental data and with other CYP models supports the atomic detail models of the membrane-bound *T. brucei* CYP51

In the simulation of the membrane-bound ligand-free *T. brucei* CYP51, the tilt angle of the transmembrane helix to the normal of the bilayer was 20.1 $\pm$ 6.5° which is similar to the transmembrane helix tilt angle of 17 ° measured for this helix from CYP2B4 by solid state NMR spectroscopy [108]. In the model of membrane-bound *T. brucei* CYP51, the F-G loop region (residues 206-224) is embedded in the membrane which is consistent with experiments on other CYPs [103, 105, 106]. Table

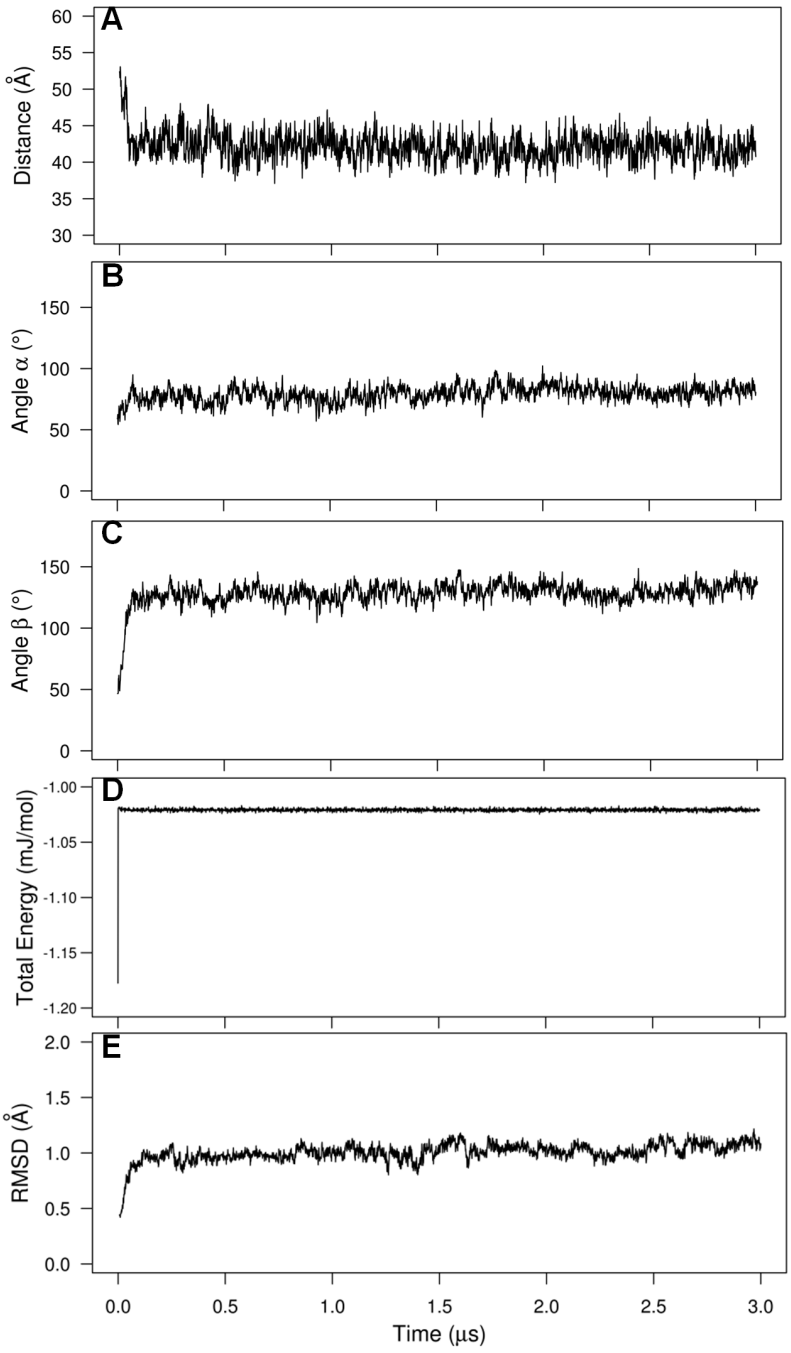

Figure 3.4: The fluctuations of the three measured parameters (d, α and β) during one of the ten coarse-grained simulation of *T. brucei* CYP51 in POPC bilayer. The system was considered to be equilibrated when all three parameters were fluctuating stably.

3.2 shows a comparison of the model of the membrane-bound *T. brucei* CYP51 with experimental data and the models of CYP2C9 [29] and CYP3A4 [111]. The CYP51 model is mostly in good agreement with these experiments and shows about the

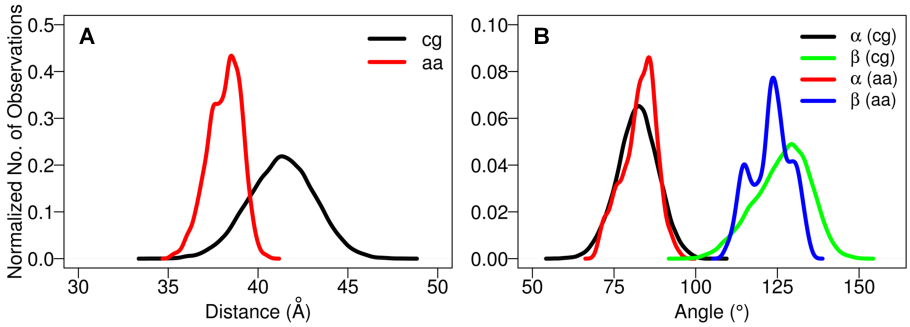

Figure 3.5: The normalized distribution of the three measured parameters (d, α and β) in the coarse-grained and all-atom simulations of CYP51. (A) The distance between the center of mass of the membrane and the protein. The black line shows the distribution of the coarse-grained simulation and the red line, the all-atom simulation. (B) The angle α between the vector V1 and the z-axis and the angle β between the vector V2 and the z-axis. The black and red lines show angles α and β in the coarse-grained simulation, respectively. The green and blue lines are angle α and β in the all-atom simulation respectively. The distribution of these three parameters is similar to those of the membrane-bound model of CYP2C9 with an unstructured F-G loop (CYP2C9_2) [29].

same degree of agreement with experiments as the CYP2C9 and CYP3A4 models. The measured heme tilt angle (between the heme plane and the normal to membrane plane, as shown in Figure 3.3B) of CYP3A4 was 59.7 $\pm$ 4.1° in experiments [111]. This angle was 63.1 $\pm$ 4.5° in our simulation of the ligand-free membrane-bound *T. brucei* CYP51, this is close to the experimental value (Table 3.2) and that in models of six human drug-metabolizing CYPs (CYP1A2, 67 $\pm$ 6°; CYP2A6, 69 $\pm$ 5°; CYP2C9:,61 $\pm$ 4°; CYP2D6, 72 $\pm$ 6°; CYP2E1, 60 $\pm$ 5°; and CYP3A4, 56 $\pm$ 5°) [112]. The membrane-bound model of *T. brucei* CYP51 also agrees quite well with the predicted orientation (d, 41.6 Å; α, 83.7°; β, 131.8°) in the Orientations of Proteins in Membranes (OPM) database for CYP51 although OPM predicted a different orientation for CYP2C9 with the A helix region (residues: 45-66) not embedded in the membrane [29, 137]. After we built our models of *T. brucei* CYP51, the structure of the full-length *Saccharomyces cerevisiae* CYP51 was determined by crystallography [109]. The orientation of our model in lipid bilayer is consistent with this structure of a fungal CYP51. The computed B-factors of the membrane-bound CYP51 in the simulation were compared with those of the crystal structure and the simulated soluble form (Figure 3.6). The trends are consistent for the three systems although, as expected, the computed B-factors vary over a wider range in the simulations. The F-G loop region has lower mobility in the membrane-bound simulation because this region is embedded in the membrane.

Residue Number[i]	Exp. Data[*]	Exp. CYP	CYP51		CYP2C9_M1		CYP2C9_M2		CYP3A4	
			Loc.[ii]	Agree[iii]	Loc.	Agree	Loc.	Agree	Loc.	Agree
1-27	M[a]	2B2	M	+	M	+	M	+	M	+
28	M[a] / C[b]	2B2, 2B4	C	+	C	+	C	+	M	+
29-40	M[a]	2B2	M	+	M & C	+-	M	+	M & C	+-
42-49	M[a]	2B2	M & C	+-	M	+	M & C	+-	M & C	+-
63-75	C[a]	2B2	C	+	M	-	M	-	M & C	+-
95-100	M[a]	2B2	C	-	C	-	C	-	C	-
110-118	C[a]	2B2	C	+	C	+	C	+	C	+
124-133	C[a]	2B2	C	+	C	+	C	+	C	+
186-193	C[a]	2B2	C	+	C	+	C	+	C	+
209-218	M[a]	2B2	M & C	+-	M	+	M	+	M	+
220-227	C[a]	2B2	M	-	M & C	+-	C	+	M & C	+-
310-318	C[a]	2B2	C	+	C	+	C	+	C	+
390-406	C[a]	2B2	C	+	C	+	C	+	C	+
39,72	M[c]	21A1	M	+	M	+	M	+	M	+
221	C[c]	21A1	M	-	M	+-	C	+	M	-
83, 123,190, 235, 341	C[c]	21A1	C	+	C	+	C	+	C	+
373[iv]	M[c]	21A1	C	-	M	+	M	+-	C	-
Heme Tilt Angle (○)[v]	59.7 ± 4.1[d]	3A4	63.1 ± 4.5	+	43.2 ± 3.4	-	38.5 ± 5.4	-	61.5	+
	38-78[c]	17A1, 21A1		+		+		+		+
Height_1 (Å)[vi]	35 ± 9[e]	2B4	50.5 ± 2.2	-	47.9 ± 1.7	-	47.1 ± 1.9	-	58.0	-
Height_2 (Å)[vii]			38.1 ± 0.9	+	33.6 ± 0.7	+	41.0 ± 1.0	+	39.6	+

Table 3.2: Comparison of membrane-bound *T. brucei* CYP51 model with the experiments, two human CYP2C9 models with different F-G loop structures [29] and a CYP3A4 model [111]. (i) The residue numbers are adjusted from the CYP2C9 residue numbers according to the alignment of CYP51 with CYP2C9 (Figure S2) and the alignment of CYP51 with CYP3A4 using the "align" function of PyMOL [116]. (ii) The location of the residues with respect to the membrane: "M" indicates that the residues are either in the membrane or in the head group region of the membrane. "C" indicates that the residues are in the cytosol. "M & C" indicates that the fragment is partly in the membrane and partly in the cytosol. "M / C" indicates that there are experiments showing a particular residue is in the membrane while other experiments showed that this residue is in the cytosol. (iii) These columns show whether the model is in agreement with the experiments. "+": the model is in good agreement with the experiments. "+-": the model is partly in agreement with the experiments. "-": the model is not in agreement with the experiments. (iv) Residue 373 was shown to be deeper in the membrane than the residues 39 and 72. (v) "Heme Tilt Angle" is the angle between the heme plane and the normal of the membrane plane. For CYP51 and the two models of CYP2C9, the heme tile angle is computed as an average over snapshots from the production runs of the simulations. For CYP3A4, the angle is computed for the final model. (vi) "Height_1" is the height of the CYP above the lipid bilayer measured by atomic force microscopy in the experiments and the distance between the z-coordinate of the highest non-hydrogen atom of the protein and the average z-coordinate of the phosphate group of the upper leaflet of the lipid bilayer. (vii) "Height_2" is the distance between the center of mass of the protein and the center of mass of the membrane. (*) References are (a): [102], (b): [107], (c): [104], (d): [111] and (e): [101] and experiments were carried out for the CYP listed in the third column.

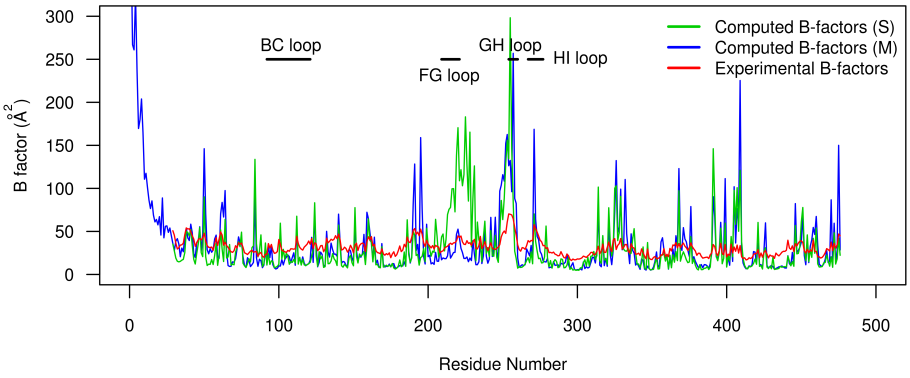

Figure 3.6: Comparison of B-factors of the ligand-free forms of *T. brucei* CYP51 in the crystal structure, the soluble simulation and the membrane-bound simulation. The B-factors of the soluble form (S) were computed for the last 9 ns and those of the membrane-bound form (M) were computed for the last 15 ns of the simulations of the ligand-free models. The F-G loop is embedded in the lipid bilayer thus this region is more rigid in the simulation of membrane-bound *T. brucei* CYP51 than that of the soluble form. This is because the motions of the residues are restricted by the lipid bilayer.

3.3.2 Dynamics of CYP51 and Implications

3.3.2.1 Lower mobility in the active site of CYP51 but not higher mobility at the ligand access tunnel entrance

Comparison of the crystal structures suggested that CYP51 is less mobile than other CYPs [7, 96]. Therefore, we compared the computed B-factors in the ligand-free simulations of the four CYPs studied.

The simulations of soluble models of CYPs are converged shortly after equilibration, for example, in the simulation of the soluble model of ligand-free *T. brucei* CYP51 (Figure 3.9). In the simulations of the soluble CYPs, the average computed B-factor of *T. brucei* CYP51 (30.0 $\pm$ 32.4 Å^2), averaged over all non-hydrogen atoms and the last 9 ns of the simulations, and of human CYP51 (24.1 $\pm$ 22.7 Å^2) is similar to that of CYP2C9 (29.4 $\pm$ 39.8 Å^2 and 31.5 $\pm$ 39.4 Å^2) and of CYP2E1 (28.8 $\pm$ 33.2 Å^2 and 27.9 $\pm$ 39.0 Å^2) (Figure 3.7A). In the membrane-bound simulations, the average B-factor (over the last 12 ns) of the globular domain of *T. brucei* CYP51 (35.7 $\pm$ 33.2 Å^2) is also similar to that of CYP2C9 (32.1 $\pm$ 33.2 Å^2 and 32.5 $\pm$ 33.2 Å^2) (Figure 3.7B).

As the computed B-factors vary widely along the sequence, we next compared the B-factors of two groups of residues, the substrate binding cavity residues and the tunnel 2f entrance residues [100]. Residues of these two groups on *T. brucei* CYP51 are shown in Figure 3.8 and the corresponding residues of human CYP2C9

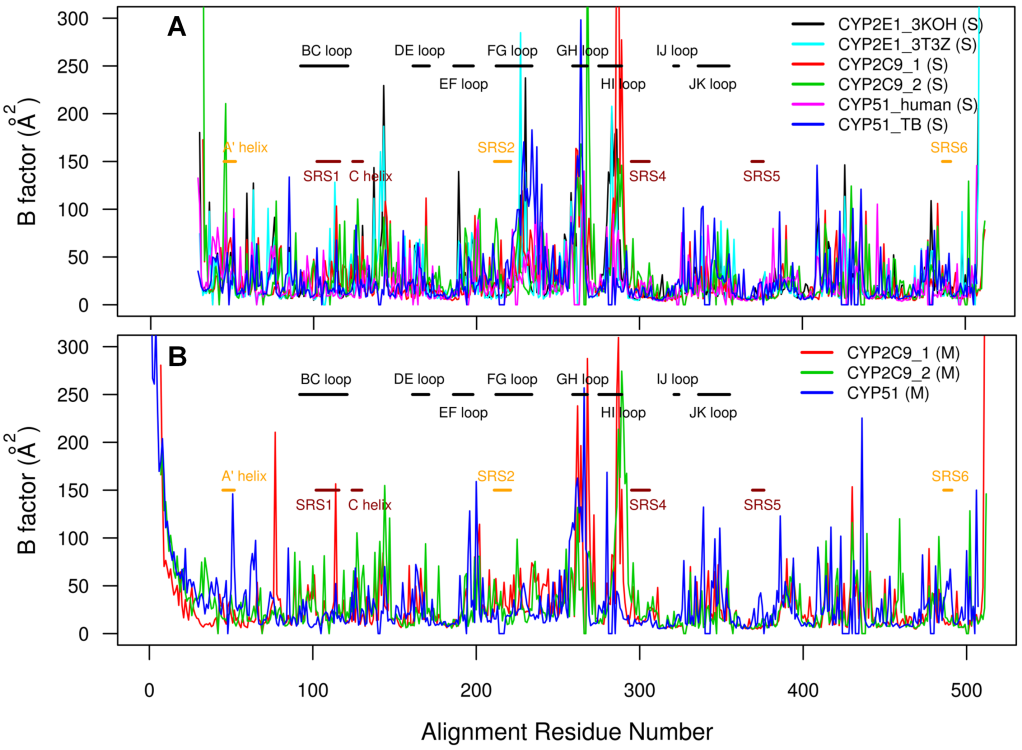

Figure 3.7: Comparison of the B-factors of the non-hydrogen atoms of different CYPs with the loop regions (black), the ligand binding cavity residues (dark red) and ligand tunnel opening residues (orange) labeled. The residue numbers of the CYPs correspond to the PROMALS3D alignment (Figure 3.5). (A) Comparison of B-factors averaged over the last 9 ns of the two simulations with soluble ligand-free forms of the two human CYP2C9 models, the two simulations with the two human CYP2E1 models and the simulations of T. brucei CYP51 and human CYP51. (B) Comparison of the B-factors averaged over the last 12 ns of the simulations with the membrane-bound ligand-free forms of CYP2C9 and CYP51.

and human CYP2E1 were taken from the alignment (Figure 3.10). The substrate binding cavity consists of the substrate recognition site 1 (SRS1, αB'-B'/C loop, residues 102-116), the C helix (residues 124-130), SRS4 (residues 284-295) and SRS5 (residues 355-361) [100, 138]. The ligand access tunnel 2f has entrance residues including αA' (residues 45-52), SRS2 (αF', residues 208-214) and SRS6 (β4 hairpin, residues 457-462). The average B-factors of the Cα atoms of these regions were computed and are given in Table 3.3. If the average of the computed B-factors of the Cα atoms was relatively low, the region was considered to have low mobility.

In the simulations of the soluble proteins, the average computed B-factor of the Cα atoms of the binding cavity residues of *T. brucei* CYP51 (9.1 $\pm$ 2.8 Å^2) and human CYP51 (8.1 $\pm$ 3.0 Å^2) is similar to that in the simulations of the two human CYP2E1 models (8.0 $\pm$ 3.2 Å^2 and 9.3 $\pm$ 5.0 Å^2) but lower than that in

CYP	Environment[a]	SRS1	C helix	SRS4	SRS5	A' helix	SRS2	SRS6
T. brucei CYP51	S	11.9 ± 2.0	8.7 ± 1.4	7.1 ± 1.1	6.8 ± 2.2	18.1 ± 4.9	39.9 ± 18.6	18.2 ± 9.2
human CYP51	S	8.7 ± 2.6	7.4 ± 1.5	9.1 ± 4.0	5.8 ± 1.3	40.7 ± 11.9	19.1 ± 8.0	10.4 ± 3.6
human CYP2C9_1	S	23.2 ± 8.6	10.1 ± 2.7	5.6 ± 1.2	5.9 ± 1.1	19.6 ± 13.0	17.0 ± 3.9	11.3 ± 4.8
human CYP2C9_2	S	19.8 ± 9.3	13.7 ± 4.8	13.8 ± 7.9	4.8 ± 0.9	11.2 ± 5.7	20.4 ± 15.1	6.3 ± 0.7
human CYP2E1_1	S	8.7 ± 2.1	9.7 ± 2.5	5.8 ± 2.2	8.5 ± 5.0	20.7 ± 7.2	9.0 ± 2.1	8.0 ± 1.4
human CYP2E1_2	S	12.5 ± 5.5	11.3 ± 2.3	5.3 ± 1.9	7.8 ± 4.5	12.5 ± 7.3	8.3 ± 3.5	9.0 ± 3.5
T. brucei CYP51	M	11.5 ± 2.6	8.3 ± 0.8	8.8 ± 3.7	9.8 ± 6.2	37.5 ± 14.3	16.9 ± 2.1	32.7 ± 15.1
human CYP2C9_1	M	15.8 ± 5.0	7.4 ± 0.8	13.0 ± 6.9	6.1 ± 1.2	10.4 ± 4.6	20.0 ± 5.3	7.5 ± 2.0
human CYP2C9_2	M	16.4 ± 9.1	16.8 ± 8.6	15.1 ± 8.2	5.5 ± 0.7	10.5 ± 2.4	18.8 ± 15.9	7.6 ± 2.2

Table 3.3: Average computed B-factors (Å^2) of Cα atoms in specific regions of the simulated CYPs. The B-factors for the soluble simulations were computed from the last 9 ns of the two simulations with soluble ligand-free forms of the two human CYP2C9 models, the two simulations with the two human CYP2E1 models, and the simulations of *T. brucei* CYP51 and human CYP51. The B-factors of the membrane-bound simulations were computed from the last 12 ns of the simulations with the membrane-bound ligand-free forms of CYP2C9 and CYP51. (a) Simulations were performed without a lipid bilayer (S, soluble model) or with a lipid bilayer (M, membrane-bound model).

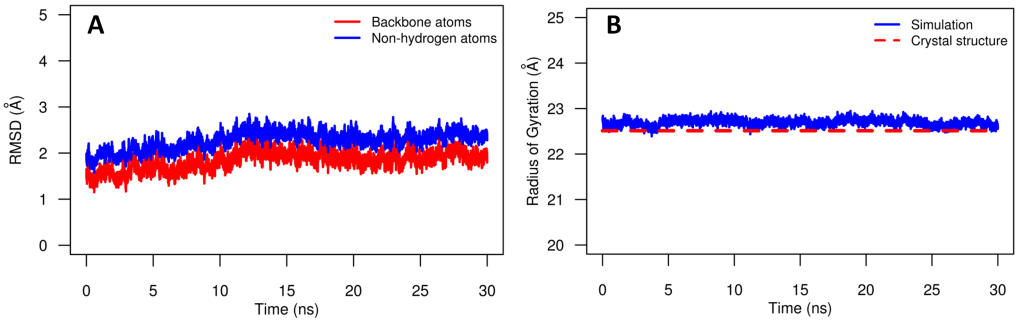

Figure 3.9: Plots of the RMSD and radius of gyration during the 30 ns simulation of the soluble ligand-free *T. brucei* CYP51. (A) RMSD plot of backbone atoms (red) and non-hydrogen atoms (blue). (B) Radius of gyration plot of non-hydrogen atoms. The radius of gyration of the crystal structure is shown by a dotted line.

the simulations of the two human CYP2C9 models (12.7 $\pm$ 9.6 Å^2 and 14.2 $\pm$ 8.9 Å^2) as shown in Figure 3.7A. In the membrane-bound simulations, the substrate binding cavity also has lower mobility in the simulation of the membrane-bound model of *T. brucei* CYP51 (9.8 $\pm$ 3.7 Å^2) than in those of human CYP2C9 (11.8 $\pm$ 6.1 Å^2 and 14.0 $\pm$ 8.6 Å^2). In both the soluble and membrane-bound simulations of CYP51, the binding cavity residues have low mobility (average B-factor less than 10 Å^2). The mobility of SRS1 of CYP51 is lower than that of CYP2C9 in both soluble and membrane-bound simulations and one CYP2E1 simulation but similar to that in the other CYP2E1 simulation. The C helix region of the binding cavity has lower mobility in CYP51 than CYP2C9 and CYP2E1 in the soluble simulations. In the membrane-bound simulations, the C helix of *T. brucei* CYP51 is less mobile than the model 2 of CYP2C9. The C helix of model 1 of CYP2C9 has lower mobility in this region because this region is partly embedded in the membrane. The SRS4 and SRS5 regions of all the CYPs studied have low mobility (average B-factor mostly less than 10 Å^2) in all the simulations.

The average computed B-factor of the 2f tunnel entrance residues of *T. brucei* CYP51 (25.4 $\pm$ 15.6 Å^2) and of human CYP51 (23.1 $\pm$ 15.1 Å^2) is higher than that of human CYP2E1 (12.8 $\pm$ 7.3 Å^2 and 9.9 $\pm$ 5.3 Å^2) and of human CYP2C9 (16.9 $\pm$ 9.3 Å^2 and 13.7 $\pm$ 11.2 Å^2) (Figure 3.7A). That of CYP51s is also higher than that of the 2a tunnel entrance residues of human CYP2E1 (18.9 $\pm$ 14.2 Å^2 and 20.1 $\pm$ 16.1 Å^2) and of human CYP2C9 (18.7 $\pm$ 8.1 Å^2 and 19.8 $\pm$ 12.2 Å^2). In the membrane-bound simulations, the 2f tunnel entrance residues (αA', SRS2 and SRS6) also have higher mobility in *T. brucei* CYP51 (29.3 $\pm$ 14.6 Å^2) than the 2a tunnel entrance residues (β1, B-B' loop and G' helix) in human CYP2C9 (17.8 $\pm$

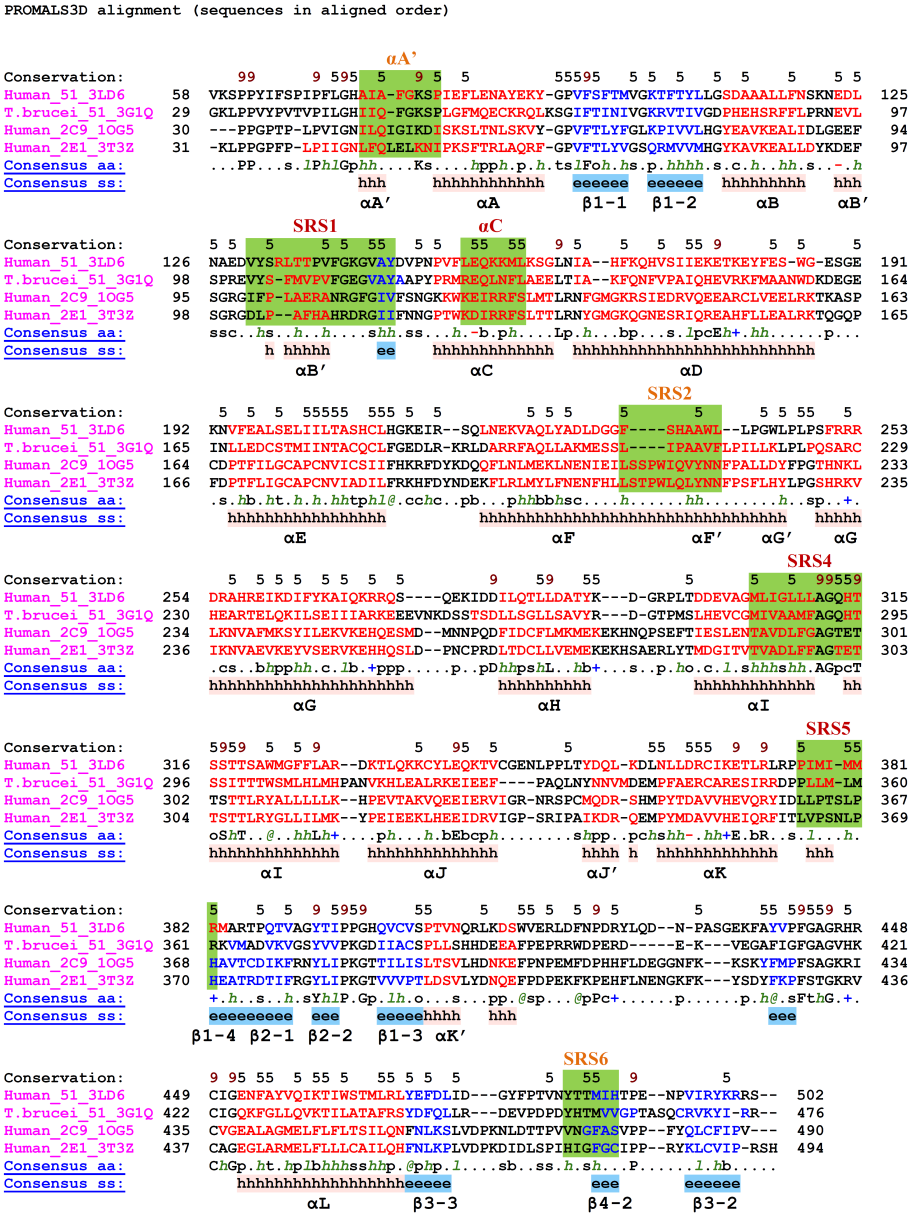

Figure 3.10: The PROMALS3D crystal structure-based alignment of the sequences of CYP51, CYP2C9 and CYP2E1 [134]. The binding cavity residues and tunnel 2f entrance residues are shown in green colored blocks. The aligned helices are in red and the aligned beta sheets in blue. The score of the conserved amino acids is shown in the "conservation" row. The well-conserved residues are labeled in the "consensus_aa" row. Conserved residues are uppercase letters. Consensus amino acid symbols are: "l": aliphatic (I, V, L); "@": aromatic (Y, H, W, F); "h": hydrophobic (W, F, Y, M, L, I, V, A, C, T, H); "o": alcohol (S, T); "p": polar residues (D, E, H, K, N, Q, R, S, T); "t": tiny (A, G, C, S); "s": small (A, G, C, S, V, N, D, T, P); "b": bulky residues (E, F, I, K, L, M, Q, R, W, Y); "+": positively charged (K, R, H); "-": negatively charged (D, E); "c": charged (D, E, K, R, H). The conserved secondary structures (consensus_ss) are represented by red "h" for α-helices and blue "s" for β-strands.

8.0 Å^2 and 17.3 $\pm$ 14.1 Å^2). The A' helix and SRS6 of *T. brucei* CYP51 are more mobile than those of CYP2C9 and of CYP2E1 (Table 3.3).

The differences between the computed B-factors of the CYPs studied are small and the overall mobility of these CYPs does not differ a lot. However, the binding cavity residues of CYP51s are in general less mobile than those of CYP2C9 and CYP2E1, and the tunnel entrance residues of CYP51s more mobile than those of CYP2C9 and CYP2E1.

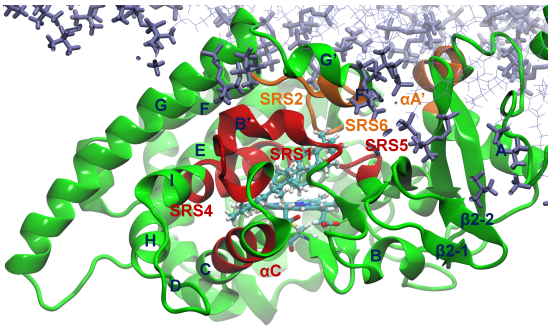

Figure 3.8: The substrate binding cavity residues (dark red) and the ligand tunnel entrance residues (orange) shown in the last snapshot of the simulation of the membrane-bound ligand-free T. brucei CYP51. The substrate binding cavity residues are residues 102-116 (SRS1, αB'-B'/C loop), 124-130 (αC), 284-295 (SRS4, αI) and 355-361 (SRS5). The residues at the entrance of ligand tunnel 2f include αA' (residues 45-52), SRS2 (αF', residues 208-214) and SRS6 (β4 hairpin, residues 457-462). The ligand MCP is shown in cyan. The bilayer is shown in violet.

3.3.3 Ligand Tunnels in CYP51

Figure 3.11 shows the tunnels between the active site and the protein surface [29]. Tunnel 2f (between helices A', F" and the tip of the β4 hairpin), leading to the membrane, and tunnels 2b (between the B-B' loop and the β1-2 and β1-4 sheets), S (the solvent tunnel, between F, I helices and the β4 hairpin) and W (the water tunnel, helix C and the β-bulge segment), leading to the solvent, were observed to open in simulations of membrane-bound and soluble forms of the *T. brucei* CYP51. Tunnel 2a (between the F-G loop, B-B' loop and β1-1 sheet), pointing to the bilayer, and tunnels 2c (between the G and I helices, B'-C loop), 2ac (between the B' helix and G helix) and 2e (through the B-C loop) were closed throughout the simulations of the membrane-bound form of *T. brucei* CYP51. Some opening of tunnels 2e and 2ac was seen in the simulations of the soluble form of *T. brucei* CYP51 (Figure 3.12).

3.3.3.1 Protein flexibility influences the opening and closing of ligand tunnels

In the simulations of *T. brucei* CYP51, dynamic opening and closing of the ligand tunnels were observed. The 2f tunnel is open in the crystal structures and was mostly open in the simulations (for 10-97% of the snapshots analyzed). The S

tunnel is closed in the crystal structure (PDB id: 3G1Q) but was found to be open in the simulations (for 2-73% of the snapshots analyzed). These two tunnels were very often open in the simulations of the ligand-free soluble form and the ligand-free membrane-bound forms. However, although these two tunnels were open in the simulation of the MCP-bound membrane-bound form of *T. brucei* CYP51, they were mostly closed in that of the MCP-bound soluble form. This indicates that the membrane stabilized the opening of tunnels 2f and S when the ligand was bound in the active site. The common entrance residues of tunnels 2f and S (SRS2 and SRS6) are mobile in the MD simulations (discussed previously), and thus, these two tunnels were able to open and close in the simulations. The B-C loop and C helix modulate the opening and closing of most family 2 tunnels, including 2a, 2b, 2c, 2e and 2ac. These tunnels are closed in the crystal structures and because of the low mobility of the B-C loop and C helix region (discussed previously), these tunnels were mostly closed in the simulations (Figure 3.12). However, tunnel W,

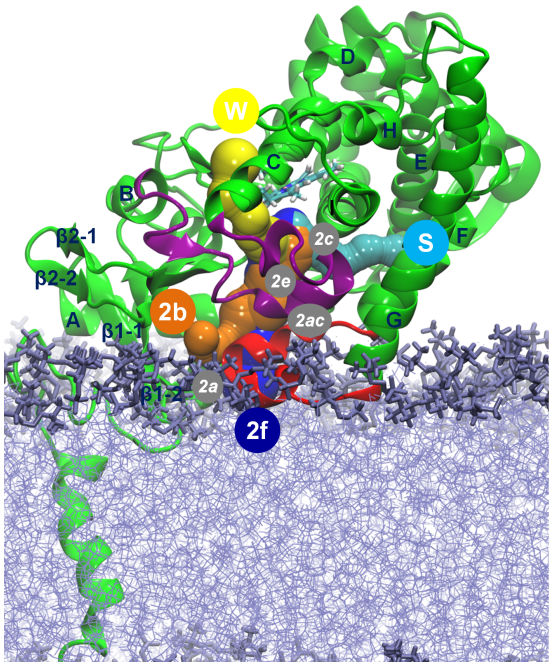

Figure 3.11: Cartoon representation of *T. brucei* CYP51 showing the ligand tunnels of the membrane-bound CYP51 analyzed in CAVER3.0 [32]. The important secondary structure elements are labeled. The B-C loop is shown in purple and the F-G loop in red. The bilayer is shown in violet. Tunnels that opened at least in one snapshot during the 60 ns simulation of the membrane-bound ligand-free CYP51 are shown by sphere representations and labeled by circles with white letters. Tunnel 2f (blue) leads to the membrane and tunnels 2b (orange), S (cyan), and W (yellow) lead to the solvent. Tunnels 2a, 2c, 2e and 2ac that were not open in the 60 ns simulation are labeled by white italic letters in gray circles.

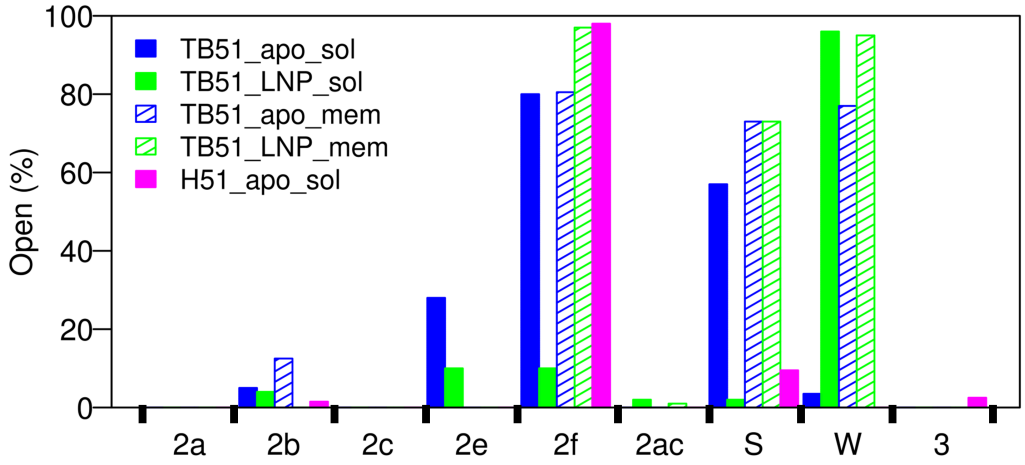

Figure 3.12: Percentage of the snapshots in which the ligand tunnels were open in the simulations is shown for the ligand-free soluble *T. brucei* CYP51 (solid blue), MCP-bound soluble *T. brucei* CYP51 (solid green), ligand-free membrane-bound *T. brucei* CYP51 (hatched blue), MCP-bound membrane-bound *T. brucei* CYP51 (hatched green) and ligand-free soluble human CYP51 (solid magenta). In CYP51s, the tunnels are located as follows: 2a: between the F-G loop, B-B' loop and β1-1 sheet, b: between the B-B' loop and the β1-2 and β1-4 sheets, 2c: between the G and I helices, B'-C loop, 2ac: between the B' helix and G helix, 2e: through the B-C loop, 2f: between helices A', F" and the tip of the β4 hairpin, 3: between the F and G helices, S: between F, I helices and the β4 hairpin and W: helix C and the β-bulge segment.

which is open in the crystal structures (Table 3.4), stayed open throughout most of the simulations.

3.3.3.2 Tunnels 2f and S can serve as major ligand tunnels in CYP51

The open tunnels of CYP51 are different from those of the human drug-metabolizing CYPs. In the simulations of CYP2C9, all the common tunnels (2a, 2b, 2c, 2ac, 2e, and S) were mostly open in a least one of the eight simulations of the two models of the soluble and membrane-bound CYP2C9 models. The 2f tunnel was not open in most of the CYP2C9 simulations [29]. The ligand tunnel analysis of the simulations of the membrane-bound models of six human drug metabolizing CYPs (CYP1A2, CYP2A6, CYP2C9, CYP2E1 and CYP3A4) shows that tunnel S was open in all six simulated CYPs, tunnels 2b and 2c were open in five of the simulated CYPs, tunnel W in four, and tunnels 2e and 2f in three [112]. However, tunnels 2f, S and W were found to be wide open in all our simulations of *T. brucei* CYP51 and the other tunnels were mostly closed. In the simulation of human CYP51, only tunnels 2f and S were wide open.

The 2f tunnel has been proposed to serve as the substrate access channel for CYP51 [7, 94]. Our simulations support the idea that tunnel 2f can serve as the

ligand entrance tunnel for hydrophobic ligands, such as the CYP51 substrates, to enter the buried active site from the membrane, where the sterol biosynthetic pathway takes place. The ligands can exit by either tunnel 2f to the membrane or tunnel S to the solvent, depending on the properties of the ligands.

3.3.4 Open Water Tunnel and Multiple Conformations of Heme Propionate Groups

3.3.4.1 B-C loop rigidity keeps the W tunnel open

The W tunnel is directed toward the proximal surface of CYP51, and the residues lining this tunnel are mostly in the proximal region of the B-C loop and the B and C helices. This tunnel cannot serve as a substrate/product tunnel because it is on the proximal (opposite from the substrate binding cavity) side of CYP51, which is the expected binding face of CYP51 for the CYP reductase. The W tunnel was found open only in the simulations of the membrane-bound models of six human drug-metabolizing CYPs but was closed in the simulations of the soluble models [112]. The open tunnel W in the membrane-bound form of CYP51 suggests that it is possible that tunnel W has to be open to facilitate the interaction with the reductase. This is different from the crystal structures of most human CYPs that have this tunnel closed. In the human drug-metabolizing CYPs, the opening and closing of this tunnel can be observed because the proximal side is flexible. However, in CYP51, because of the rigidity of the C helix and β-bulge segment (Figure 3.7), the residues lining the W tunnel are not able to facilitate the opening and closing motions of this tunnel. As a result, this tunnel stays open in both the crystal structures (Table 3.4) and simulations (Figure 3.12).

3.3.4.2 The two heme propionate groups adopt different conformations

Conformational changes of the A-propionate group have been observed in experimental and computational studies of mutants of the prokaryotic cytochrome P450 BM3 [139, 140]. In these studies, the heme A-propionate group was shown to adopt different conformations upon ligand binding. However, conformational changes of the two propionate groups have not been reported in MD simulations of wild-type CYPs. Here, we discuss the different conformations of the two propionate groups of CYP51 in crystal structures and in our simulations of the wild-type *T. brucei* CYP51.

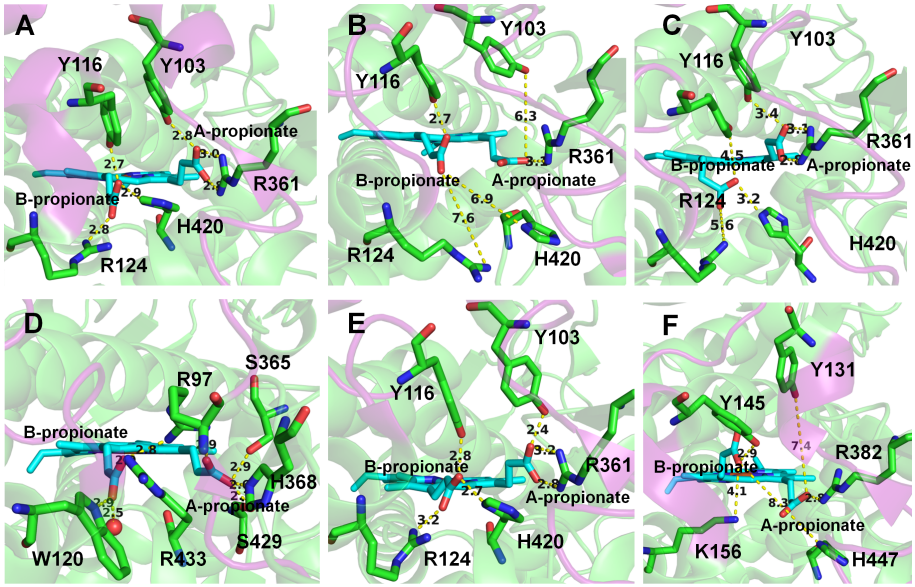

Figure 3.13: Different conformations and hydrogen bonding interactions of the two heme propionate groups in the crystal structures and simulations. The distance between the two heme propionate groups and important hydrogen bonding partners is shown. (A) The most predominant conformation of the propionate groups observed in the crystal structures of CYP51 enzymes. The coordinates are from *T. brucei* CYP51 (PDB id: 3G1Q). (B, C) Two alternative conformations of the propionate groups seen in the snapshots at 60 ns and 15 ns in the simulations of the membrane-bound ligand-free and MCP-bound *T. brucei* CYP51, respectively. (D) The conformation in the crystal structure and simulations of CYP2C9 (PDB id: 1R9O). (E, F) Two different conformations in the crystal structures of *T. cruzi* (PDB id: 2WX2) and human CYP51 (PDB id: 3JUV), respectively.

The B-propionate group of the heme connects the W tunnel and the active site. The two heme propionate groups show different conformations in the CYP51 crystal structures from different organisms (Table 3.4 and Figure 3.13). Three crystallographic conformations of the two heme propionate groups are shown in Figures 3.13A, E and F and two conformations from the simulations in Figures 3.13B and C. Although small conformational changes of the A-propionate group have been seen in the mutants of cytochrome P450 BM3 [139, 140], such high flexibility of the A-propionate group, and the flexibility of the B-propionate group has not been observed in simulations of other wild-type CYPs. In our simulations, we found that the heme propionate groups of *T. brucei* CYP51 are able to oscillate between different conformations (Figure 3.14) but they prefer the conformation shown in Figure 3.13B which was shown to be occupied in all crystal structures of the membrane-bound CYPs except those of CYP51 and CYP46A1 [24]. The conformation in Figure 3.13C was not found in the crystal structures but was observed in

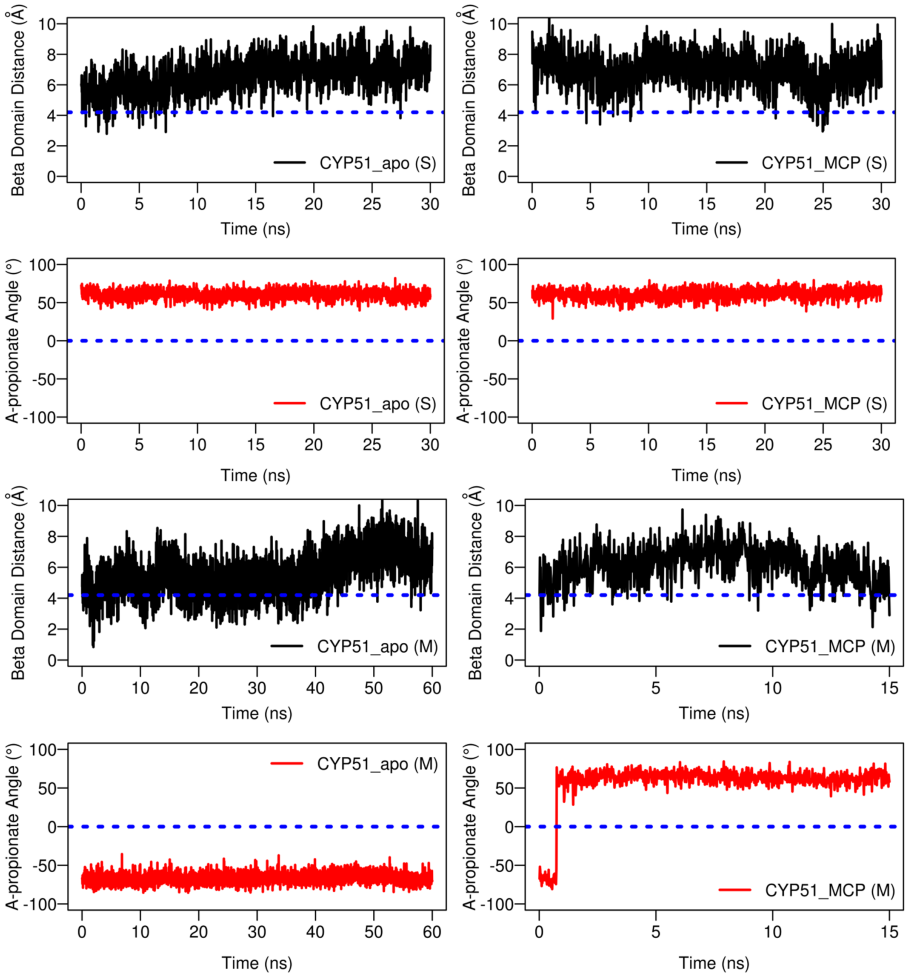

Figure 3.14: The A-propionate angle to the heme plane and the distance of the beta-domain to the heme plane. The angles and distances were plotted for all the four simulations of *T. brucei* CYP51. The blue dotted line is the two values that were proposed by Denisov et al. [24]. In their studies, all the membrane-bound CYPs except CYP51 and CYP46A1 have the A-propionate angle less than 0 which is in the proximal side, and all the membrane-bound CYPs except CYP51 have the beta-domain height smaller than 4.2 Å.

our simulations. The propionate groups might have different conformations when binding to different ligands in the active site.

The conformational changes of the A-propionate group of cytochrome P450 BM3 have been shown to be an effect of mutations but not substrate binding and suggested to influence heme reduction, catalytic efficiency and electron transfer [140]. Our simulations suggest that the heme propionate groups of CYP51 have higher flexibility than those of most of the other CYPs. The high flexibility of the two propionate groups may compensate the rigidity of the binding cavity residues

of CYP51. This flexibility may be important for ligand binding, electron transfer and heme reduction in CYP51.

3.3.4.3 Opening of the W tunnel is affected by the conformation of the heme B-propionate group and is facilitated by ligand binding in the active site

Tunnel W stayed open to a 1.4 Å radius probe in the simulations of *T. brucei* CYP51. We checked the width of tunnel W using CAVER with a bigger probe radius (2.0 Å) for the simulations of *T. brucei* CYP51. The influence of the conformation of the heme propionate groups on the width of tunnel W was analyzed. In the ligand-free simulations, tunnel W stayed closed to the 2.0 Å radius probe in the soluble simulations and was open in 3.5% of snapshots from the membrane-bound simulation. The heme propionate groups mostly adopted a conformation similar to that in Figure 3.13B. However, in the ligand-bound simulations, they adopted a conformation similar to Figure 3.13C and tunnel W was open to the 2.0 Å radius probe in 28% of the soluble simulation and 65% of the membrane-bound simulation.

This analysis suggests that the conformation of the B-propionate group can influence the opening width of tunnel W. In the presence of ligands in the active site, tunnel W tends to be more open. Hence, the presence of a substrate in the active site might facilitate the water transfer between the active site and the solvent necessary for the catalytic cycle.

3.3.4.4 A hydrogen bond network determines the conformation of the heme propionate groups

When the two heme propionate groups are in different conformations, they have different hydrogen bonding partners (Figure 3.13 and Table 3.4). The two propionate groups in CYP51 have fewer hydrogen bonding partners than the propionate groups in the other CYPs studied. In the crystal structures of CYP51, the A-propionate has a maximum of three hydrogen bonds (two with R361) whereas the B-propionate has a maximum of three hydrogen bonds (Table 3.4 and Figure 3.13A, E and F). However, in CYP2C9 (PDB id: 1R9O), the A-propionate forms hydrogen bonds with R97, S365, H368 and S429 whereas the B-propionate forms hydrogen bonds with R97, W120, R433 and one water molecule (represented by a red sphere) (Figure 3.13D). Fewer hydrogen bonding partners might be one reason for the flexibility of the two heme propionate groups in CYP51.

Organism	PDB id	Ligand[i]	Accessibility[ii]	Heme A-propionate		Heme B-propionate	
				Conformation[iii]	H-bonding[iv]	Conformation[iii]	H-bonding[iv]
T. brucei	3TIK, 3P99, 2WV2, 4G3J	+	+	Distal	Y103, R361	Proximal	Y116, R124, H420
	3G1Q	-	+	Distal	Y103, R361	Proximal	Y116, R124, H420
	2X2N	+	+	Distal	Y103, R361	Parallel	Y116, H420
	3GW9, 4G7G	+	+	Distal	R361	Proximal	Y116, R124, H420
	4H6O, 3KSW	+	+	Distal	Y103, R361	Parallel	H420
T. cruzi	2WX2	+	+	Distal	Y103, R361	Parallel	Y116, H420
	3KHM	+	+	Distal	R361	Proximal	R124, H420
	3K1O	+	+	Distal	Y103, R361	Distal	Y116, H420
L. infantum	2WUZ	+	-	Distal	Y103, R361	Proximal	Y116, R124, H420
	3L4D	+	+	Distal	Y103, R361	Proximal	Y116, R124, H420
Human	3JUS, 3LD6	+	-	Distal	Y131, R382	Distal	Y145, K156, H447
	3JUV	-	+	Proximal	R382	Distal	Y145
M. tuberculosis H37Rv	2W09[v]	+	+	Distal	R326	Proximal	—-
	2W0A[v]	+	+	Distal	Q72, R326	Distal	H392
	2W0B	+	+	Distal	Y76, R326	Proximal	K97, H392
	2VKU	+	+	Distal	Y76, R326	Distal (R)[vi]	Ligands
	2CI0[v]	+	+	Distal	R326	Distal&Proximal[vii]	H392
	2CIB[v]	+	+	Distal	R326	Distal	H392
M. tuberculosis	2BZ9, 1H5Zv	-	+	Distal	R326	Proximal	—-
	1U13	-	+	Distal	Y76, R326	Proximal	K97, H392
	1X8V[v]	+	+	Distal	Y76, R326	Proximal	K97, H392
	1EA1, 1E9X	+	+	Distal	Y76, R362	Proximal	R95, H392

Table 3.4: Accessibility of the water tunnel to solvent and hydrogen bonding of the two heme propionate groups in the crystal structures of CYP51. If there is more than one chain, chain A is used. (i) with (+) or without (-) ligand in the active site. (ii) The water tunnel (tunnel W) is accessible (+) or inaccessible (-) to the solvent (with a probe radius of 1.4 Åin PyMOL). (iii) The propionate is on the distal side (Distal) or the proximal side (Proximal) of the heme plane, or roughly parallel (Parallel) to it. (iv) The hydrogen bonding partners of the heme propionate groups. A hydrogen bond is considered to be formed when the distance between the two non-hydrogen atoms is less than 3.3 Å. The residue number is the residue number in the corresponding PDB structures. "—-": the propionate has no hydrogen bonding with surrounding residues. "Ligands": the propionate makes hydrogen bonds to the ligands. (v) There are residues missing in the B-C loop and C helix region of the crystal structures. 2W09, 2W0A, 2CI0 and 1X8V have the B'-C loop missing. 2CIB has the B'-C loop and the beginning of the C helix missing and 1H5Z has the C helix missing. (vi) The B-propionate has rotated 120 degrees towards the I helix because of the interaction with the two ligands (one in the binding site and the other in the ligand tunnel 1). (vii) There are two conformations of the B-propionate group in the crystal structure. One conformation is on the distal and one on the proximal side of the heme plane.

Conformational changes of the two propionate groups, which are driven by the changes of the interactions with neighboring residues, were seen in our simulations of *T. brucei* CYP51. The orientation of the A-propionate group is mainly driven by the hydrogen bonding interaction with R361. The A-propionate group changed its conformation from that in Figure 3.13A to that in Figure 3.13B when the two hydrogen bonds with R361 were disrupted by water molecules in the active site in the ligand-free simulation. However, when the hydrophobic ligand MCP is bound in the active site, the water molecules were expelled from the active site. To regain the two hydrogen bonds with R361, the A-propionate group changed its conformation from that in Figure 3.13B to that in Figure 3.13C. Whereas the conformation of the A-propionate group is largely determined by the interaction with one residue (R361), the conformation of the B-propionate group is determined by its interaction with several neighboring residues, including R124 and H420. In the simulation of the membrane-bound MCP-bound *T. brucei* CYP51, the transition of the B-propionate group from the conformation in Figure 3.13B to that in Figure 3.13C was facilitated by reforming the hydrogen bond with H420.

The hydrogen bonding network between the two heme propionates and the neighboring binding cavity residues plays a crucial role in maintaining the catalytic activity of CYPs [141, 142]. Changes of the hydrogen bond network may affect the binding of ligands, the electron transfer process and the release of water after the catalytic reaction.

3.3.4.5 Functional implications for drug design

The rigidity of the binding cavity residues, from the comparison of CYP51 (*T. brucei* and human), CYP2C9 (human) and CYP2E1 (human), is consistent with the high substrate selectivity of CYP51. The rigidity restricts the reorientation of the binding site residues upon ligand binding. Thus, only ligands that can fit rather perfectly into the binding cavity can bind to the active site of CYP51. This indicates that, for anti-parasitic drug design, the inhibitors have to fit the binding pocket of CYP51 in a similar way to the substrates to have inhibitory effects.

The higher mobility of the tunnel entrance residues of CYP51 than that of other CYPs can facilitate anti-parasitic drug design. The reason for the high mobility could be that as the substrates of CYP51 are large compared with those of CYP2C9 and CYP2E1, the tunnel entrance residues have to open wider to allow the large substrates to enter the active site of CYP51. This indicates that the design of large

molecules that are able to enter the active site of CYP51 but not of human drug-metabolizing CYPs such as CYP2C9 and CYP2E1, is possible. However, CYP3A4 is able to bind to large inhibitors and this would have to be taken into consideration.

Comparison of the dynamics of *T. brucei* CYP51 and human CYP51 indicates that there are limited differences in the SRS regions of the two proteins. Therefore, inhibitors that can only inhibit parasitic CYP51s, and not the human orthologue, may result from the differences in the substrate-contacting residues, such as F105 (L in human), M106 (T in human) and F290 (L in human), rather than differences in the flexibility of the binding cavity residues.

3.4 Conclusion

A model of the membrane-bound *T. brucei* CYP51 was built and simulated with an improved version of the protocol that has been applied to CYP2C9. The model is in good agreement with available experimental data on CYPs. The F-G loop region is more rigid in the simulation of the membrane-bound *T. brucei* CYP51 than in that of the soluble *T. brucei* CYP51 because this region is embedded in the membrane.

We found that, unlike other membrane-bound CYPs (represented by CYP2C9 and CYP2E1) that have high flexibility in the active site and a large range of ligands, the CYP51 active site is relatively rigid in simulations of both soluble and membrane-bound forms. Longer and repeated simulations would be desirable to gain a complete understanding of the dynamics of different CYPs. However, despite their limited length, our simulations show distinct dynamics in the different CYP forms.

Opening and closing of ligand tunnels were observed in CYP51. The simulations suggest that the 2f tunnel is used as a ligand access tunnel and that both 2f and S can potentially be used as ligand egress tunnels. The simulations of the membrane-bound form of *T. brucei* CYP51 indicate that ligands need to enter tunnel 2f from the membrane and can exit via 2f back to the membrane or via S to the solvent. High flexibility is seen in the residues at the surface entrance to tunnel 2f in *T. brucei* CYP51. The rigid B-C loop keeps tunnel W open in both crystal structures and simulations of *T. brucei* CYP51. The conformation of the two heme propionate groups affects the opening width of tunnel W. Tunnel W opened wider in the presence of ligands which indicates a role for this tunnel in water delivery between the active site and the bulk solvent. The hydrogen bonding network

determines the conformation of the two heme propionate groups and can further affect the binding and unbinding of ligands. The conformational changes of the heme's two propionate groups may be important for ligand binding, electron transfer and heme reduction of CYP51.

Access of the ligands to the binding site of CYP51 is facilitated by the flexibility of the 2f tunnel but the relative rigidity of the binding site means that high complementarity in shape and chemical properties is required for binding in the active site. The tunnels provide another potential target binding site for drugs apart from the active site. Both tunnel 2f and tunnel W are more open in CYP51 than in the other human CYPs studied, and might serve as sites for CYP51-specific ligands.

From the comparison of the dynamics of *T. brucei* CYP51 and human CYP51, the reason for the specific inhibition of parasitic CYP51 does not appear to be the difference in the flexibility of the active site but may arise from the differences in the interaction properties of the binding site residues.

4

Ligand Tunnels in P450s

4.1 Introduction

The sterol 14-α demethylase (14DM, CYP51) enzyme plays an important role in sterol biosynthesis [7]. The inhibition of this enzyme has been used as a strategy to develop anti-fungal drugs and, more recently, anti-parasitic drugs [7, 95, 97, 100, 96, 99]. As a member of the cytochrome P450 (CYP, P450) superfamily, eukaryotic CYP51 shares most of the common features of the CYPs, including: (i) embedding in the membrane via immersion of part of its globular domain and a single trans-membrane helix, (ii) a buried active site which requires the ligands to enter and exit from the active site through certain pathways (ligand tunnels), and (iii) two electrons need to be transferred from the redox partners for the catalytic cycle. However, distinct features of CYP51 compared to other CYPs have been observed in experiments and in molecular dynamics (MD) simulations. These features include high substrate specificity, binding site rigidity, different open tunnels between the active site and the protein surface, and plasticity of the two heme propionate groups [7, 40]. Because of the key role of CYP51 for anti-parasitic drug design, it is important to study the above mentioned features of CYP51. In this paper, we focus on the study of how the ligands can enter and egress from the active site of *T. brucei* CYP51, and consider both soluble and membrane-bound forms of the protein. Investigations of the ligand access and egress pathways can help to understand the ligand binding mechanism and ligand specificity of CYP51 and thus can help to design specific parasitic CYP51 drugs, that avoid deactivation and side effects caused by other human drug metabolizing CYPs.

The mappings of ligand tunnels that have been found in different P450s to *T. brucei* CYP51 are, according to Wade's nomenclature [12]: tunnel 2a (between the F-G loop, B-B' loop and β1-1 sheet) and tunnel 2f (between helices A', F" and the tip of the β4 hairpin) which lead to the membrane, and tunnels 1 (between the C, D, H and L helices), 2b (between the B-B' loop and the β1-2 and β1-4 sheets), 2c (between the G and I helices, B'-C loop), 2ac (between the B' helix and G helix), 2e (through the B-C loop), S (the solvent tunnel, between F, I helices and the β4 hairpin) and W (the water tunnel, helix C and the β-bulge segment) which lead to the solvent (Figure 4.1).

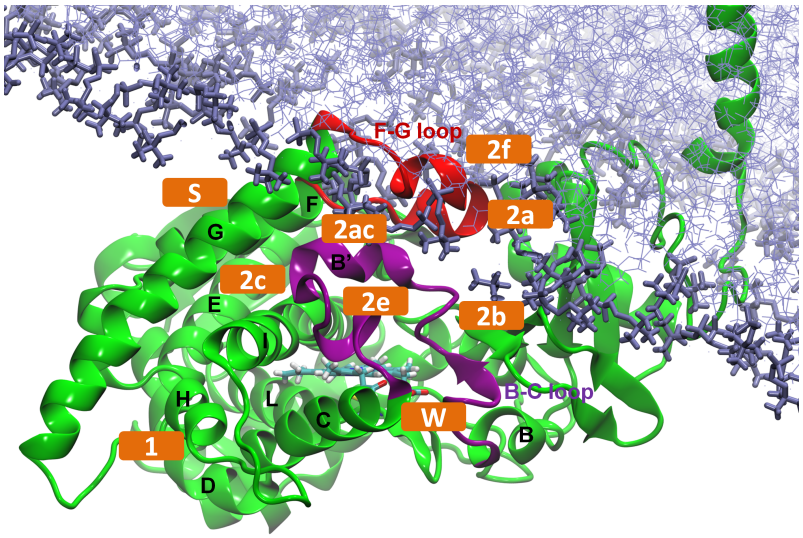

Figure 4.1: Membrane-bound model of ligand-free *T. brucei CYP51* [40]. The important tunnels defined in Wade's nomenclature are labeled (orange). The B-C loop is colored in purple and the F-G loop in red. The heme is shown with sticks representation and helices structures elements are also labeled.

Tunnels 2f, S and W have been shown to be open to a probe with a radius of 1.4 Å, corresponding to the size of a water molecule, in standard MD simulations [40]. However, it is computationally challenging to simulate ligand exit in a feasible time in standard MD simulations. In order to make the simulation of ligand exit computationally possible, the random acceleration molecular dynamics (RAMD) simulation method was developed [16, 17]. In RAMD simulations, an external, randomly oriented force is applied to the center of mass of the ligand to accelerate the ligand exit. As a result, the ligand is able to exit the active site on the time scale of a few hundreds of ps to 2 ns and multiple trajectories can be generated to explore alternative egress routes and mechanisms. RAMD has been successfully used to study the ligand egress pathways of many soluble forms of CYPs

such as P450cam, CYP2B1, CYP2C5, CYP3A4, CYP2E1, CYP2A6 and CYP2C9 and membrane-bound forms of CYP19 [16, 17, 34, 27, 2, 28, 143, 26, 21, 31, 110]. The preferred ligand tunnels in these simulations are mostly the family 2 tunnels including 2a, 2b, 2c, 2e and 2ac. These are all lined by the B-C loop. Because of the differences of CYP51 from the other CYPs, it is interesting to study the ligand tunnels of CYP51, especially in the membrane-bound form.

Here, we address the questions: Can the open tunnels of CYP51 in standard MD simulations be used by inhibitors, substrates and products to enter and exit the active site? Are different ligand tunnels used in the soluble and membrane-bound forms? Which residues gate and modulate the opening and closing of the ligand tunnels? Do the ligand tunnels differ between *T. brucei* CYP51 and human CYP51?

To answer these questions, we built membrane-bound models of ligand-bound CYP51 for four different ligands. These were based on our model of ligand-free *T. brucei* CYP51 in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer [40]. The four ligands are two inhibitors, N-[(1R)-1-(2,4-dichlorophenyl)-2-(1H-imidazol-1-yl)ethyl]-4-(5-phenyl-1,3,4-oxadiazol-2-yl)benzamide (VNI) [7] and 14α-methylenecyclopropyl-Δ7-24,25-dihydrolanosterol (MCP, LNP) [96], the natural substrate of *T. brucei* CYP51 [144], obtusifoliol (OBT) (4α,14-dimethyl-5α-ergosta-8,24(28)-dien-3β-ol), and the product with the 14α-methyl group removed, δ4α-methyl-5α-ergosta-8,14,24(28)-trien-3β-ol (OBT_DM). The structures of these four compounds are shown in Figure 4.2. We also built models of soluble forms of *T. brucei* CYP51 and human CYP51 and simulated them with VNI, LNP, OBT and OBT_DM, and with OBT and OBT_DM, respectively.

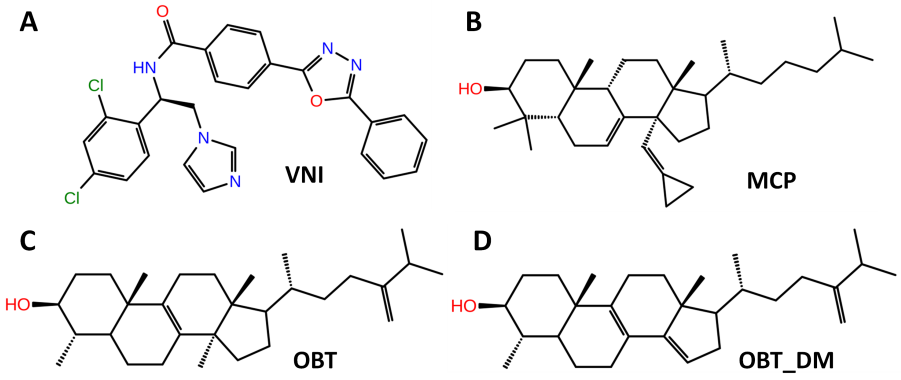

Figure 4.2: Chemical structures of the four ligands studied. The hydroxyl groups are colored in red, nitrogens atoms in blue and chloride atoms in green.

In our simulations, we found that tunnel 2f was the predominant tunnel for all four ligands in both the soluble and the membrane-bound forms of *T. brucei* CYP51. In *T. brucei* CYP51, the solvent (S) tunnel, tunnel 1 and the water (W) tunnel were also used and the common tunnels found in the other CYPs [12, 2, 28] were barely used. The conserved residues F105 and M460 play an important role in modulating the opening of tunnel 2f and S. In human CYP51, tunnel 2f and tunnel 1 were used.

4.2 Materials and Methods

4.2.1 Standard MD Simulations

The membrane-bound model of ligand-free CYP51 in a lipid bilayer of 602 POPC molecules was built from the crystal structure of ligand-free *T. brucei* CYP51 (PDB ID: 3G1Q) using a protocol combining coarse-grained and all-atom simulations as described in [40, 29]. The coarse-grained simulations were performed to find the predominant orientation of CYP51 above the membrane and the all-atom simulations were then performed to refine the membrane-bound model and investigate the dynamics of the system with this predominant orientation. For the soluble models, because of the rigidity of the binding site residues of *T. brucei* CYP51, two of the four ligands (VNI and MCP) were superimposed onto the binding site of ligand-free CYP51 by using the ligand-bound crystal structures (PDB ID: 3GW9 and 3P99), respectively. Coordinates (sdf files, one conformation) of the other two ligands (OBT and OBT_DM) were downloaded from PubChem (CID 65252 and 443237), respectively. They were positioned in the active site so as to adopt a similar orientation to MCP, which is a sterol-like inhibitor. For membrane-bound models of *T. brucei* CYP51, the ligands were superimposed onto the active site of the last snapshot of equilibration from the corresponding equilibrated soluble ligand-bound models. For the simulations of soluble human CYP51, the substrate (OBT) and product (OBT_DM) were superimposed onto the binding site of ligand-free human CYP51 (PDB ID: 3LD6 with crystal ligand removed) by using the OBT-bound and OBT_DM-bound structures of *T. brucei* CYP51. The partial atomic charges of the four ligands were derived from Hartree-Fock 6-31G* *ab initio* calculations and restrained electrostatic potential fit (RESP) on the R.E.D server [121, 120, 145]. Other parameters were derived with the Antechamber

program of AmberTools12 [54]. The final AMBER parameters files are given in the Appendix.

System[i]	Environment[ii]	Method[iii]	No. of Trajectories[iv]	rMin (Å)[v]	Length (ns)[v]
CYP51 + VNI (*T. brucei*)	SOL	MD	1	—	2
		RAMD	30	0.025	—
			30	0.050	—
	MEM	MD	1	—	15
		RAMD	30	0.025	—
			30	0.050	—
CYP51 + MCP (*T. brucei*)	SOL	MD	1	—	2
		RAMD	30	0.025	—
			30	0.050	—
	MEM	MD	1	—	15
		RAMD	30	0.025	—
			30	0.050	—
CYP51 + OBT (*T. brucei*)	SOL	MD	1	—	2
		RAMD	30	0.025	—
			30	0.050	—
	MEM	MD	1	—	3.75
		RAMD	30	0.025	—
			30	0.050	—
CYP51 + OBT_DM (*T. brucei*)	SOL	MD	1	—	2
		RAMD	30	0.025	—
			30	0.050	—
	MEM	MD	1	—	3
		RAMD	30	0.025	—
			30	0.050	—
CYP51 + OBT (human)	SOL	MD	1	—	2
		RAMD	30	0.025	—
			30	0.050	—
CYP51 + OBT_DM (human)	SOL	MD	1	—	2
		RAMD	30	0.025	—
			30	0.050	—

Table 4.1: (i) The CYP51-ligand complexes studied. The full names of the ligands are given in the Introduction section and their structures are shown in Figure 4.2. (ii) The simulations were of soluble models of CYP51 (SOL) and membrane-bound models of CYP51 (MEM). (iii) Simulation method: "MD", standard molecular dynamics; "RAMD", random acceleration molecular dynamics. (iv) "1" indicates that the standard MD simulation was performed once to produce the input snapshots for RAMD. The last snapshot of the equilibration and the last snapshot of the production run were used for RAMD. (v) The length of the standard MD production simulation is listed. The exit time of different ligands via tunnel 2f in RAMD simulations varies, see Figure 4.6, but the maximum time of one RAMD simulation was set to 2 ns.

The ff99SB force field for the protein and the general amber force field (GAFF) for the ligands and POPC lipids [125] were used in the simulations of CYP51s [50]. The TIP3P water model was used for all simulations. The simulations of soluble models were performed with the Amber11/12 software [49] and those of membrane-bound models with NAMD2.9 [55]. The equilibration of the soluble models was performed as described in Chapter 3.2.2. In the production run of the soluble models, the non-bonded cutoff was set to 12 Å. Berendsen temperature control and Langevin dynamics were used to maintain the temperature at 300 K and the pressure at 1 bar with a timestep of 1 fs. For the membrane-bound models, the same simulation procedure with the same parameters as in ref. [40, 29] was applied. Production runs were performed with a timestep of 1.5 fs. In the production run, the non-bonded cutoff was set to 10 Å. Langevin dynamics and the Nosé-Hoover Langevin Piston method with a surface tension of 60 dyn/cm applied to the lipid bilayer plane were used to maintain the temperature at 310 K and the pressure at 1 bar. In all the simulations, all bonds to hydrogen atoms were constrained with the SHAKE algorithm [61].

In the (12 ns) simulation of OBT_DM in the membrane after the RAMD simulation, the last snapshot of the RAMD simulation was taken as the starting structure with all the velocities and periodic box information from that RAMD simulation. The production run was then performed directly with the same parameters as standard MD simulations of membrane-bound models of *T. brucei* CYP51. The simulation of only OBT_DM in the membrane was started from the snapshot after the RAMD simulation. The snapshot at 7.5 ns of the 12 ns simulation with ligand OBT_DM in the membrane was used as the starting structure of the RAMD simulation. A 7.15 ns production run was performed after removing CYP51 without minimization and equilibration.

4.2.2 RAMD Simulations

Random acceleration molecular dynamics (RAMD) as implemented in NAMD2.9 was used to study the ligand egress pathways from the buried active site of CYP51 [16, 17, 34, 27, 28]. The same parameters as in the standard MD simulations were used for the RAMD simulations except that the pressure in the simulations of soluble CYP51 was controlled by a Berendsen barostat at 1 bar and a timestep

of 2 fs was used. In RAMD simulations, the external acceleration, which, in all our simulations is 0.035 kcal/Å·g, is applied with a random initial direction. The distance traveled by the ligand was monitored at 100 fs (50 steps with a timestep of 2 fs) time intervals. If the distance was greater than a pre-defined minimum RAMD distance of 0.025 Å or 0.050 Å, the direction of the force was maintained. If not, the direction of the force was changed to another randomly chosen direction. When the accumulated distance traveled by the ligand was larger than the distance between the exterior of the protein, defined as 30 Å for soluble models and 40 Å for membrane-bound models, and the active site, the RAMD simulation was terminated and the ligand was considered "out". But if the ligand did not exit in 2 ns, the trajectory was recorded as "No exit". For each model studied, the last snapshot from the equilibration and one snapshot (500 ps for the soluble models, 15 ns for the membrane-bound VNI- and LNP-bound models, 3.75 ns for the membrane-bound OBT-bound model and 3 ns for the membrane-bound OBT_DM-bound models) from the production run were used as initial structures for performing RAMD simulations.

4.2.3 Analysis

The concatenation and analysis of the trajectories were performed with the cpptraj tool of AmberTools12 [54]. The trajectory was visualized in VMD1.9 [118]. The figures were made with VMD, PyMOL [116] and Maestro [146]. To plot the ligand 2D structures, the SMILES identifiers of VNI and MCP were taken from the PDB structure (PDB ID: 3GW9 and 3P99) and those of OBT and OBT_DM from the PubChem database.

4.3 Results and Discussion

4.3.1 Predominant Ligand Tunnels

For all four ligands (VNI, MCP, OBT and OBT_DM) in both soluble and membrane-bound forms of *T. brucei* CYP51, the 2f tunnel was found to be the predominant ligand exit tunnel in the RAMD simulations (Table 4.1, Figure 4.3). The 2f tunnel is open in the crystal structures of parasitic CYP51s and was proposed to be the major ligand tunnel of the CYP51 family [7, 109]. We found that the 2f tunnel widened transiently to allow the ligands to exit from this tunnel and that this

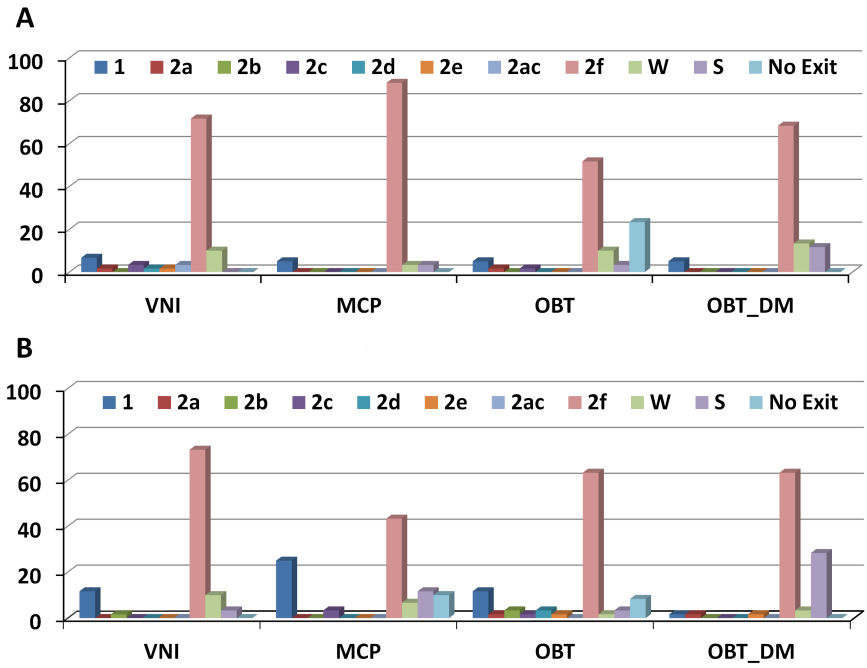

Figure 4.3: The percentage of ligand exit events along each of the ligand egress tunnels in the 60 RAMD simulations of each system. (A), soluble models and (B), membrane-bound models of *T. brucei* CYP51.

widening was facilitated by the flexible movements of the residues at the 2f tunnel entrance.

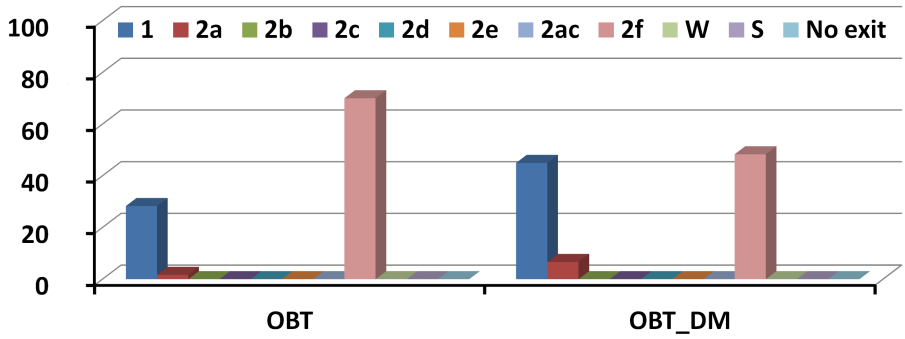

Figure 4.4: The percentage of ligand exit events along each of the ligand egress tunnels in the 60 RAMD simulations of the soluble models of OBT- and OBT_DM-bound human CYP51.

Tunnel S, tunnel 1 and tunnel W were also observed to be ligand egress routes in the simulations of *T. brucei* CYP51 but were not used as often as tunnel 2f (Figure 4.3). The exit times of ligands via these tunnels are shown in Figure 4.7. Tunnel W was used in the simulations of *T. brucei* CYP51 but the possibility of this tunnel to serve as a ligand tunnel *in vivo* is very low because the entrance of this tunnel is on the proximal side which is proposed to be the binding interface of the

reductase [147]. However, in the simulations of soluble models of human CYP51, tunnel 2f was mostly used and tunnel 1 was often used, tunnel 2a was rarely used and tunnel S and tunnel W were not used at all. This shows different tunnel usage in human CYP51 in comparison to *T. brucei* CYP51 (Figure 4.4). Tunnel S was found to be open in standard MD simulations of other CYPs including CYP2C9, CYP3A4 and CYP2D6 (20, 36, 37) and in RAMD simulations of human aromatase (CYP19) [110]. Tunnel 1 has rarely been observed as a ligand exit tunnel in standard MD simulations [12]. However, the fact that tunnel 1 was often used in the RAMD simulations of the soluble models of human CYP51, shows the potential of tunnel 1 to serve as a ligand tunnel. Interestingly, in the crystal structure of *Mycobacterium tuberculosis* H37Rv CYP51 (PDB ID: 2VKU), one bis(4-hydroxyphenyl) methanone ligand (with two possible conformations) is trapped in the middle of tunnel 1 (Figure 4.5) [148]. This suggests that tunnel 1 can serve as a ligand exit tunnel although in this crystal structure, the opening of tunnel 1 results from the I helix being kinked towards the distal side (away from heme). The I helix is not fully conserved in *M. tubercolosis*, *T. brucei* and human CYP51, see Appendix. However, the three residues of the kinked region are relatively well conserved with "AGQ" in *T. brucei* and human CYP51 and "AGH" in *M. tuberculosis* CYP51 and the following four residues are completely conserved. In our simulation, this large conformational change of the I helix was not needed to allow the ligands to exit from tunnel 1.

In the RAMD simulations of human CYP51 with OBT and OBT_DM, it took much shorter times (below 0.5 ns) for the ligand to exit from the active site in comparison to those of *T. brucei* CYP51 (Figure 4.6&4.8). This indicates that via either tunnel 2f or tunnel 1, ligands can exit more easily and quicker from the active site of human CYP51 than from that of *T. brucei* CYP51.

4.3.2 Ligand Tunnels in the Membrane-bound Model of *T. brucei* CYP51

The association of *T. brucei* CYP51 with the membrane did not significantly increase the average exit time through tunnel 2f of any of the ligands in our RAMD simulations (Figure 4.6). In almost 95% of the egress trajectories via tunnel 2f, the ligands entered the membrane after leaving the protein. For the other egress events via tunnel 2f, the ligands entered the solvent.

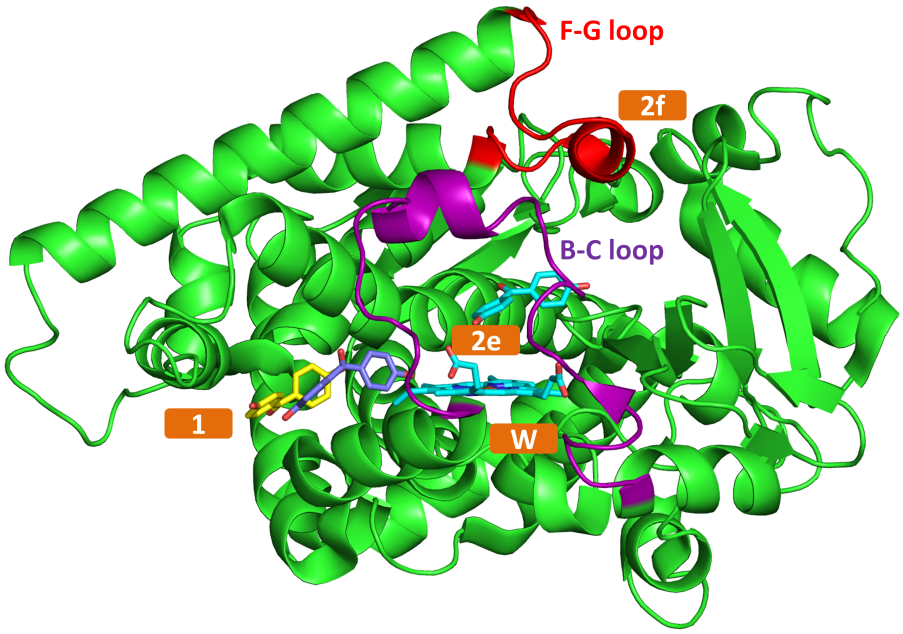

Figure 4.5: One ligand (bis(4-hydroxyphenyl)methanone, DBE) (shown in stick representation in yellow and violet) with two conformations is trapped in tunnel 1 in the crystal structure of *M. tuberculosis* CYP51 (PDB id: 2VKU). Tunnels 1, 2e, W and 2f are labeled and tunnels 1, 2e and W are open to a water sized probe in this structure.

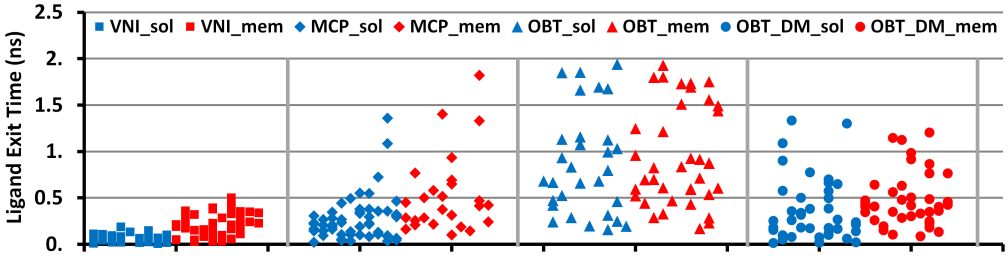

Figure 4.6: The time (ns) required for the ligands to egress through tunnel 2f in *T. brucei* CYP51 in each RAMD trajectory. The RAMD simulations with the soluble models are shown in blue and those with the membrane-bound models are shown in red. The average times of soluble simulations for exit of VNI, MCP, OBT and OBT_DM are 54 $\pm$ 40 ps, 271 $\pm$ 244 ps, 865 $\pm$ 556 ps and 354 $\pm$ 340 ps, respectively. The average times of membrane-bound simulations for exit of VNI, MCP, OBT and OBT_DM are 196 $\pm$ 121 ps, 527 $\pm$ 424, 945 $\pm$ 534 and 484 $\pm$ 295 ps, respectively.

The crystal structures show that the hydrophilic head of the ligands in the active site faces towards the membrane. This means that when the ligands enter the membrane from the protein, the ligands have the hydrophilic head pointing towards the middle of the bilayer. This is in the opposite orientation to the lipids with their polar head groups on the bilayer surface (Figure 4.9A). Ligand reorientation would be necessary for the ligand to exist in the membrane. To investigate whether such reorientation was possible, an additional standard MD simulation

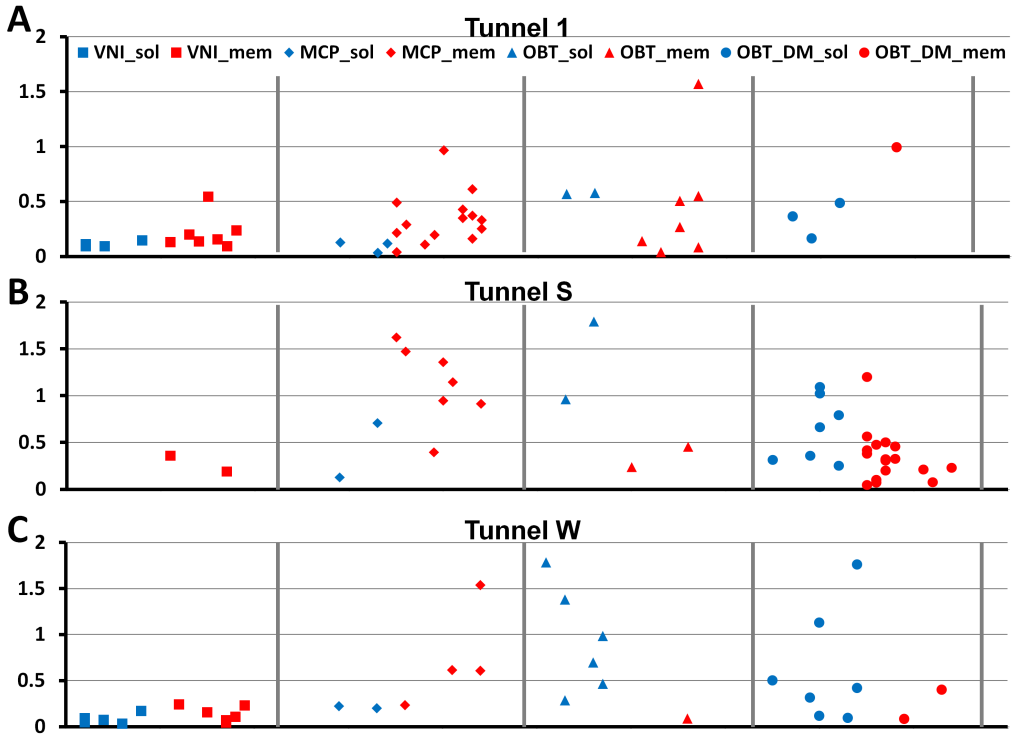

Figure 4.7: The time (ns) required for the ligands to egress through tunnels 1 (A), S (B) and W (C) in *T. brucei* CYP51 in each RAMD trajectory. The RAMD simulations with the soluble models are shown in dark blue and those with the membrane-bound models are shown in red.

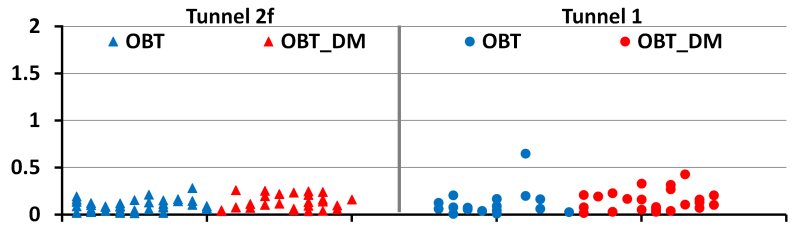

Figure 4.8: The time (ns) required for the ligands to exit through tunnels 2f and 1 in the soluble forms of human CYP51 in RAMD simulations with the substrate (OBT, blue) and the product (OBT_DM, red). The average time for exit of OBT via tunnel 2f and tunnel 1 and, OBT_DM via tunnel 2f and 1 are 95 $\pm$ 58 ps, 124 $\pm$ 150 ps, 132 $\pm$ 77 ps and 143 $\pm$ 106 ps, respectively.

with the natural product OBT_DM in the membrane in the presence of the protein was performed starting from the end snapshot of a RAMD simulation. In this simulation, after ca. 6.5 ns, OBT_DM rotated roughly 100° and changed its orientation from hydrophilic head down in the middle of the bilayer to up towards the headgroup layer. This orientation is similar to that of the POPC lipids (Figure 4.9B). A further simulation (7.15 ns length) without the protein allowed the OBT_DM to orient in the same way as the lipids and to position its hydrophilic head in the head group region of the bilayer (Figure 4.9C) after 150 ps and it kept this orien-

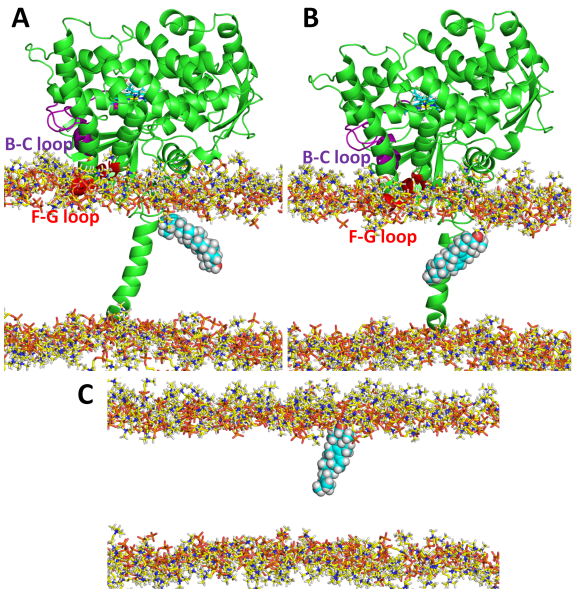

Figure 4.9: The change in orientation of OBT_DM in the membrane upon exit from *T. brucei* CYP51. (A) At the end of the RAMD simulation of *T. brucei* CYP51, the OBT_DM was orientated with its hydrophilic head down in the middle of the bilayer. (B) After a subsequent standard MD simulation of 6.5 ns, OBT_DM had changed its orientation in the membrane by about 100° so that its hydrophilic head pointed up towards the lipid head groups. (C) A further subsequent simulation without the protein starting with the snapshot of (B) allowed the hydrophilic part of OBT_DM to orient in the head group regions of the phospholipid bilayer. Only the head groups of the POPC bilayer are shown for clarity.

tation for the rest of the simulation. These results suggest that the substrates of CYP51 can come from the membrane before the catalytic reaction and as products re-enter the membrane afterwards, and reorient for the next step in catalysis.

4.3.3 Tunnel Entrance Residues and Ligand Specificity

Both *T. brucei* and human CYP51 uses tunnel 2f as the predominant ligand tunnel but they have different affinity to different inhibitors such as VNI [97]. The difference between ligand specificity may result from the entrance residues of tunnel 2f which are not conversed in the two proteins as shown in Figure 4.10. The secondary structure of the F-G loop region differs in the two proteins and the F-G loop region of *T. brucei* CYP51 is closer to the A helix region than the human CYP51. The tunnel entrance residues of tunnel 2f are F48, V77, P210, V213 and F214 (F214 is only identified in *T. brucei* CYP51, not in human CYP51) in *T. brucei* CYP51 and F77, F105, H236 and W239 in human CYP51.

The tunnel entrance residues of tunnel 1 (Figure 4.11) are better conserved than those of tunnel 2f (Figure 7). These residues are Q126, F129 and M284 in

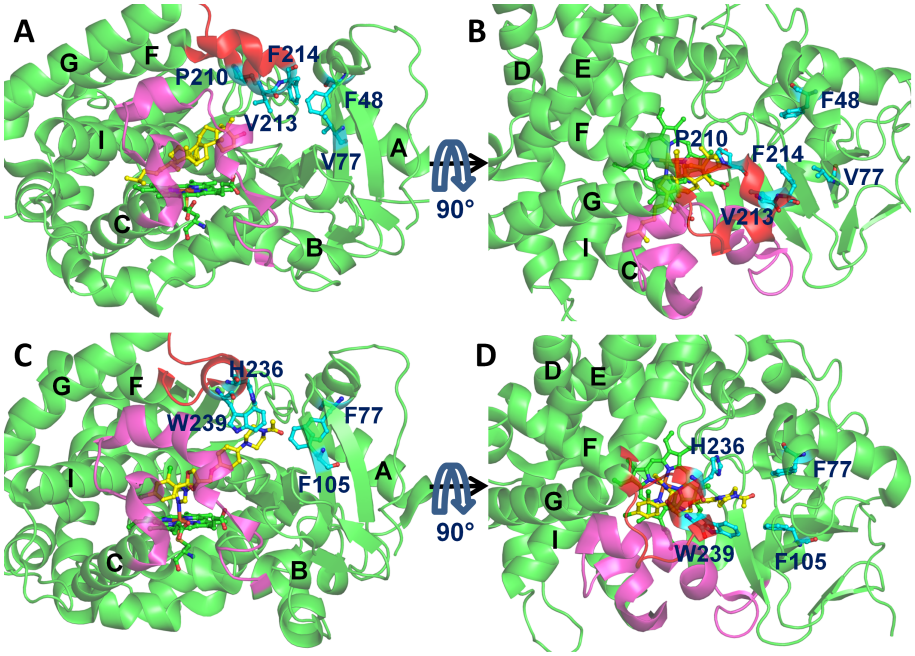

Figure 4.10: Comparison of tunnel entrance residues of tunnel 2f of *T. brucei* CYP51 (PDB ID: 3P99) and human CYP51 (PDB ID: 3LD6). Ligands and heme cofactor are represented by stick representation and in yellow and green, respectively. The B-C loop is in magenta and F-G loop in red. Tunnel entrance residues are labeled in dark blue and important secondary structures are labeled in black.

T. brucei CYP51 and Q155, M158 and M304 in human CYP51. When ligands exit from tunnel 1, both F129 in *T. brucei* CYP51 and M158 in human CYP51 need to undergo large side chain conformational changes to allow ligand passage via tunnel 1. However, conformational change of the side chain of M is easier than that of F which may result in ligand exit more often from tunnel 1 in human CYP51 than in *T. brucei* CYP51 (Figures 4.3&4.4).

The tunnel entrance residues of tunnel S are shown in Figure 4.12. These residues are well conserved in *T. brucei* and human CYP51. However, tunnel S is used in *T. brucei* CYP51 but not used in human CYP51. This is not due to the tunnel entrance residues but the openness of the tunnel. Tunnel S was found to be mostly closed in human CYP51 but to be open in *T. brucei* CYP51 in standard MD simulations as shown in ref. [40]. The low openness of tunnel S in dynamics of human CYP51 results in that this tunnel is not used in human CYP51.

Tunnel W was used in *T. brucei* CYP51 but is not considered as a ligand tunnel because tunnel W leads to the interface between P450 and CPR. Tunnel W is not conversed in *T. brucei* and human CYP51 (Figure 4.13). Tunnel W was not used in human CYP51. Tunnel entrance residues of tunnel W in human CYP51 are more

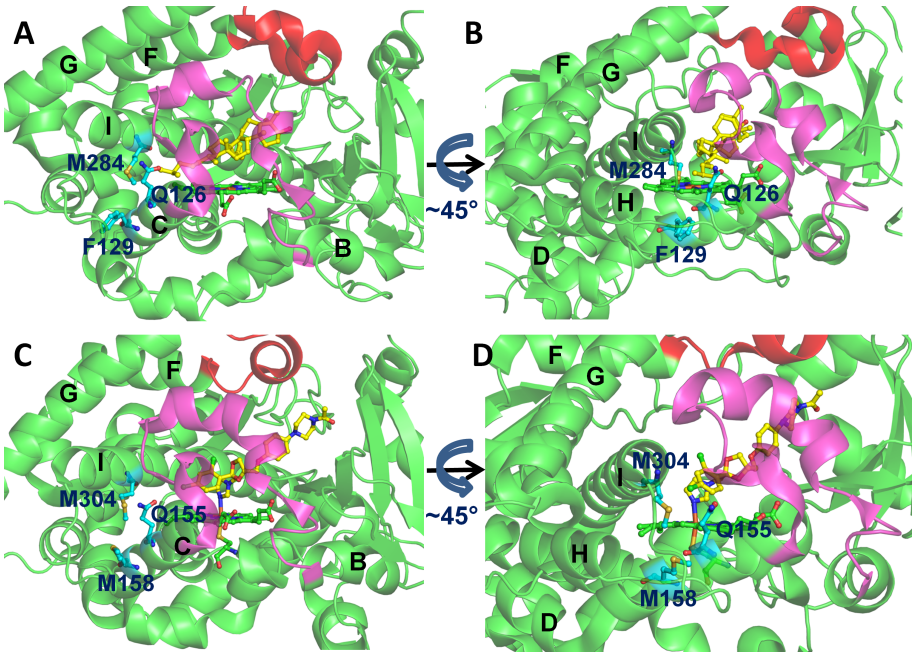

Figure 4.11: Comparison of tunnel entrance residues of tunnel 1 of *T. brucei* CYP51 (PDB ID: 3P99) and human CYP51 (PDB ID: 3LD6). Ligands and heme cofactor are represented by stick representation and in yellow and green, respectively. The B-C loop is in magenta and F-G loop in red. Tunnel entrance residues are labeled in dark blue and important secondary structures are labeled in black.

polar than those in *T. brucei* CYP51 (Figure 10). This can make ligands of CYP51 which are normally hydrophobic ligands more difficult to exit in human CYP51 than in *T. brucei* CYP51. In standard MD simulations, tunnel W was found to be open in *T. brucei* CYP51 and closed in human CYP51 [40].

4.3.4 Gating Residues of Ligand Tunnels

The tunnels through which ligand egress was most often observed in the RAMD simulations of *T. brucei* CYP51 were tunnels 2f and S. Tunnel 2f was open at the beginning of the RAMD simulations but not wide enough for the large ligands studied to exit. Tunnel S was closed (Figure 4.15A&B). During the ligand exit, tunnel 2f opened wider and tunnel S opened from the closed conformation (Figure 4.15C&D). The opening of the two tunnels allowed the exit of the ligands.

By analyzing the RAMD trajectories, we found two residues (F105 and M460) that played an important role in the opening of these two tunnels and they are considered to be the gating residues of these two tunnels. Long distances (more than 5 Å) between these two residues correspond to open tunnels that can accom-

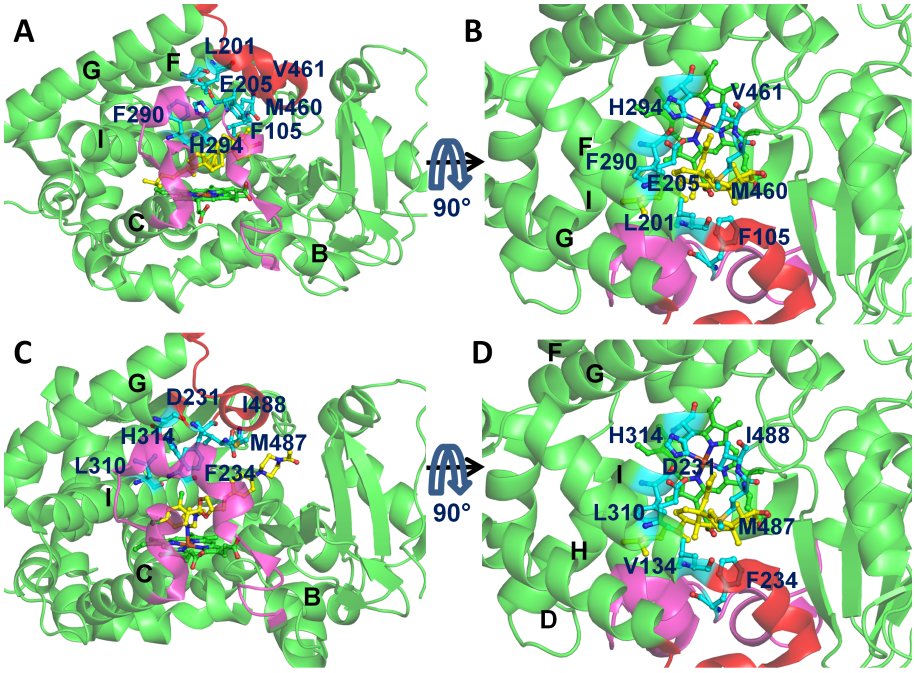

Figure 4.12: Comparison of tunnel entrance residues of tunnel S of *T. brucei* CYP51 (PDB ID: 3P99) and human CYP51 (PDB ID: 3LD6). Ligands and heme cofactor are represented by stick representation and in yellow and green, respectively. The B-C loop is in magenta and F-G loop in red. Tunnel entrance residues are labeled in dark blue and important secondary structures are labeled in black.

modate the entrance and exit of the ligands. When the ligands are in the active site, they can help to keep the ligands in the active site during the catalytic reaction (Figure 4.14A&C) by interacting with each other. Both residues are phyla-specific in the CYP51 family. F105 is L in fungi and animals and I in *T. cruzi* and M460 is L in fungi [96]. In human CYP51, F105 is V and M460 is M. The mutations F105Y and F105A lead to the loss of activity of *Mycobacterium tuberculosis* CYP51 (39). The F105Y or F105A mutant may influence the binding of substrates in the active site. In our RAMD simulations, the concerted movement of the side chains of F105 and M460 can result in either tunnel 2f or tunnel S opening for ligand passage (Figure 4.14B&D), depending on the directions of the side chain conformational changes. Because tunnel 2f leads to the membrane and tunnel S leads to the solvent, these two gating residues can then determine whether the ligands are released into the membrane or into the solvent.

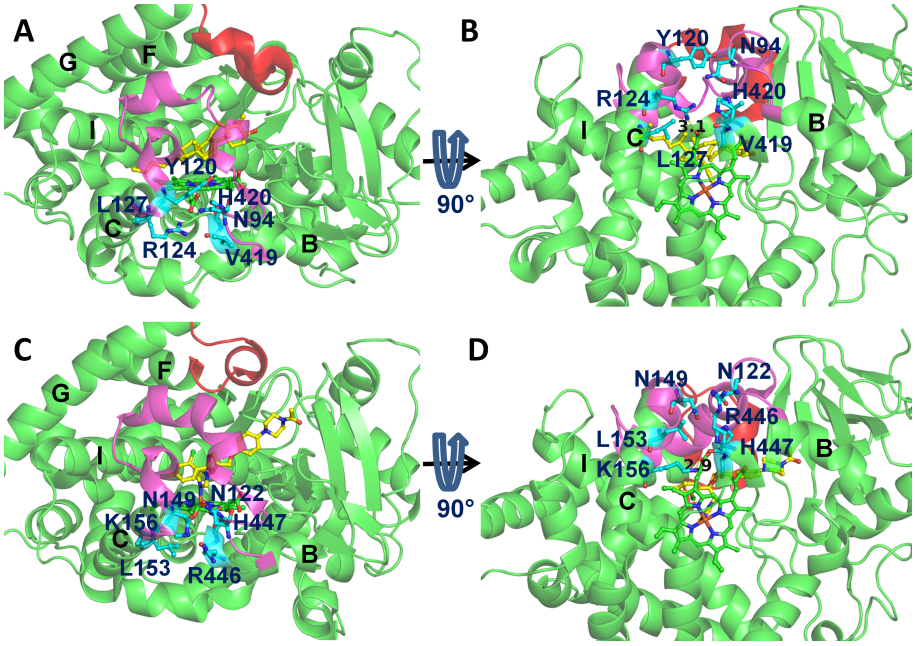

Figure 4.13: Comparison of tunnel entrance residues of tunnel W of *T. brucei* CYP51 (PDB ID: 3P99) and human CYP51 (PDB ID: 3LD6). Ligands and heme cofactor are represented by stick representation and in yellow and green, respectively. The B-C loop is in magenta and F-G loop in red. Tunnel entrance residues are labeled in dark blue and important secondary structures are labeled in black.

4.4 Conclusion

The access and egress pathways of four ligands (VNI, LNP, OBT and OBT_DM) from the buried active site of *T. brucei* CYP51 and OBT and OBT_DM from that of human CYP51 were studied with RAMD simulations. Here, by simulating ligand egress, we find that the preferred exit route of ligands is via tunnel 2f in both soluble and membrane-bound forms of *T. brucei* CYP51 and soluble forms of human CYP51. In the RAMD simulations of *T. brucei* CYP51, tunnels S, 1 and W can also serve as the egress pathway for the ligands to the solvent but is much less frequently observed than tunnel 2f. For human CYP51, tunnels S and W were not used for ligand egress in any of the simulated trajectories. However, tunnel 1 was used more often in human than in *T. brucei* CYP51, suggesting that tunnel 1 may serve as a ligand tunnel to the solvent in human CYP51. The use of tunnel 1 is supported by the crystal structure of *Mycobacterium tuberculosis* CYP51 which shows one ligand in tunnel 1. The other tunnels are less likely to be used as ligand tunnels. Tunnel entrance residues may be the key for ligand specificity of *T. brucei* and human CYP51. The difference between tunnel entrance residues of tunnel 2f

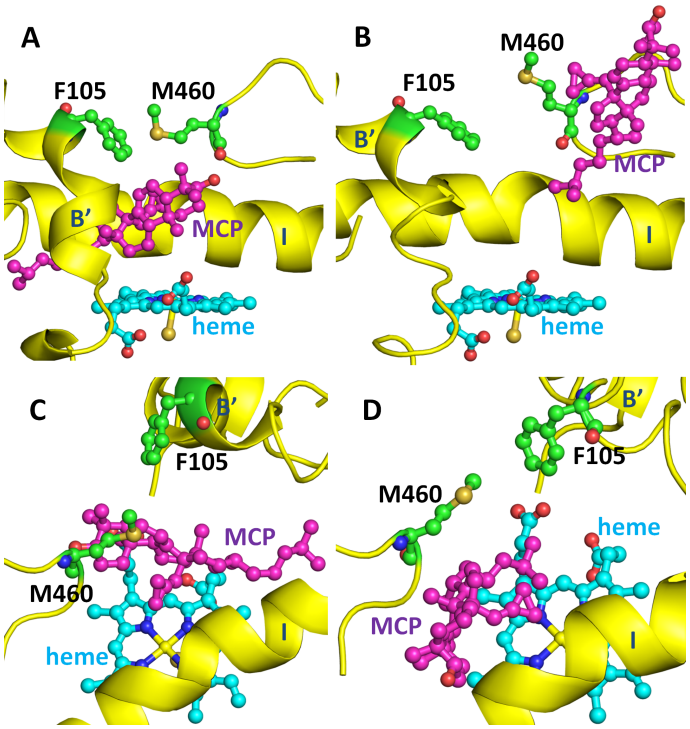

Figure 4.14: Reorientation of the gating residues of tunnel 2f and tunnel S, F105 and M460, in *T. brucei* CYP51 during egress of MCP. The side chain orientations of F105 and M460 are shown before (A) and during ligand exit (B) via tunnel 2f and before (C) and during ligand exit (D) via tunnel S. The secondary structure is shown in yellow and the heme in cyan. Corresponding snapshots of different views of the protein with a surface representation are shown in Figures 4.15A-D.

may result in different specificity of *T. brucei* and human CYP51 towards the same ligand, such as VNI. The substitution of M to F in human CYP51 may result in the more often use of tunnel 1 than *T. brucei* CYP51. In *T. brucei* CYP51, binding site residues F105 and M460 are the gating residues of tunnel 2f and tunnel S. They can not only modulate the opening of the two tunnels but also modulate whether the ligands should go to the membrane or the solvent.

Despite reducing protein flexibility, the presence of the membrane bilayer did not obviously increase the time needed for the ligand exit in our RAMD simulations of membrane-bound models compared to the soluble models of CYP51. In all our simulations of membrane-bound models of *T. brucei* CYP51, the ligands preferred tunnel 2f as the exit tunnel and entered the membrane after leaving the protein. Therefore, in CYP51, the results suggest that the substrates come from the membrane and the products go back to the membrane after the catalytic reaction.

Our study helps to understand the ligand access and egress mechanism of CYP51. Analysis of tunnel entrances residues show differences between *T. bru-*

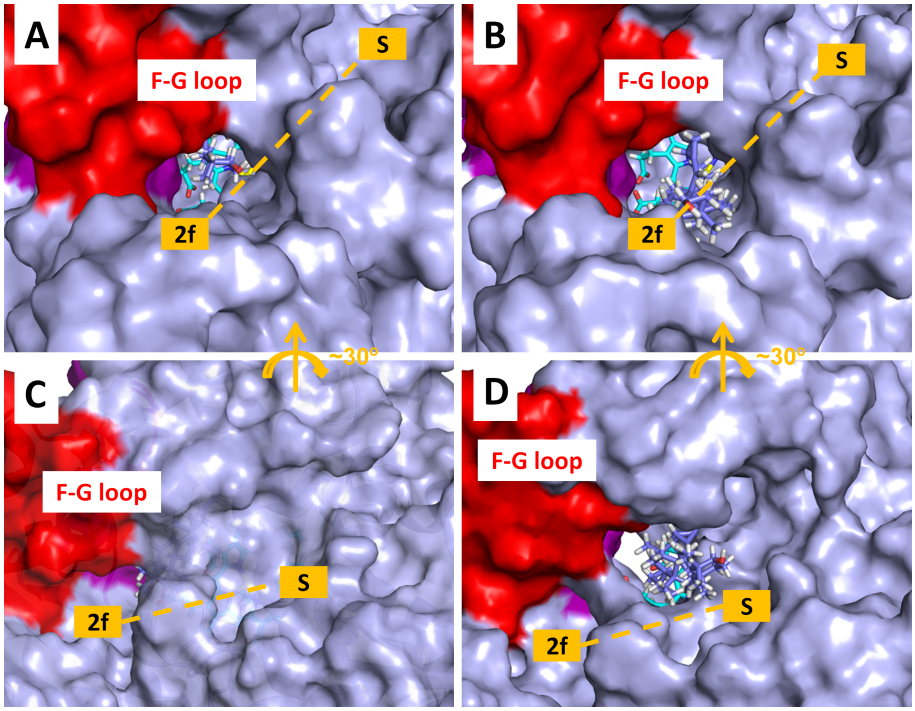

Figure 4.15: Tunnel opening during egress of MCP from *T. brucei* CYP51. The protein is shown with a solvent accessible molecular surface representation. The heme is shown in stick representation (cyan carbons) and the ligand MCP in stick representation (violet carbons). The F-G loop is in red and the B-C loop in purple. First snapshots of RAMD simulations are those taken after standard MD simulation. (A) First snapshot of a RAMD trajectory with tunnel 2f open. (B) Snapshot from the same trajectory at 132 ps showing that tunnel 2f has opened wider to allow the MCP to exit. (C) First snapshot of another RAMD trajectory with the tunnel S closed to a 1.4 Å radius probe. (D) Snapshot from the same trajectory at 902 ps showing that tunnel S has opened wide enough to allow MCP to exit. The white space below the heme center shows the open tunnel W.

cei and human CYP51 which give insights into the design of parasite-specific inhibitors.

5

Interactions of Cytochrome P450s and Their Redox Partners

5.1 Introduction

Cytochrome P450s (CYPs, P450s) are heme-containing enzymes and exist in all biological kingdoms. P450s play an important role in drug metabolism and sterol biosynthesis. The most predominant catalytic function of P450s is to catalyze monooxygenation of organic compounds. For the catalytic reaction, electrons need to be transferred to the active site from the redox partners of P450s. There are various redox partners of P450s. In prokaryotes, redox partners of P450s include putidaredoxin (Pdx), cindoxin (Cdx) and ferredoxin (Fdx). In eukaryotes, these include adrenodoxin (Adx), cytochrome *b5* and cytochrome P450 reductase (CPR). Two electrons are transferred to the active site of P450s in two steps, one electron per step. It is generally believed that two electrons are transferred from the same kind of redox partner but the second electron may also come from a different kind of redox partner, for example, in mammals, the second electron may come from CPR or cytochrome *b5* [149]. CPR has different conformations including a closed conformation (Figure 5.1A), a partially open conformation (Figure 5.1B) and an open conformation (Figure 5.1C). The closed conformation is used for the internal electron transfer of CPR and the open conformation is used to transfer electrons to other proteins such as P450s.

The electron transfer process of P450s has been studied a lot during the last decades and a basic understanding of the system has been achieved. Complexes of P450s and their redox partners have been crystallized by various groups [150, 151,

152, 153]. The complex of the P450BM-3 domain and the flavodoxin-like flavin domain (FMN domain) was the first crystallized complex of P450s and their redox partners [150]. P450BM-3 is a bacterial P450 system that has the P450 domain fused to a mammalian-like FMN domain in a single polypeptide chain [154]. This structural information gives initial insights into how P450s can interact with their redox partners. The flavin mononucleotide (FMN) cofactor locates at the interface between the P450BM-3 domain and the FMN domain, and thus can transfer electrons via salt bridges to the heme center, which is the active site of P450s. However, the P450BM-3 domain and the FMN domain were separated and crystalized which may be the reason for the complex not being functional for electron transfer and catalytic reactions. The complex of mitochondrial P450 11A1 and adrenodoxin (Adx) was crystallized later [151]. A thorough study of interactions of P450cam and putidaredoxin (Pdx) has been carried out by Tripathi et al. [152]. The complexes of both substrate-free and camphor-bound P450cam and putidaredoxin were crystallized. The conformational changes of the binding interface with Pdx upon camphor binding and a water-mediated H-bonded network for electron transfer were observed. An NMR study was carried out by Hiruma et al. [153] to study the complex of P450cam and Pdx. They were also able to crystallize the complex and the crystallized complex is very similar to those obtained by Tripathi et al. Two pathways with strong electrostatic coupling were proposed and the electron transfer rate was calculated.

For complexes of eukaryotic P450s and CPR, there is no such complex crystallized yet. However, mutation studies have been performed to study these interactions. Im et al. summarized the experimental studies of interactions of CYP2B4 with both cytochrome P450 reductase and cytochrome *b5* [149]. Mutations of residues on the positively charged patch of the proximal side of CYP2B4, which include R122A, R126A, R133A, R422A, K433A and R443A, significantly decreased Kd of the complex as shown in Table 5.2 [149]. They also proposed important interface residues on CPR, which are D113, E116, E142 and D144. A modeled complex which was based on these mutation data was also built using PyMOL [149]. Later on, the same group was able to crystallize the open conformation of CPR [156]. This crystal structure gives a structural basis for studying the interactions between P450s and CPR using computational methods. A mutation study by Kenaan et al. showed two hydrophobic but surface exposed residues (V267 and L270) on CYP2B4 that are important for the CYP-CPR interaction [157]. These

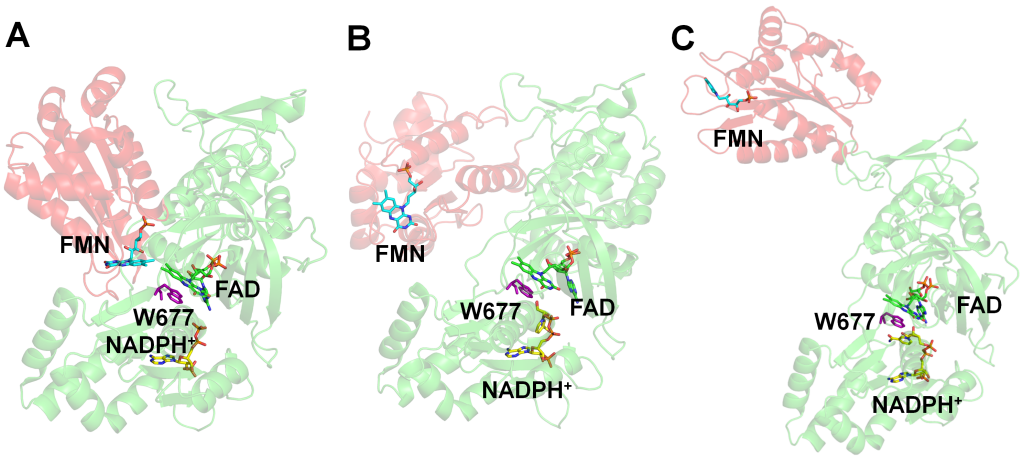

Figure 5.1: Different conformations of CPR: closed conformation of human CPR (PDB ID: 3QE2) [155] (A), partially open conformation of rat CPR (PDB ID: 3ES9) [156] (B) and open conformation of a model of human CPR (protocol described in Methods section) (C). The FMN domain is in red and the FAD domain in green. The FMN (cyan), FAD (green) and NADPH (yellow) cofactors are represented with sticks representation, so is the gating residue (W677, magenta) for internal electron transfer.

two residues are not on the proposed proximal binding site of P450s but in the region of the H and I helices. Mutations D113A, E115A and E116A on rat CPR affect binding of CPR to its electron acceptors. Thus, these residues may reside on the binding site of CPR to CYP2B1 [158].

Comparison of interactions between mammalian P450s and CPR has been extensively performed by Backes et al. [159, 160, 161, 162, 163]. They showed that in a system of CYP2B4, CYP1A2 and CPR, CPR has higher affinity to CYP1A2, and that CYP2B4 and CYP1A2 can form P450-P450 complexes [159]. Hazai et al. built a model of the complex of CYP2C9 and CPR based on the crystal structure of the complex of P450-BM3 and the FMN domain [164]. A model of CYP2D6 and human CPR on a membrane was built and molecular dynamics simulations were performed to study the dynamics of the system [165]. The complex of P450-BM3 and the FMN domain was used to build a model of CYP2B4 and the FMN domain of reductase [166]. The positively charged clusters of 343R-345K and 433K-434R of CYP2B4 were suggested to interact with the negatively charged clusters of 207D-208D-209D and 113D-115E-116E of CPR, respectively.

However, due to the experimental difficulties and the complexity of the systems, there are still a lot of open questions about the interactions between P450s and CPR. For example: How do different P450s interact with CPR? How do different P450s compete for CPR? Are there any differences in the affinity between

different P450s for CPR? What is the possible route for the electron transfer between CPR and different P450s?

These questions are hard to answer by experiments and thus computational methods provide a complementary way to tackle these problems. Here, we use a combination of Brownian dynamics and molecular dynamics simulations to try to answer these questions. With this approach, our simulations are able to reproduce the experimental data and then we used this approach to investigate the interactions between various P450s and CPR. Our simulations show that different P450s bind to CPR using the same interface, which is on the proximal side of P450s, but with different affinities (inferred from binding energies). P450s with ligands bound in the active site are shown to have higher affinity to bind to CPR than the ligand-free forms.

5.2 Methods

The procedure described in [83] was used to dock CPR to different P450s. The procedure consists of two steps: the first step is to use Brownian dynamics simulations to generate encounter complexes of the two proteins using biochemical data. The following molecular dynamics (MD) simulation using implicit solvent is then performed to refine complexes obtained from Brownian dynamics simulations.

5.2.1 Protein Structure Preparation

Crystal structures of complexes of P450s and their redox partners are used to evaluate the use of Brownian dynamics simulations for systems of P450s and their redox partners. The complexes used are: (1) the crystal structure of the complex of the P450BM-3 domain and the flavin mononucleotide (FMN) domain (PDB ID: 1BVY), and (2) that of mitochondrial CYP11A1 and adrenodoxin (PDB ID: 3N9Y).

Crystal structures of P450s are used as starting structures for docking. The PDB structures used for different P450s are shown in Table 5.1. When there are ligands bound to the active site, ligands are removed when performing simulations. The removal of ligands will have little influence on results of rigid-body Brownian dynamics simulations. When there are missing residues that are not on the possible binding interfaces for CPR of crystal structures, Brownian dynamics simulations are performed without adding these missing residues. However, for

CYP	Organism	PDB id	Ligand[i]	Resolution (Å)	Missing Residues[ii]
1A1	Human	4I8V	BHF	2.60	No
1A2	Human	2HI4	BHF	1.95	No
2A6	Human	2FDU	D1G	1.85	No
2B4	Rabbit	1SUO	CPZ	1.90	No
2C9	Human	1OG2	—	2.60	No
		1R9O	FLU	2.00	Yes
2D6	Human	2F9Q	—	3.00	Yes
		3TBG	RTZ	2.10	No
2E1	Human	3T3Z	9PL	2.35	No
3A4	Human	1TQN	—	2.05	Yes
		3NXU	RIT	2.00	Yes
51	*T. brucei*	3G1Q	—	1.89	No
		3P99	LNP	3.00	No
51	Human	3LD6	KKK	2.80	No

Table 5.1: CYP crystal structures used for docking to CPR. (i) Ligands in crystal structures are: BHF: 2-phenyl-4H-benzo[H]chromen-4-one, D1G: N,N-dimethyl(5-(pyridin-3-yl)furan-2-yl)methanamine, CPZ: 4-(4-chlorophenyl)imidazole, RTZ: 10-2-[(2R)-1-methylpiperidin-2-yl]ethyl-2-(methylsulfanyl)-10H-phenothiazine, 9PL: (3S,4R)-3-ethyl-4-[(1-methyl-1H-imidazol-5-yl)methyl]dihydrofuran-2(3H)-one, LNP: (3alpha,9beta,10alpha,13alpha)-30-cyclopropylidenelanost-7-en-3-ol, KKK: ketoconazole. (ii) The missing residues do not include missing residues on the N- and C-termini.

the refinement using molecular dynamics simulations, these residues are modeled using the Modeller software as described in [29].

For CPR, chain B of PDB ID 3ES9, which is the open conformation of the rat CPR, was used as the template to model the human CPR in the open conformation (the alignment of human and rat CPR in Appendix). The FMN domain and FAD domain of the crystal structure of human CPR (PDB ID: 3QE2) were superimposed onto the chain B of rat CPR individually. The hinge region (residues 231-243) of human CPR was also superimposed onto the hinge region of the crystal structure and the missing residues (TGEE, 236-239) were then modeled in VMD.

PDB2PQR [117, 167] was used to add hydrogens to crystal structures as well as the van der Waals radii and partial charges of amino acids. Proteins were protonated at pH 7. The AMBER force field was used [42]. Generated files are in PQR format. PQR files are PDB files with additional columns for the atomic charge and radius of each atom. The parameters for the heme center were assigned as in [119] and [29] and were added to the PQR file manually. The partial charges of cofactors were derived using a two-stage RESP fit method: RESP-A1A (HF/6-31G*) of the R.E.D server [121]. The parameter files of the two ligands were

generated using the Antechamber program of AmberTools 1.5 with the generalized Amber force field (GAFF) [54]. Hydrogens of cofactors were added using the tleap program of AmberTools 1.5 and derived charges for cofactors were added to the PQR files.

5.2.2 Brownian Dynamics Simulations

The electrostatic potential of the proteins was calculated with the adaptive Poisson-Boltzmann software (APBS) by solving the non-linearized Poisson-Boltzmann equation [76]. In the APBS calculation, 193 grid points were used for all three dimensions (x, y and z) with a 1 Å spacing. The ionic strength was set to 150 mM. A dielectric constant of 2.0 was used for the protein and of 78.0 was used for the solvent. Effective charge sites were defined using the effective charges for macromolecules in solvent (ECM) method [75]. Effective charge sites are on the charged amino acids and cofactors (O1A, O2A, O1D and O2D atoms of heme and, O1P and O2P atoms of FMN).

Rigid-body Brownian dynamics simulations were performed using the Simulation of Diffusional Association (SDA 7) software package. In SDA simulations, the larger protein is fixed, whereas the smaller protein is mobile with both rotational and translational diffusional motions. 20000 SDA runs were performed. For the exclusion grid, which was only applied to the first protein, to avoid clashing of proteins, a spacing of 0.25 Å was used in simulations and the solvent probe radius was set to 1.4 Å. At the beginning of simulations, the second protein was put at a random orientation 150 Å away from the first protein. It diffuses under electrostatic interactions. When the two proteins are close to each other and form an encounter complex, this encounter complex is recorded only when the reaction criterion is satisfied. In this case, the distance between the FE atom of the heme and the N5 atom of the FMN cofactor is smaller than 25 Å. This constraint was applied to ensure that the cofactors are close enough for electron transfer. 2000 encounter complexes were recorded.

The 2000 recorded complexes which have the lowest total interaction energy were then clustered into ten clusters using the protocol described in [83]. A distance matrix was constructed by calculating the backbone root mean square deviation (RMSD) of each docked protein configuration from each other. At the first step, each docked structure was assigned to a separate cluster. The closest clusters were then merged in the following steps until the number of clusters, in this case

ten clusters, was obtained. A representative structure of each cluster was selected which has the lowest RMSD to the centroid of the cluster.

5.2.3 Molecular Dynamics Refinement

Molecular dynamics simulation was used to refine cluster representatives after rigid-body docking to introduce flexibility into the proteins. AMBER14 was used to perform molecular dynamics simulations. The docked CPR was combined with the P450 into one file after SDA docking. Parameters of cofactors were generated using the procedure mentioned in Chapter 3.2.2. The tleap program of AmberTools13 was used to generate the topology file of the complex. In minimization, the maximum number of minimization cycles was set to 1000 and it was switched from steepest descent to conjugate gradient after 500 steps. The modified generalized Born (GB) implicit solvent model was used for solvent and the GB salt concentration was 100 mM [168]. The generalized Born/surface area (GB/SA) method was used. In production runs, the GB implicit solvent model was also used. Langevin dynamics with collision frequency γ=2 was used for temperature coupling and no pressure scaling was used. No periodic boundary condition was applied. Bonds involving hydrogen were constrained using the SHAKE method. A timestep of 1 fs was used and the production run was performed for 100000 steps. In the production run, the system was heated from 30 K to 300 K in the first 10000 steps and the temperature was kept constant thereafter.

5.2.4 Interface Residues

Interface residues were identified using the "InterfaceResidues" module of PyMOL [116]. The cutoff value was set to 7.5 Å. The residues of one protein that were within 7.5 Å of surface residues of the other protein were defined as interface residues.

5.3 Results and Discussion

5.3.1 Validation of SDA for docking redox partners to P450s

A key advantage of SDA, in comparison to many other docking softwares and web servers, is the possibility to use biochemical data as constraints in simulations.

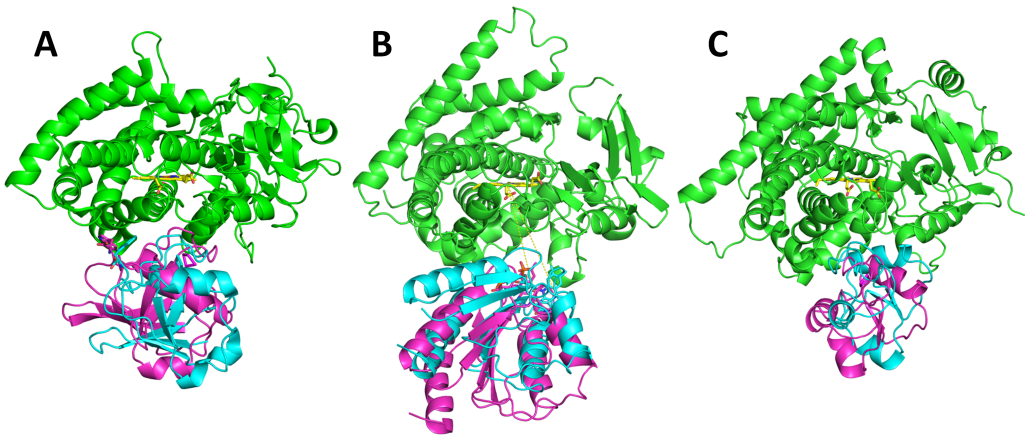

Figure 5.2: Comparison of SDA docked complexes of P450 (green) and electron transfer partner (cyan) by Motiejunas et al. [83] (A) and in this study (B&C) with structures of the crystal complexes (magenta) of (A) cytochrome P450cam and putidaredoxin (Pdx) (PDB ID: 4JWS), (B) cytochrome P450-BM3 domain and FMN domain (PDB ID: 1bvy), and (C) cytochrome P450 11A1 and adrenodoxin (Adx) (PDB ID: 3N9Y).

SDA has been successfully applied to predict protein complexes in combination with biochemical data [169, 82, 81]. SDA was used to predict the complex of P450cam and putidaredoxin in 2007 [79]. The orientation of Pdx in the predicted complex is similar (ca. 15° difference along x axis) to that in the crystal structure of the complex published by Tripathi et al. in 2013 (Figure 5.2A) [152]. SDA and subsequent molecular dynamics (MD) refinement were also used by Madrona et al. [170] to predict the complex of P450cin and cindoxin. This shows that SDA simulation followed by molecular dynamics refinement can be applied to predict encounter complexes of P450s and their redox partners.

In this study, the two available complexes of P450s and their redox partners are used to assess SDA: the complex of CYP11A1 and adrenodoxin (Adx) and that of the P450BM-3 domain and the FMN domain. For the complex of CYP11A1 and Adx, the representative of the first lowest energy cluster is classified as an acceptable CAPRI model [171] evaluated using the crystal complex (Figure 5.2C). This representative has 19% native contacts and is considered to be acceptable because it has more than 10% native contacts. SDA was able to predict a similar orientation of the FMN domain to the crystal structure of the complex of P450BM-3 domain and FMN domain (Figure 5.2B), although it is not classified as a CAPRI acceptable model because the docked complexes have 0% native contacts. This could due to the fact that the crystal complex is not active and the FMN domain may have a slightly different binding interface *in vivo* because of restrictions due to

CYP2B4	Kd (μM)[i]	Cluster 1		Cluster 2		Cluster 3	
		Energy[ii]	Size[iii]	Energy[ii]	Size [iii]	Energy [ii]	Size [iii]
WT	0.02 ± 0.02	-9.3 ± 1.0	5977	-10.3 ± 1.4	4996	-8.7 ± 0.6	3383
R122A	0.23 ± 0.08	-1.3 ± 0.5	550	-1.1 ± 0.3	263	-1.3 ± 0.3	238
R126A	0.27 ± 0.08	-2.5 ± 0.9	1419	-2.7 ± 0.7	1025	-2.7 ± 0.7	598
R133A	1.2 ± 0.57	-2.5 ± 0.9	2159	-2.0 ± 0.5	2219	-2.8 ± 0.6	1662
K433A	1.1 ± 0.81	-1.7 ± 0.5	1489	-1.8 ± 0.5	965	-1.5 ± 0.6	576

Table 5.2: Measurement of the binding affinities of CYP2B4 and rat CPR. (i) Experimental data taken from [172]. (ii) Average total interaction energy (k_BT) of all cluster members. (iii) Number of structures of the cluster.

its fusion with the P450BM-3 domain. Another reason could be that SDA generates diffusional encounter complexes, not fully bound complexes.

We then used SDA to compare experimental data on binding affinities when residues on the proximal side of CYP2B4 were mutated. The interactions of CYP2B4 with both cytochrome b5 and cytochrome P450 reductase have been extensively studied by Waskell et al. [149]. Their mutagenesis data show that when one of R122, R126, R133, R422, K433 and R443 was mutated, the Kd of the complex was increased dramatically [149]. Negatively charged interface residues on CPR are E92, E142, D147 and D207 which are important for the interaction with on CYP2B4. For comparison, the total energies of the top three clusters of SDA were used. As shown in Table 5.2, with each of the single mutations, the computed energy was much higher (less favorable) than the wild type. Although having the same binding interface on CYP2B4, the orientation of CPR differs (Figures 5.4&5.3). There are far fewer encounter complexes recorded for the mutants than the wild type, meaning that the mutations make the association less favorable. This result also shows the possibility of using the total interaction energy from SDA to compare binding affinities of systems of P450s and CPR.

From our assessment and previous SDA usage, we believe that SDA together with molecular dynamics refinement is a method to study interactions of P450s and CPR. This will allow us to investigate interactions of different P450s and CPR, compare the binding affinities of different P450s and compare the binding affinities of P450s with and without ligands in the active site.

5.3.2 Interactions of CYP2B4 and CPR

By using SDA and MD refinement, we were able to generate complexes that are similar to the models built by Waskell et al. using mutagenesis data [149]. As

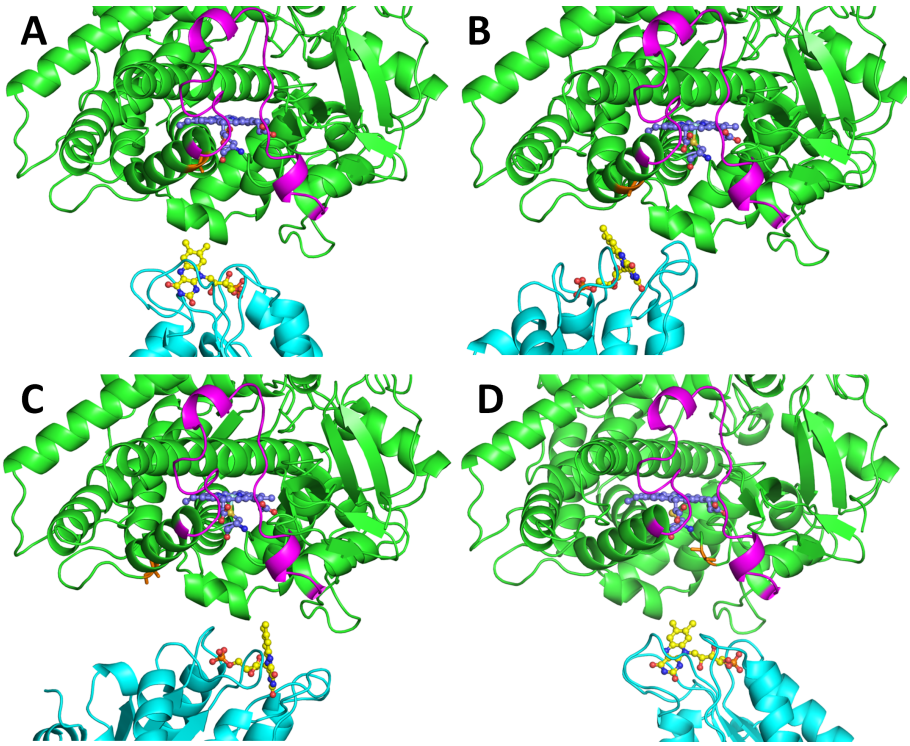

Figure 5.3: SDA recorded complexes of different mutants on the proximal side of CYP2B4: R122A: (A), R126A: (B), R133A: (C) and K433A (D). CYP2B4 is in green, CPR in cyan, the heme center in purple and FMN in yellow. The recorded complex of wild-type is shown in Figure 5.4. All structures shown are representative structures of the first cluster and the mutated residues are shown in orange.

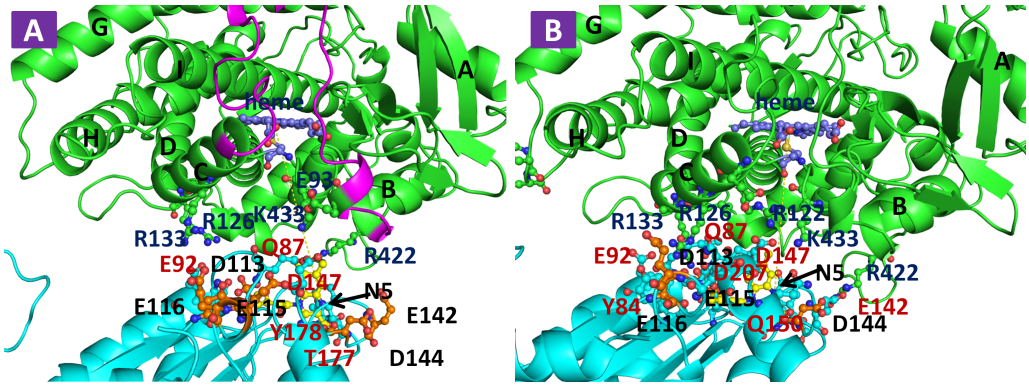

Figure 5.4: Complexes of CYP2B4 (green) and the human CPR (cyan) generated by SDA without (A) and with (B) MD refinement. The heme and FMN cofactors are show in purple and yellow, respectively. The B-C loop of CYP2B4 is colored in magenta. Interface residues on the proximal side of CYP2B4 are labeled and colored in dark blue and those of CPR in red. The interface residues (D113, E115, E116, E142, D144) of CPR which were proposed to be interface residues experimentally but are not identified in this study as interface residues are shown in orange and labeled in black. The distance between the FE atom of heme and the N5 atom (labeled) of FMN is 23.2 Å in (A) and 20.0 Å in (B).

shown in Figure 5.4, positively charged interface residues on CYP2B4 in our models are R122, R126, R133, R422 and K433, which is in very good agreement with the previously mentioned mutagenesis data. Not all the interface residues that were proposed (D113, E116, E142 and D144) are on the binding interface of CPR [149].

It should also be noticed that the distance between the electron transfer atoms, the iron (FE) atom of the heme cofactor and the nitrogen (N5) atom of the FMN cofactor, shortened from 23 Å to 20 Å during MD refinement. This shows that the SDA docked encounter complex is the favored encounter position for both proteins. The distance of 20 Å is close enough for electron transfer. Electrons can be transferred from FMN to intermediate residues and then to the heme center.

5.3.3 Interactions of CYP51s and CPR

Sterol 14α-demethylase (CYP51) plays an important role in sterol biosynthesis. It can be used as a target for antiparasitic drugs and thus it is crucial to understand differences between human and parasitic CYP51s to develop parasite specific inhibitors. Here, we docked the human CPR to both human and *T. brucei* CYP51s.

The average total interaction energy of the top three clusters of encounter complexes of human CYP51 and human CPR is -18.1 k_BT and that of *T. brucei* CYP51 and human CPR is -8.5 k_BT. The difference of the total interaction energy indicates that human CYP51 has higher affinity for human CPR than *T. brucei* CYP51. If the association of CYP51 and CPR is the rate limiting step of the catalytic reaction, this result can help to explain the observation by Lepesheva et al. [6] that the efficiency of reduction for human CYP51 is higher than that for *T. brucei* CYP51 when using rat CPR for electron transfer as shown in Figure 4 of ref. [6]. Rat CPR has 92% sequence identity to human CPR and the residues surrounding the FMN cofactor are well conserved. However, rat CPR has only 24% identity to *T. brucei* CPR and the interface residues for the P450s are not conserved. The reason for the higher efficiency of human CYP51 may be the faster association of human CYP51 indicated by the more positively charged proximal side than *T. brucei* CYP51 as shown in Figure 5.5.

Interface residues of *T. brucei* CYP51 and human CPR are shown in Figure 5.6. Only three positively charged residues are identified on the interfaces of CYP51 of both complexes in Figure 5.6.

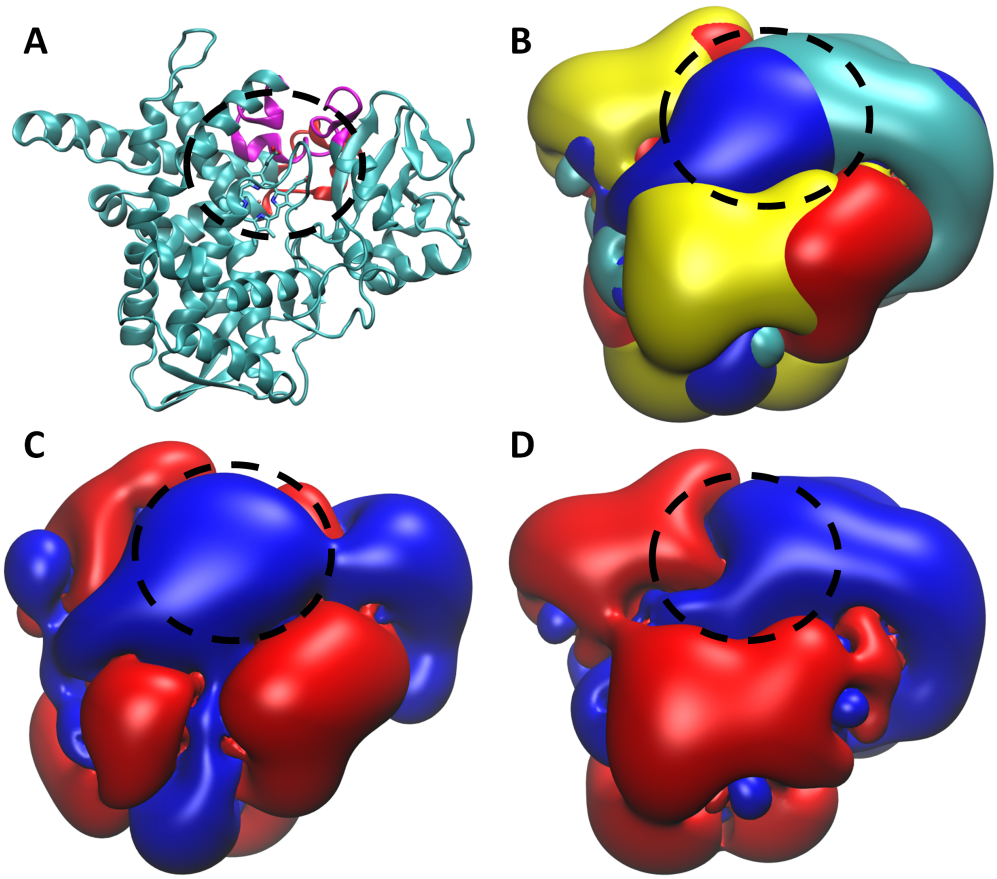

Figure 5.5: Comparison of electrostatic potentials of *T. brucei* and human CYP51. The approximate proximal side is labeled by dashed circles. (A) The structure of *T. brucei* CYP51 with B-C loop in magenta and F-G loop in red. (B) The comparison of electrostatics of the proximal side of *T. brucei* and human CYP51. For *T. brucei* CYP51, positively and negatively charged patches are in cyan and yellow, respectively. For human CYP51, those patches are in blue and red, respectively. (C) and (D) are electrostatics (-0.1 and 0.1 kcal/mol e) of human and *T. brucei* CYP51 with blue for positive and red for negative patches. The ionic strength was 150 mM at pH 7.

5.3.4 Interactions of Human Drug-metabolizing P450s and CPR

To study interactions of human drug-metabolizing CYPs with CPR, we used SDA to dock crystal structures of CYP1A2, CYP2A6, CYP2C9, CYP2D6, CYP2E1 and CYP3A4 to CPR. The trend of the energy can help to analyze affinities of P450s with CPR as shown previously. The electrostatic potential of P450s are shown in Figure 5.7. As shown in Figure 5.8, CYP1A2 has much more favorable energy in the SDA recorded encounter complexes with CPR and thus has much higher affinity to bind to CPR than CYP2B4. This is in good agreement with the experimental data by Kelley et al. [159], in which they showed that in a mixture of CYP2B4, CYP1A2 and CPR, CPR has higher affinity to bind to CYP1A2 than

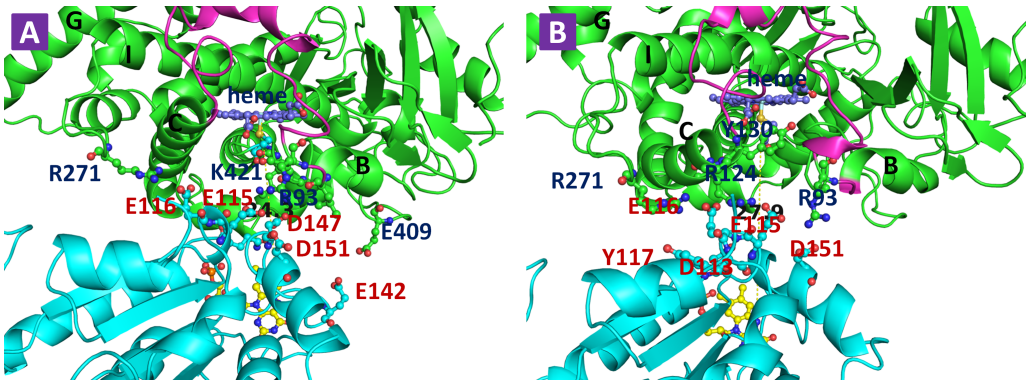

Figure 5.6: Complexes of *T. brucei* CYP51 (green) and the human CPR (cyan) generated by SDA without (A) and with (B) MD refinement. The heme and FMN cofactors are show in purple and yellow, respectively. The B-C loop of CYP51 is colored in magenta. Interface residues on the proximal side of CYP51 are labeled and colored in dark blue and those of CPR in red. The distance between the FE atom of heme and the N5 atom of FMN is 24.3 Å in (A) and 27.8 Å in (B).

CYP2B4. Amongst these P450s, CYP1A2, CYP2A6 and CYP3A4 have the most favorable binding energy and thus highest affinity. The ligand-bound crystal structures of CYP2D6 and CYP3A4 should have higher affinity to bind to CPR which indicates a higher affinity for P450s to bind CPR when substrates are bound to the active site. The ligand-bound crystal structure of CYP2C9 has lower affinity to CPR than the ligand-free structure, possibly because of the high inhibitory effect of flurbiprofen.

CYP2A6, CYP2D6 and CYP2E1 are chosen as representatives to investigate interactions between them and CPR. Figure 5.9 shows interactions of CPR with these three human drug-metabolizing P450s. The same as previously mentioned CYP2B4 and CYP51, they all bind to the FMN domain of CPR using the proximal side. Although CPR binds using the same interface, the orientation and distance to P450s differ when binding to different P450s. Figures 5.9A&B show the interface of CYP2A6 and CPR. The highest number of interface residues of these three P450s (Figure 5.9) was identified which is consistent with the lowest binding energy of CYP2A6 and CPR. The proximal side of CYP2A6 is highly positively charged and this can be the reason for its strong interaction with CPR. The interaction of CYP2D6 and CPR (Figures 5.9C&D) is similar to the model built by Sündermann et al. [165]. For CYP2E1, it is similar to CYP2A6, although the distance between heme and FMN increased after MD refinement.

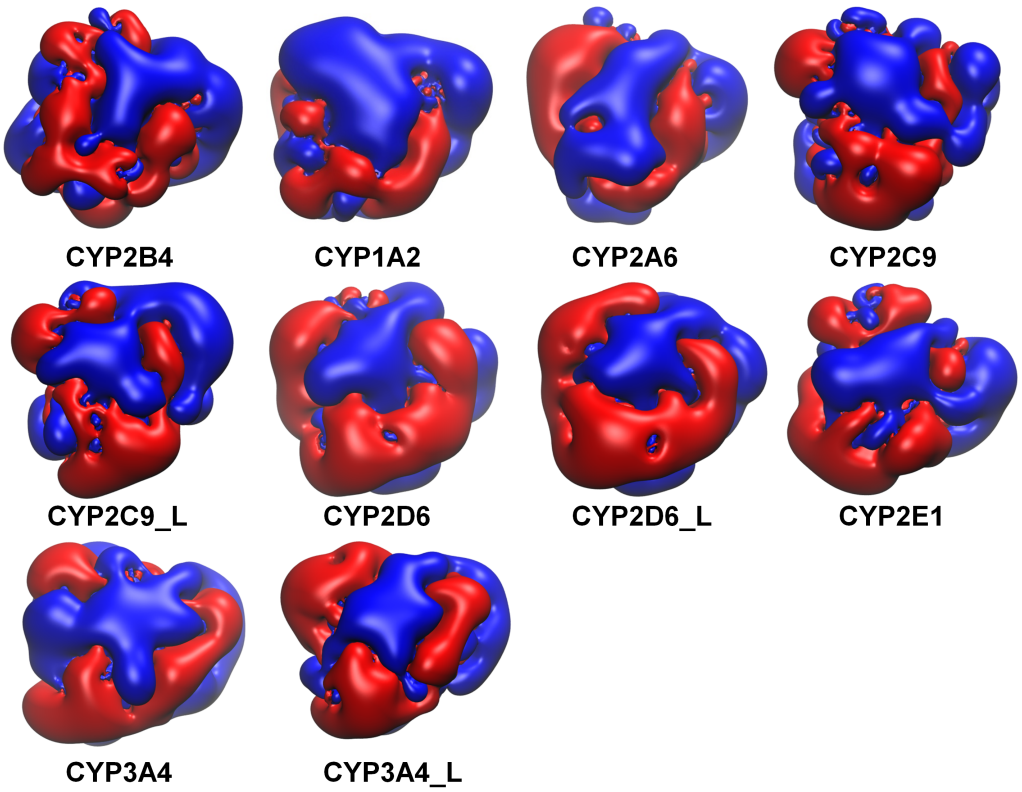

Figure 5.7: Electrostatic potential of the proximal side of P450s (-0.1 and 0.1 kcal/mol e) with positive charges in blue and negative charges in red. All P450s have the same view as in Figure 5.5A.

5.4 Conclusion

Interactions of different P450s with cytochrome P450 reductase were studied using the SDA 7 software, which uses Brownian dynamics to generate encounter complexes, and followed by molecular dynamics simulations in AMBER to refine encounter complexes of SDA. SDA was able to reproduce encounter complexes of P450s and their redox partners, which are in good agreement with crystal structures of corresponding complexes. In the generated encounter complexes of different P450s with CPR, binding interfaces are the positively charged proximal side of P450s and the negatively charged patch next to the FMN cofactor of CPR.

MD refinement introduces flexibility to encounter complexes and refines interactions of the two proteins. In approximate 50% of the P450s studied, after MD refinement, the distance between the heme center and the FMN cofactor shortened (Figure 5.9). The distance after MD refinement should be short enough for effective electron transfer from FMN to the heme center. However, in other cases, the

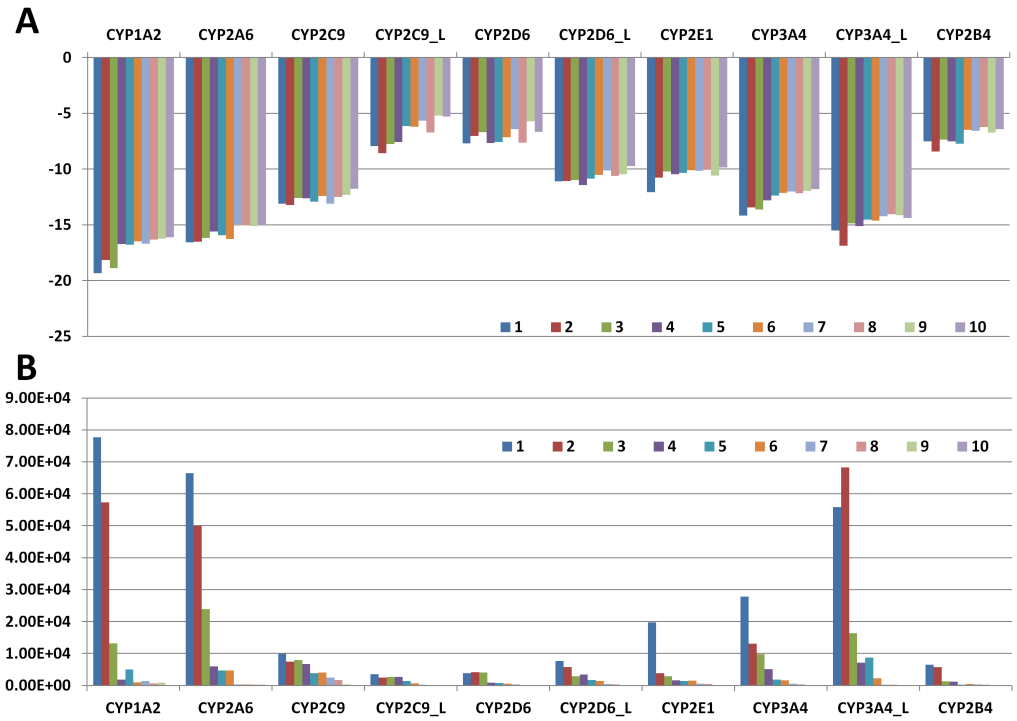

Figure 5.8: Average total energy and number of structures of all ten clusters of SDA docking.

distance between the two cofactors increased, which makes it difficult to transfer electrons. Therefore, the electron transfer may happen through a chain of residues on both CPR and P450s to transfer electrons from CPR to P450s.

Mutagenesis study suggested binding site residues of CPR (D113, E115, E116, E142 and D144) are mostly reproduced in different complexes. Our study also suggests a few important interface residues that were observed in most of the complexes, such as Q87, E92 and D147 on CPR. Further mutagenesis studies of human or rat CPR may help to define the importance of these residues.

The SDA interaction energy analysis of CYP1A2 and CYP2B4 agrees with experimental data and this analysis may help to compare binding affinities of different P450s to CPR. When a P450 has more negative SDA interaction energy for CPR, it should have high affinity for CPR. CYP1A2, CYP2A6 and CYP3A4 may have the highest affinity for CPR.

Our systematic investigation of P450s with CPR shows similarities and differences between complexes of P450s CPR. Our study may give new insights into interactions of P450s and their redox partners, help to understand the electron transfer process and thus to understand the mechanism of P450s.

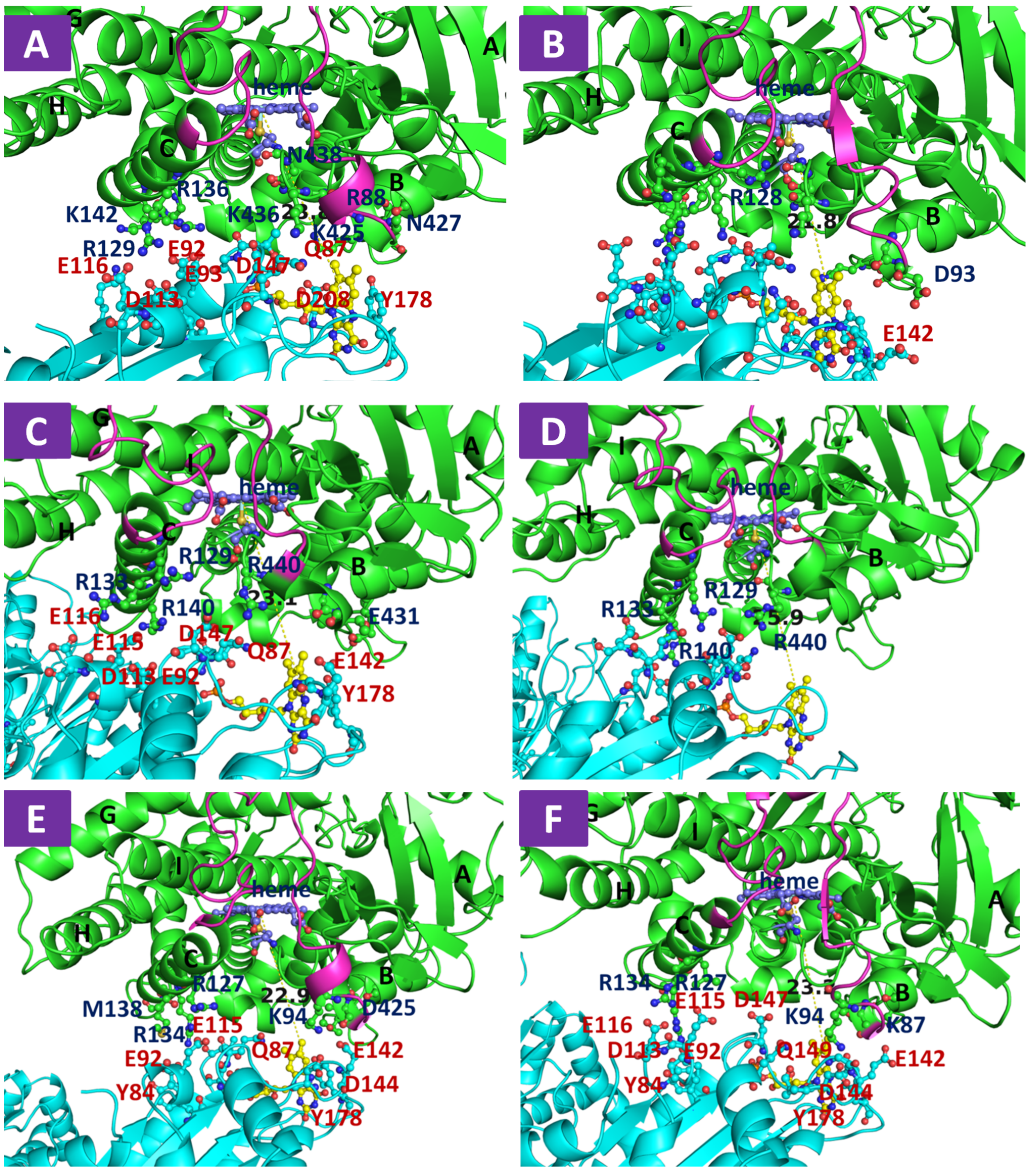

Figure 5.9: Complexes of human drug-metabolizing CYPs (green) and the human CPR (cyan) generated by SDA of CYP2A6 without (A) and with (B) MD refinement, CYP2D6 without (C) and with (D) MD refinement, and CYP2E1 without (E) and with (F) MD refinement. The heme and FMN cofactors are show in purple and yellow, respectively. The B-C loop of CYP51 is colored in magenta. Interface residues on the proximal side of CYPs are labeled and colored in dark blue and those of CPR in red. The distance between the "FE" atom of CYPs and the "N5" atom of the FMN cofactor is 23.8 Å in (A), 21.8 Å in (B), 23.1 Å in (C), 25.9 Å in (D), 22.9 Å in (E) and 23.2 Å in (F).

SDA generates diffusional encounter complexes. If large conformational changes are required for the electron transfer process, SDA and MD refinement will not allow to simulate such conformational changes. However, we believe that our study provides basic understanding of interactions between P450 and CPR.

6

Simulation of a membrane-bound model of *T. brucei* CYP51 and human CPR

6.1 Introduction

As discussed in Chapter 5, for many P450s in eukaryotes, the redox partner is cytochrome P450 reductase (CPR). CPR consists of three domains: the FMN domain which contains the flavin mononucleotide (FMN), the FAD domain with the flavin adenine dinucleotide (FAD), and the NADPH domain with the nicotinamide adenine dinucleotide phosphate (NADP). In eukaryotes, both proteins locate in the endoplasmic reticulum (ER) membrane anchored with a trans-membrane helix. The association of the globular domains of the two proteins is mainly driven by electrostatic interactions [147]. The positively charged proximal side of P450 interacts with the negatively charged FMN domain of the CPR. However, so far, no crystal structure has been solved of their complex, and little is known about their interactions with the membrane or possible conformational changes of the proteins upon binding or during the catalytic cycle.

Understanding the P450-CPR interaction in the membrane is of great interest. First of all, it can help to investigate the catalytic mechanism of P450s in the native membrane-bound form. It can give insights into the association and dissociation of the two proteins in the membrane environment. Furthermore, the bound complex can be used to study ligand access and egress tunnels in P450s which can be

influenced by the protein-protein interactions. Lastly, it can help to improve the prediction of drug metabolism by P450s and the design of drugs to bind to P450 drug targets.

Many studies have been carried out for the soluble forms of P450s and their redox partners as discussed in Chapter 5. However, very little is known about the structures of complexes of P450s and CPR in their native membrane-bound form, especially, due to the difficulties of doing experiments with membrane-bound proteins. Computational methods have been used to build models of different P450s in the membrane. These models are shown to be in good agreement with experimental data [29, 30, 24, 111]. With these computational methods, it is also possible to build models of membrane-bound forms of complexes of P450s and CPR.

A systematic multi-scale simulation protocol employing coarse-grain and atomic detail models has been established to model and simulate P450 in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer and this has been used to model human CYP2C9 in a POPC bilayer [29]. Other researchers have also recently built and simulated models of P450s in bilayers including for CYP2C9 [30], CYP3A4 [24, 111], CYP19 [110] and other human drug-metabolizing P450s [112]. Sündermann et al. [165] performed a simulation of a complex of CYP2D6 and the open form of CPR in a bilayer. This complex was however built manually and the time scale of the simulation (10 ns) was too short to investigate the dynamics of the equilibrated complex in the membrane.

Here, we present a new multi-scale simulation approach to model the ligand-bound *T. brucei* CYP51 and human CPR in a POPC bilayer. We chose *T. brucei* CYP51 because it is an important drug target for antifungal and antiparasitic drug design as discussed in Chapter 3 and 4. The human CPR was used because there is no crystal structure of *T. brucei* CYP51 and the human CPR has been shown to be able to transfer electrons to *T. brucei* CYP51 [144]. Studying the interaction of these two proteins can also help to design *T. brucei* CYP51 specific inhibitors. This approach combines Brownian dynamics with the Simulation of Diffusional Association (SDA 7) software to generate encounter complexes of CYP51 and CPR, and coarse-grain and all-atom simulations to equilibrate and simulate the complexes on the membrane.

Using this protocol, we built a model of CYP51 and CPR in complex on a POPC bilayer. After a simulation of more than 200 ns, a stable complex was

obtained. The distance between the heme center of CYP51 and the FMN cofactor (semiquinone form) of CPR shortened in the simulation from 21.3 Å to 19 Å, which is short enough for effective electron transfer. We proposed a possible electron transfer pathway, from the FMN cofactor, to the LYS421 residue, then to the heme-coordinating CYS422 residue, and finally to the heme center.

6.2 Methods

The newly developed protocol is shown in Figure 6.1. The procedure to use the SDA software to generate encounter complexes has been discussed in Chapter 5. The first cluster representative after MD refinement of the simulation discussed in Chapter 5.3.3 was used as the starting structure of the CYP51-CPR encounter complex. The membrane-bound model of CYP51 on a POPC bilayer was discussed previously in Chapter 3. The last snapshot of the simulation of membrane-bound form of *T. brucei* CYP51 as discussed in Chapter 3 was used as the other starting structure.

With the two initial structures, the CYP51 of the encounter complex was superimposed onto the model of membrane-bound CYP51. Missing N-terminus residues 1-68 of CPR including the trans-membrane helix (residues 23-43) and the missing loop between the trans-membrane helix and the FMN domain (residues 44-68) were modeled using Modeller and inserted into the membrane as described in [29].

The same protocol and parameters as described in Chapter 3.2.6 were used to simulate the system.

6.3 Results and Discussion

The membrane-bound model of the complex of *T. brucei* CYP51 and human CPR was simulated for 217.5 ns. The RMSD analysis of backbone atoms of CYP51, the FMN domain of CPR and the FAD domain of CPR is shown in Figure 6.2. The B-factor analysis of the last 112.5 ns (when the system fluctuated stably) of non-hydrogen atoms of the whole system (both CYP51 and CPR), CYP51, the FMN domain of CPR and the FAD domain of CPR is shown in Figure 6.3 and the radius of gyration analysis of the last 112.5 ns of those is shown in Figure 6.4.

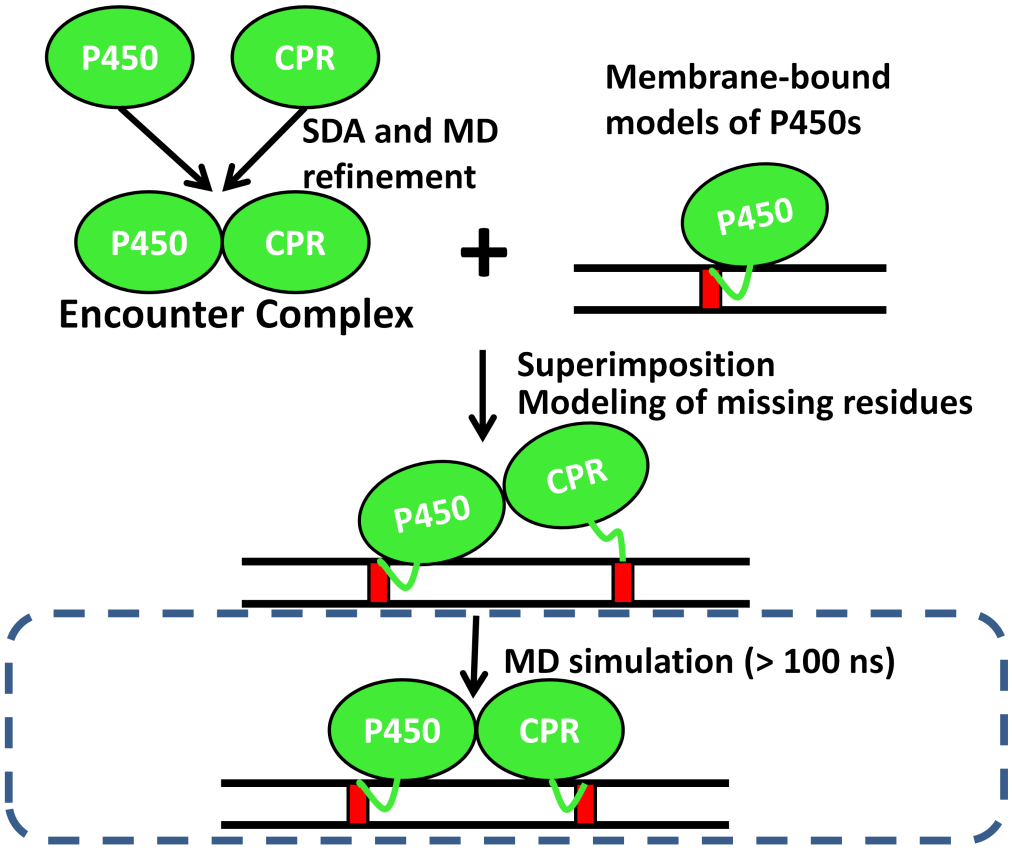

Figure 6.1: Protocol to build and simulate membrane-bound models of CYP51 and CPR.

6.3.1 The Complex in the Membrane

The starting structure and the equilibrated structure are shown in Figure 6.5. In Figure 6.5A, the membrane-bound *T. brucei* CYP51 was taken after MD simulation and the CPR was built onto it. In this model, the globular domain of CPR was relatively far away from the membrane, with the FMN domain interacting with the proximal side of CYP51 and the FAD domain extending away from the membrane. However, the FAD domain of CPR got much closer to the membrane in the simulation, although the distance between the center of mass of CPR and that of the membrane remained mostly similar during the simulation (Figure 6.6). The globular domain of CPR has a much larger distance to the membrane than that of CYP51 in the simulation. The reason can be that CPR is prohibited from getting much closer to the membrane because of its interaction with CYP51.

Figure 6.6 shows distance measurements. The distance between the center of mass of CYP51 and that of the membrane fluctuated stably during the simulation.

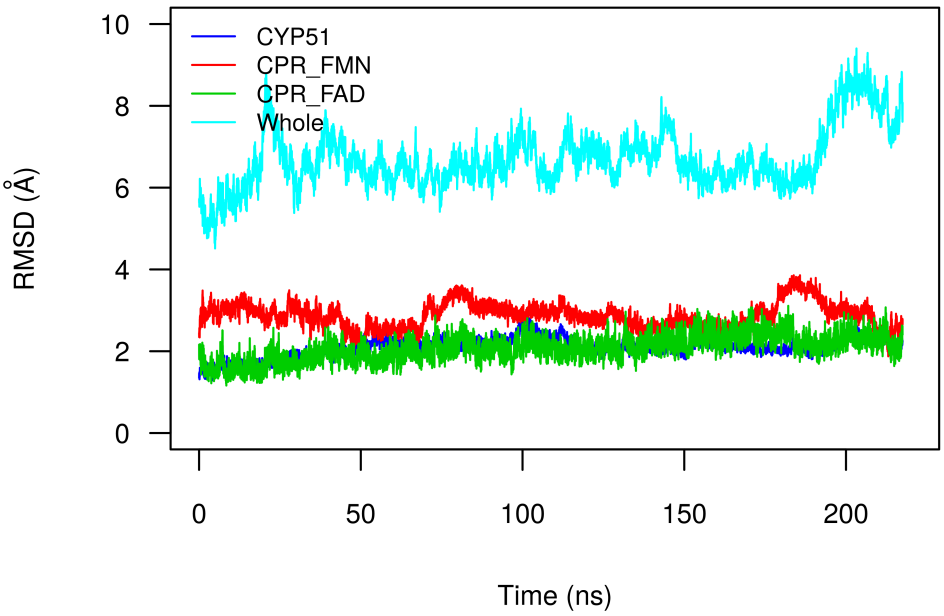

Figure 6.2: RMSD analysis of protein backbone atoms during the MD simulation of the membrane-bound model of *T. brucei* CYP51 and CPR. RMSDs Cα atoms of whole system (both proteins), CYP51, the FMN domain of CPR and the FAD domain of CPR are in cyan, blue, red and green.

Figure 6.7 shows the plot of the orientation of *T. brucei* CYP51 in the membrane. This orientation is similar to that of only *T. brucei* CYP51 in the membrane as discussed in Chapter 3 and Figure 3.5. This shows that in the simulation, CYP51 keeps its orientation in the membrane and embedded slightly deeper in the membrane in the complex form than only itself in the membrane (Figure 6.7). The heme tilt angle of *T. brucei* CYP51 in the simulation of the complex in membrane is 57.8 $\pm$ 5.0° which is in better agreement with experimental value for CYP3A4 than the membrane-bound model of only CYP51 as shown in Table 3.2.

The distance between CPR and membrane is much further than that between CYP51 and membrane. Davydov et al. have shown that P450s tend to form complexes and forming complexes can increase the affinity of P450s to bind to CPR [159, 173, 174]. The result of forming complexes with other P450s may change the orientation of P450s on the membrane, which make it easier to form complexes with CPR than monometic P450s.

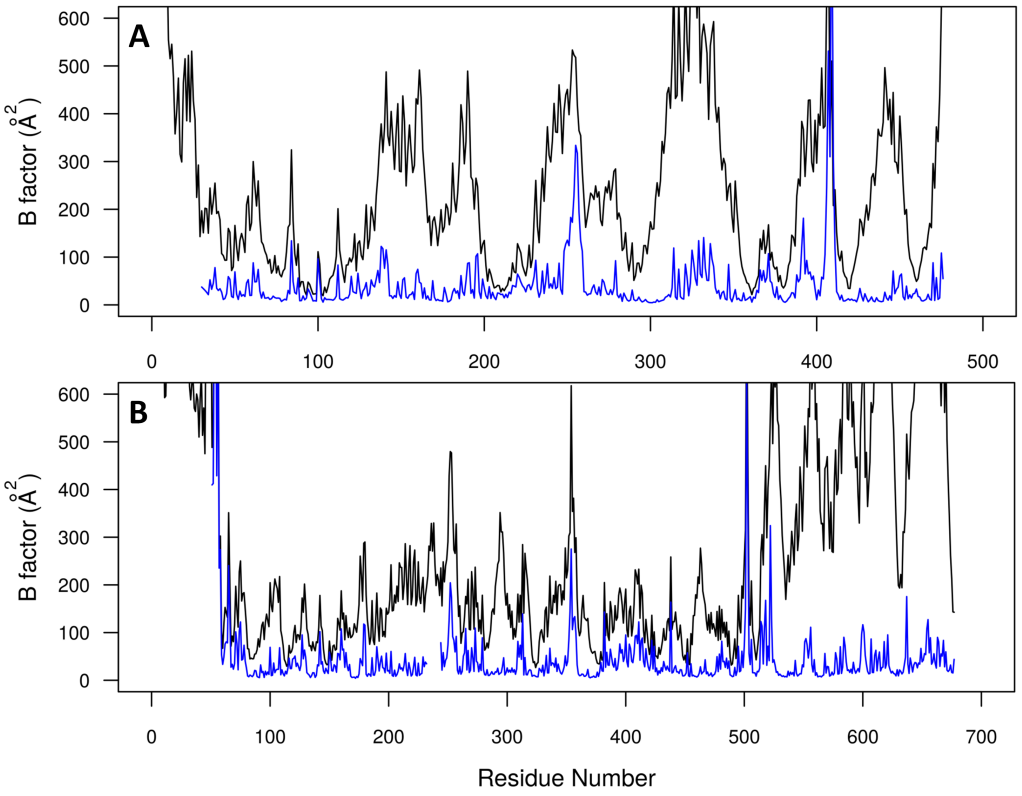

Figure 6.3: B-factor analysis of protein non-hydrogen atoms during the last 112.5 ns of the MD simulation of the membrane-bound model of *T. brucei* CYP51 and human CPR. B-factor of CYP51 when using the whole system as reference structure for superimposition (A, black) and CYP51 as reference structure for superimposition (A, blue) and that of CPR when using the whole system for superimposition (B, black) and, the FMN domain for superimposition and FAD domian for superimposition (B, blue). When using CYP51 as reference structure for superimposition, the modeled region (residues 1-29) was not computed. The FMN domain and the FAD domain of CPR were used as reference structures for superimposition to compute B-factors for the FMN domain and the FAD domain, respectively. The modeled region of CPR (residues 1-50 and 231-243) was not computed.

6.3.2 Comparison of Interface Residues with Experimental Data

There are more interface residues after 217.5 ns simulation than before the simulation as shown in Figures 5.5 and 6.8. In the simulation of the membrane-bound model of the complex, CYP51 and CPR got much closer mainly driven by electrostatic interactions. As shown in Figure 6.8, six positively charged residues (lysines and arginines) of CYP51 form strong electrostatic interactions with negatively charged residues (aspartic acids and glutamic acids) of CPR. All five negatively charged residues of CPR (D113, E115, E116, E142 and D144), which are proposed to be important for interactions of P450 and CPR, are interacting with residues of CYP51. The six positively charged residues of CYP51 used to interact with CPR lo-

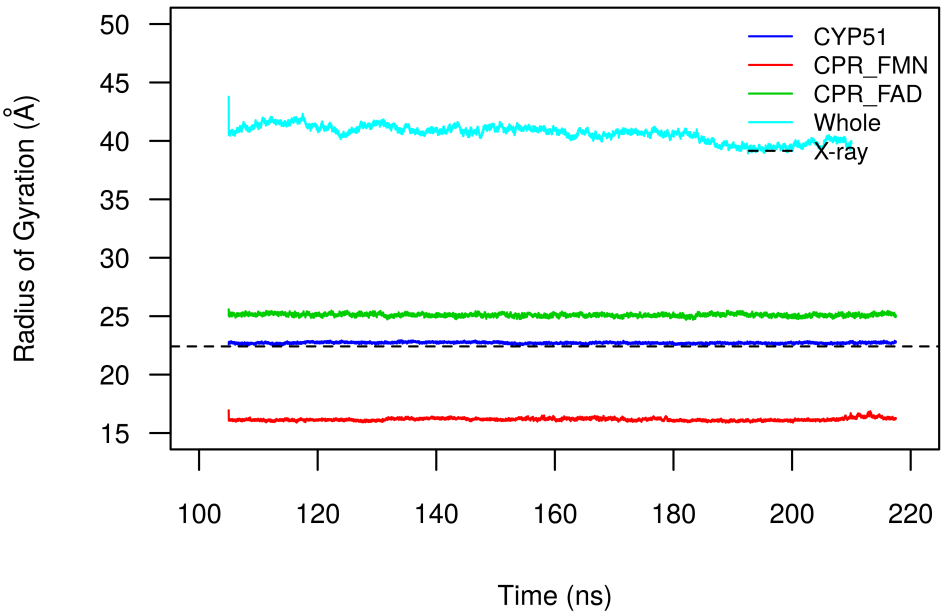

Figure 6.4: Radius of gyration analysis of protein non-hydrogen atoms during the last 112.5 ns of the MD simulation of the membrane-bound model of *T. brucei* CYP51 and CPR. The radius of gyration was computed for the whole system (cyan), CYP51 (blue) and, the FMN domain (red), the FAD domain (green) of CPR and the crystal structure of CYP51 (PDB ID: 3G1Q) (black).

cate on different secondary structures of CYP51. They are not conserved in P450s, although the positively charged proximal side of P450s is conserved.

6.3.3 Electron transfer from CPR to CYP51

The distance between the FE atom of the heme center and the N5 atom of the FMN cofactor of CPR decreased during the simulation. The distance changed from the start of the simulation with ~22 Å to the end of the simulation with ~19 Å. This distance is not as close as the distance of 14.1 Å obtained from MD simulations of the crystal structure of the complex of P450-BM3 and the FMN domain [175], but it should be short enough for effective electron transfer.

The LYS421 residue (next to the heme-coordinating CYS422 residue) directly interacts with the FMN cofactor as shown in Figure 6.8. It may form a short electron transfer pathway with CYS422 to transfer electrons from the FMN cofactor to

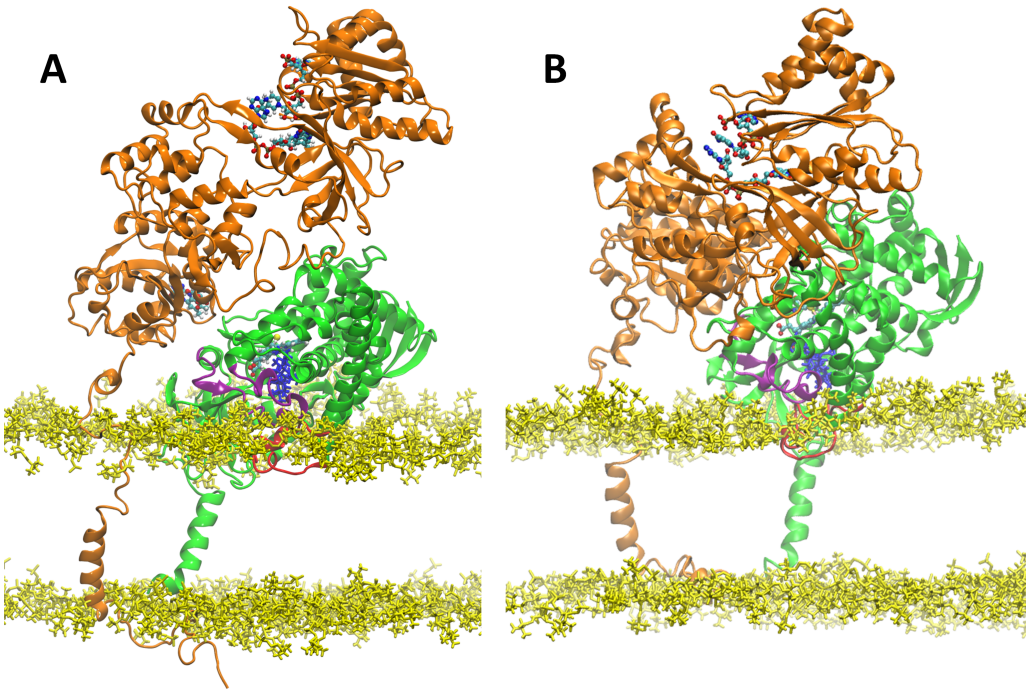

Figure 6.5: Structure of the membrane-bound model of *T. brucei* CYP51 and human CPR before (A) and after 142.5 ns MD simulation (B).

the heme. Thus this possible pathway is FMN-LYS421-CYS422-heme but further computation and mutagenesis study are required to confirm this.

6.4 Conclusion and Outlook

A model of membrane-bound form of *T. brucei* CYP51 and human CPR was built using a new multi-scale approach. Based on the model of membrane-bound *T. brucei* CYP51, the membrane-bound CPR was built upon it. The 217.5 ns simulation of the model shows that the interaction of CYP51 and CPR is stable in the membrane-bound form. The stability of the complex was driven by the electrostatic interactions as inferred by the fact that the interface residues of CYP51 are mostly positively charged lysine and arginine residues and the interface residues of CPR are mostly aspartic acid and glutamic acid residues. The distance between heme and FMN shortened in the simulation and a possible electron transfer pathway of FMN-LYS421-CYS422-heme was proposed.

The proposed protocol gives the possibility to build models of membrane-bound forms of different P450s and CPR. Having membrane-bound models of P450s and CPR can help to compare P450s, to understand selectivity of CPR upon

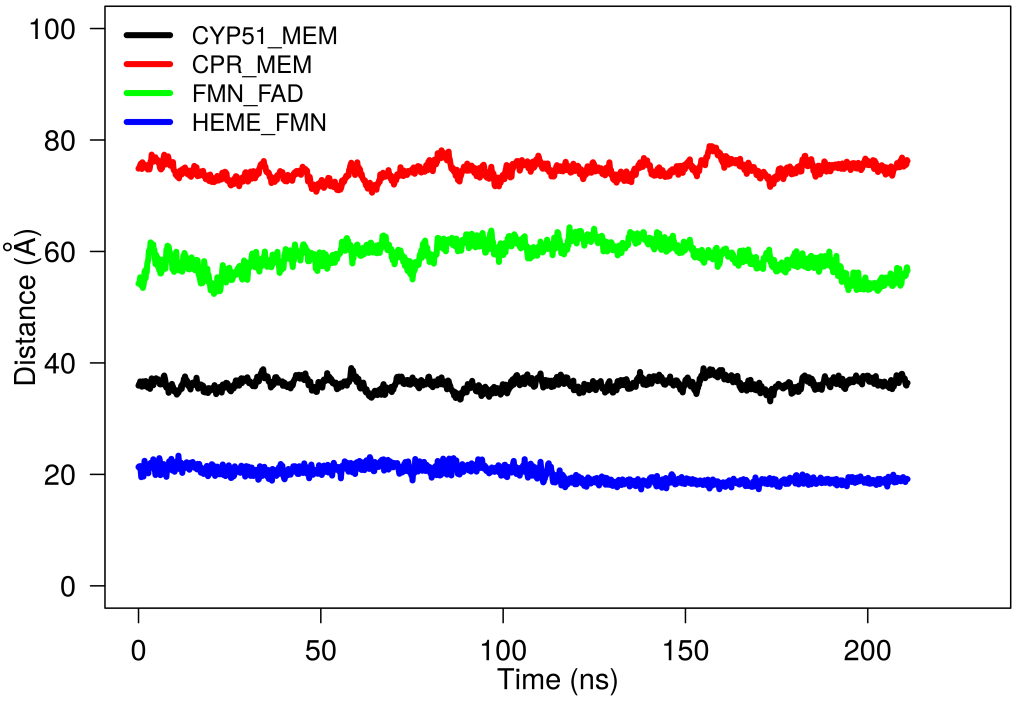

Figure 6.6: Analysis of distances between different domains of the membrane-bound model of CYP51 and CPR during the simulation. These distances are distance between CPR and membrane in red, that between the FMN domain of CPR and the FAD domain of CPR in green, that between CYP51 and membrane in black, and that between the FE atom of the heme center of CYP51 and the N5 atom of the FMN cofactor of CPR in blue.

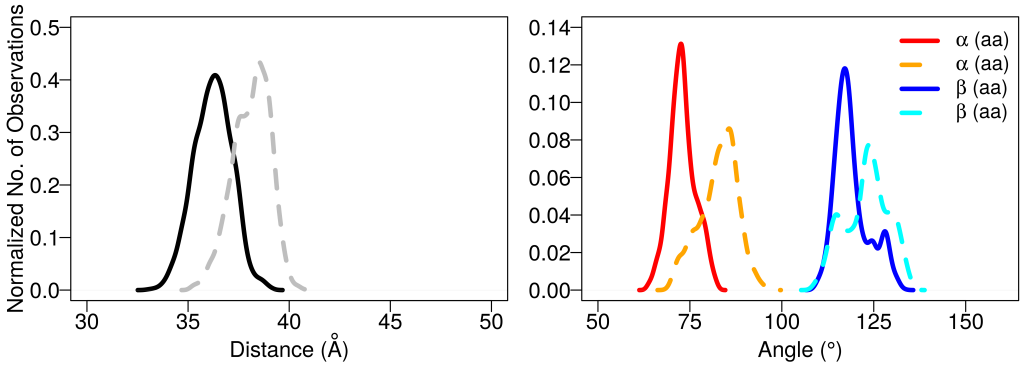

Figure 6.7: The normalized distribution of the three measured parameters (d, α and β) of *T. brucei* CYP51 in the simulation of the membrane-bound model of the complex of *T. brucei* CYP51 and human CPR in the membrane. The parameters are defined as in Chapter 3.2.7. Density plot of the distance between the center of mass of CYP51 and that of the membrane in the complex form with CPR in black and that of only CYP51 in the membrane in grey (A) and density plot of angle α of CYP51 in complex form (red) and that of only CYP51 in the membrane (orange) and, angle β of CYP51 in complex form (green) and that of only CYP51 in the membrane (cyan). The data of distance, α angle and β angle of only CYP51 in the membrane are taken from the simulation of the membrane-bound model of *T. brucei* CYP51 discussed in Chapter 3.

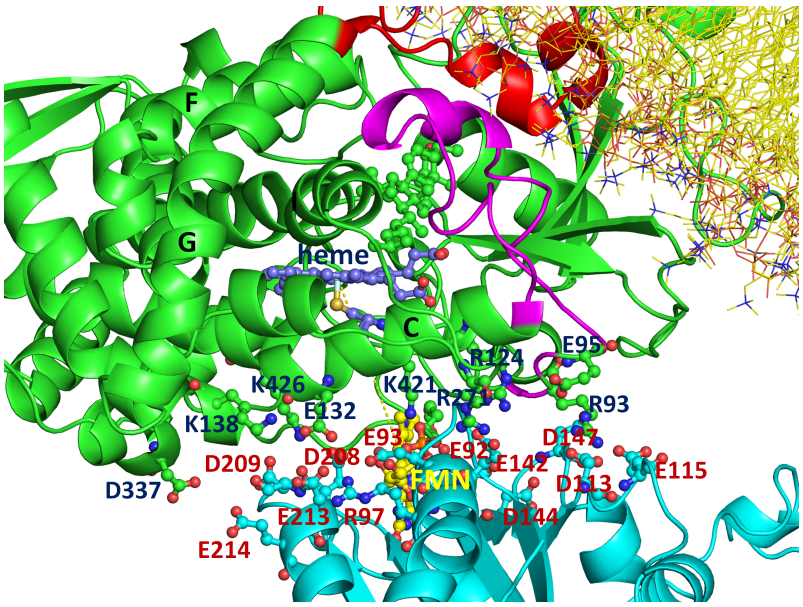

Figure 6.8: Interface residues of *T. brucei* CYP51 and human CPR after MD simulation of the membrane-bound form of the complex. CYP51 is in green and CPR in cyan. The B-C loop of CYP51 is in magenta and F-G loop in red. The heme center is in purple and the FMN cofactor in yellow. Residues that are within 7.5 Å of the other protein are selected as interface residues. Only the positively (lysine and arginine) and negatively charged (aspartic acid and glutamic acid) residues are shown and labeled. Interface residues of CYP51 are labeled in dark blue and those of CPR in dark red. The distance between the FE atom of the heme and N5 atom of FMN is 18.4 Å.

different P450s and thus to aid drug design. The model of the membrane-bound form of CYP51 and CPR is a fundamental model to understand behaviors of both CYP51 and CPR in complex form on a lipid bilayer, which is the native environment. This model may help to design drugs against parasitic CYP51s, taken their interaction with CPR into consideration.

7

webSDA, a Web Server for SDA 7

7.1 Introduction

Macromolecular interactions, such as protein-protein interactions and protein-DNA interactions, are key to the function of biological systems. Studying macromolecular diffusional association computationally can give insights into biological interactions. SDA (Simulation of Diffusional Association) is a Brownian dynamics simulation software which can be used to perform macromolecular docking, to compute bimolecular association rate constants, and to simulate molecular diffusion for systems of many macromolecules [73, 71, 85]. Intermolecular forces are computed from atomic-detail molecular structures. SDA has been used to generate diffusional encounter complexes for a wide range of macromolecular complexes, including those of barnase and barstar [71, 83], cytochrome P450 enzymes and their redox protein partners [147, 79, 170], methionine aminopeptidase and the ribosome [81], and the linker histone H5 and the nucleosome [80]. Moreover, SDA has been used to calculate association rate constants [73, 71, 176, 177] and simulate the diffusive motion of hundreds of macromolecules in solution [85, 87, 178, 88, 86]. All these functions of SDA have been extensively validated against experimental data. Although SDA is made available for standalone use, it cannot be used easily by non-experts because of its multitude of different parameters and options, many of which are interdependent. Setting these parameters correctly and understanding their function demands a good understanding of Brownian dynamics simulations and the structure of the SDA software.

webSDA is a web server which aims to eliminate these obstacles and make SDA's functionalities available to a broader audience. webSDA supports the gen-

eration of diffusional encounter complexes ("SDA docking"), the calculation of biomolecular diffusional association rate constants ("SDA association") and the simulation of the diffusive motion of several hundred macromolecules ("SDA multiple molecules"). The flexibility of the solutes can be accounted for by providing several different conformations for one of the solutes ("SDA docking" and "SDA association") or each of the solutes ("SDA multiple molecules"). It allows users to set up and run SDA simulations. webSDA automates the preparation of input files by generating input parameters automatically. If one parameter is used in multiple input files, it only needs to be defined once. Input files generated by webSDA can also be downloaded for use with the standalone version of SDA.

7.2 webSDA Workflow

The workflow of webSDA is shown in Figure 7.1.

Step 0: To begin a simulation, users may enter the PDB identifiers of the macromolecules that they wish to simulate or upload coordinate files for the macromolecules in PDB or PQR format. PQR format files contain the atomic coordinates, formatted as in PDB files from the Protein Data Bank (`www.rcsb.org/pdb`), with partial atomic charges (Q) and atomic radii (R) instead of B-factors and occupancies. If PDB files are used, PQR files will be generated using the PDB2PQR software, version 2.0.0 [117], with the AMBER force field and default parameters. Users can also generate their own PQR files using the PDB2PQR web server [167]. After uploading coordinate files, users select one of the SDA methods: "SDA docking", "SDA association" or "SDA multiple molecules". Users then follow a sequence of steps to set up and run an SDA job, guided by hints on the webSDA website. As an example, the following steps show how to run an "SDA docking" simulation.

Step 1: Select target and mobile solutes: Users need to choose a target solute and a mobile solute for the "SDA docking" job.

Step 2: Modify reaction criteria and effective charge sites: Reaction criteria are distances between defined points in the solute molecules that are required to be satisfied for an encounter complex to be recorded in an SDA simulation. The default reaction criterion for "SDA docking" is predefined for the centers of geometry of the solutes and requires that they approach within a distance equal to the sum of the maximum extensions of the two solutes plus 12 Å. The reaction criteria can be changed in the web interface and can be used to impose additional

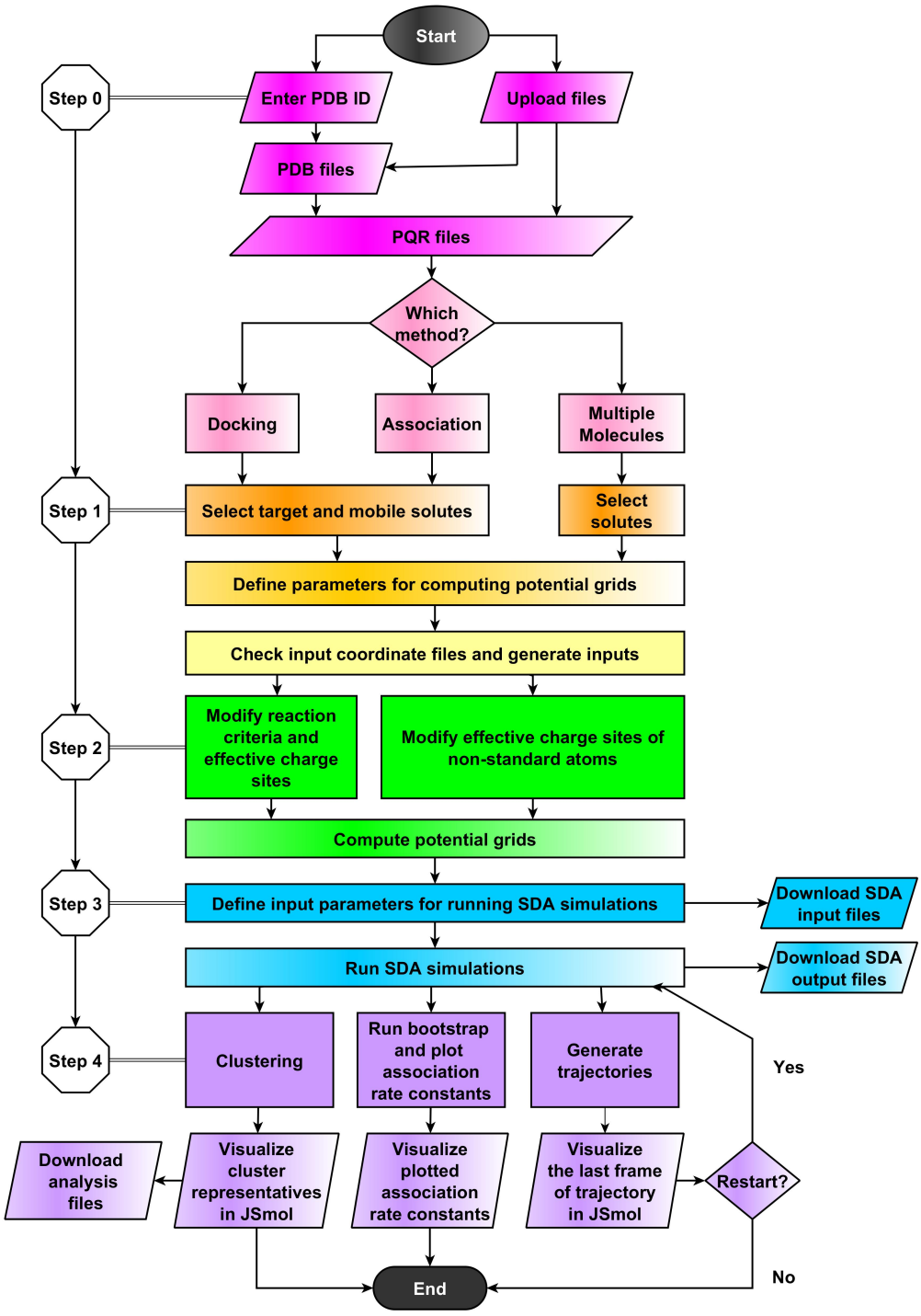

Figure 7.1: Workflow of webSDA

distance constraints which can be derived, e.g. from bioinformatics analysis or experimental data, such as FRET and mutagenesis. Effective charges are used in combination with electrostatic potential grids to approximate the electrostatic

interactions between molecules [75]. Users can modify the effective charge sites for non-standard atoms, such as found in low-molecular weight compounds, in the web interface.

Step 3: Define input parameters for running SDA simulations: The automatically generated SDA parameters are given to users. Users can modify the SDA input parameters and run SDA simulations on the server. webSDA checks the input parameters that users have changed and helps users to set correct parameters. They may upload their own files (SDA input file, reaction criteria file and effective charge sites file) to run their SDA simulations on the web server. The web server can be used to run short jobs. The SDA input file can also be generated and downloaded together with the other files to run with the standalone version of the SDA software, which is also freely available.

Step 4: Clustering: For "SDA docking", the recorded diffusional encounter complexes are clustered and the cluster representatives can be visualized with JSmol on the web server or downloaded as described below.

7.3 Output

Besides the option to download output files, webSDA offers a convenient visualization of the results. An example is given in Figure 7.2. For the results of "SDA docking", JSmol is used for visualization of the cluster representatives of the diffusional encounter complexes. This provides users with an easy way to view encounter complexes and decide which complexes to download for further analysis. For "SDA association", the calculated association rate constants are plotted using the R language, enabling users to visualize calculated association rate constants immediately. For "SDA multiple molecules", the JSmol plug-in is employed to enable users to visualize the final snapshot of their simulations.

webSDA assigns each user a universally unique identifier (UUID). This enables users to revisit all projects that they have run in the "My projects" page. There is a link associated with this page and users can send this link to colleagues to share projects.

7.4 Examples of webSDA Usage

webSDA gives results consistent with the standalone version of SDA, which has been extensively validated against experimental data [73, 81, 87]. webSDA provides example cases for each method. In Figure 7.2, we illustrate the performance of webSDA for the simulation of the association of barnase and barstar (PDB ID: 1BRS) and the simulation of many barnase molecules. With the "SDA docking" method, cluster representatives of the docked barnase-barstar complexes in good agreement with the crystal structure of the complex and results obtained by Motiejunas et al. [83] are obtained.

It should be kept in mind that "SDA docking" uses Brownian dynamics simulations to generate diffusional encounter complexes rather than fully docked complexes. The encounter complexes can be refined by molecular dynamics simulation to generate fully bound complexes, as shown in [83], but currently webSDA does not support running molecular dynamics simulation.

The "SDA association" function is another important function of webSDA. For the barnase and barstar complex at an ionic strength of 50 mM and pH 7, the association rate constant computed with "SDA association" is 4.1 x $10^8 M^{-1} s^{-1}$ (averaged from three webSDA runs: 4.0 x $10^8 M^{-1} s^{-1}$, 4.0 x $10^8 M^{-1} s^{-1}$ and 4.3 x $10^8 M^{-1} s^{-1}$), which is comparable with experimental data (3 x $10^8 M^{-1} s^{-1}$) (18) (Fig. 2) and the simulation results reported by Gabdoulline and Wade [73].

The "SDA multiple molecules" function of webSDA can be used to perform Brownian dynamics simulations of many molecules. This function of SDA has been used to study the behavior of proteins in crowded environments and is shown to be able to reproduce experimental data [85, 87, 178, 88, 86]. The long-time self-diffusion coefficients, second virial coefficients and structure factors, calculated by "DA multiple molecules", agree with experimental values.

There are several webservers [179, 180, 181, 182, 183, 184] providing similar functionality to "SDA docking" but they are often limited to proteins with standard amino acid residues. "SDA docking" supports protein, DNA and RNA, and small molecules, such as a cofactor, can be treated as part of a macromolecular structure. For the calculation of association rate constants, the TransComp server (`http://pipe.sc.fsu.edu/transcomp/`) [185] computes association rate constants for protein-protein and protein-RNA complexes. "SDA multiple molecules" is, to the best of our knowledge, the first web server for performing simulations of hundreds

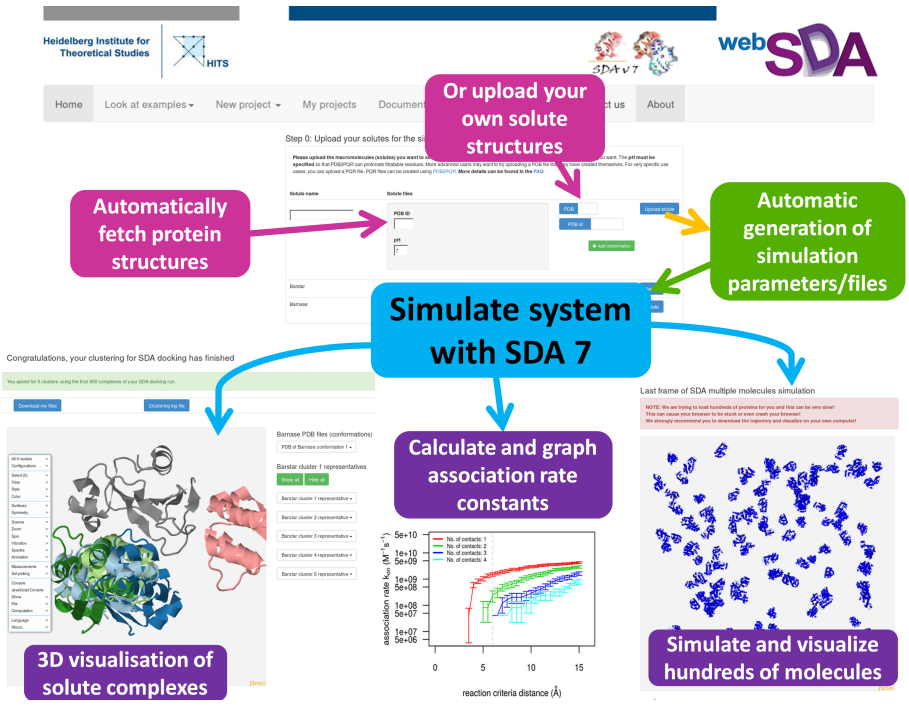

Figure 7.2: Screenshots of webSDA showing key features of webSDA

to thousands of macromolecules, which can be used to study concentrated protein solutions, and to investigate macromolecular crowding, or other phenomena involving large numbers of macromolecules.

7.5 Technical Overview

The webSDA interface uses Javascript and HTML5 to control user interface (UI) elements. The server uses the Play framework v. 2 and Apache commons.exec to manage user sessions and server resources. PDB2PQR version 2.0.0 is used to convert PDB files to PQR files [117, 167]. BioJava is used to process PDB ID and download PDB files. Python scripts (Biopython) are used to check PQR files and generate input parameters. SDA 7 (`http://mcm.h-its.org/mcm`) is used to process PQR files, to generate grids (along with the Adaptive Poisson-Boltzmann Software [76] to compute electrostatic potential grids), and to perform simulations.

7.6 Discussion and Outlook

webSDA is a web server to perform Brownian dynamics simulations of biomacromolecules. "SDA docking" allows users to generate diffusional encounter complexes of two solutes (proteins, DNA or RNA). "SDA association" can be used to compute the bimolecular association rate constant for a specified complex and can be combined with "SDA docking" when the structure of the complex is unknown. "SDA multiple molecules" is helpful for studying the diffusive behavior of hundreds of molecules.

webSDA automates the widely used SDA software, generates input parameters and files automatically, and offers user-friendly visualization of output results. webSDA is useful for beginners learning how to perform Brownian dynamics simulations as well as for experts preparing input files for running simulations. Simulation times and file sizes are limited on the webSDA server so webSDA allows the generated input files to be downloaded for carrying out production runs with SDA on local or high performance computing resources. We believe that webSDA will be an easy-to-use tool for users who are interested in using Brownian dynamics simulations to study biological systems. Currently webSDA does not support all the functions of SDA. However, webSDA has a modular structure that will facilitate the implementation of further methods and features in the future.

8

Conclusion

In this thesis, a multiscale computational study was performed for *Trypanosoma brucei* CYP51, human CYP51, interactions of P450 and CPR and the membrane-bound model of the complex of *T. brucei* CYP51 and human CPR. A webserver built for the Simulation of Diffusional Association (SDA) software was also introduced.

Using all-atom molecular dynamics simulations, the dynamics of soluble forms of *T. brucei* CYP51 was compared with those of human CYP51, human CYP2C9 and human CYP2E1. The binding site residues of both *T. brucei* CYP51 and human CYP51 are less mobile than those of CYP2C9 and CYP2E1. Whereas the tunnel entrance residues are more mobile in CYP51 than in CYP2C9 and CYP2E1. The difference between the active site residues of *T. brucei* and human CYP51 may be the key to design *T. brucei* specific inhibitors. A model of the membrane-bound form of *T. brucei* CYP51 was built using a combination of coarse-grained and all-atom simulations. Coarse-grained simulations were used to define the predominant orientation of *T. brucei* CYP51 in the membrane and this orientation was used as a starting point for the subsequent all-atom simulations. The model of membrane-bound CYP51 showed similar dynamics to the soluble model, whereas the F-G loop region which is buried in the membrane showed less mobility than the soluble model because of its interactions with lipids.

The ligand tunnels in the model of soluble *T. brucei* CYP51, the model of membrane-bound *T. brucei* CYP51 and that of soluble human CYP51 were also studied using random acceleration molecular dynamics (RAMD). Tunnel 2f serves as the predominant ligand tunnel in all three models. For *T. brucei* CYP51, when CYP51 is in the membrane, it uses tunnel 1 and tunnel S more often than in the soluble simulations. In the model of membrane-bound *T. brucei* CYP51, tunnel

2f leads to the membrane but ligands were able to exit in the membrane-bound model with a similar exit time to the soluble model. Human CYP51 uses only tunnel 1 as the second predominant ligand tunnel and exit times of ligands from the active site of human CYP51 is shorter than those of *T. brucei* CYP51. The difference between the use of ligand tunnels in *T. brucei* and human CYP51 may due to the differences in the tunnel entrance residues and the ligand tunnel gating residues.

The interaction of P450 and cytochrome P450 reductase (CPR) was studied using Brownian dynamics simulations and molecular dynamics simulations. Complexes of soluble forms of different P450s, including CYP51, CYP2B4, CYP1A2, CYP2A6, CYP2C9, CYP2D6 and CYP2E1 with CPR were investigated. P450s bind to CPR using a similar interface, the positively charged proximal side, but the orientation of CPR when binding to different P450s is different. The SDA total interaction energy is found to be useful to compare the binding affinity of P450s to CPR.

A model of the membrane-bound complex of *T. brucei* CYP51 and CPR was built and simulated. In the simulation of over 200 ns, CYP51 kept its orientation in the membrane but embedded slightly deeper into the membrane than in the simulation of only CYP51 in the membrane. The distance between the "FE" atom of the heme cofactor of CYP51 and the "N5" atom of the FMN cofactor of CPR shortened to ca. 19 Å. The protocol to build models of membrane-bound complexes of P450s and CPR can be applied to model other complexes of P450s and CPR. Our model of the membrane-bound complex of *T. brucei* CYP51 and CPR gives a basis for understanding the interactions of the two proteins in the membrane, which gives insights into the dynamics of the complex in their native membrane-bound environment.

A webserver (webSDA) was built for the Simulation of Diffusional Association (SDA 7) software package. SDA 7 was used to study the interaction between P450 and CPR. webSDA was built to improve the user-friendliness of SDA 7. The webserver automates the procedure to prepare and run SDA 7 and should be useful for beginners who want to perform Brownian dynamics simulations.

8.1 Limitations and Future Directions

8.1.1 Simulations of CYP51

The computational study of *T. brucei* CYP51 performed in this thesis gives a basis for understanding the dynamics of *T. brucei* CYP51. This simulations also provide structures of *T. brucei* CYP51 which can be used to design anti-parasitic drugs. However, the simulation were performed for less than 100 ns and this was not long enough to sample large conformational changes of *T. brucei* CYP51. The comparison of *T. brucei* CYP51 and human CYP51 shows some differences between the two proteins, mainly the difference in binding cavity residues and tunnel entrance residues. This information can be used in the design of parasite specific inhibitors. Further docking studies are required to design these inhibitors and subsequent experiments are required to test the computationally designed inhibitors.

The POPC bilayer was used in the simulations of models of membrane-bound *T. brucei* CYP51. However, biological membranes are not composed of homogenous lipids, but heterogenous lipids. Simulations with heterogeneous membranes would provide insights into the dynamics of CYP51 in biological membranes.

8.1.2 Simulations of Ligand Tunnels

The study of ligand tunnels shows how ligands can exit from the active site of both *T. brucei* and human CYP51. However, a force was applied to accelerate the exit of ligands which may introduce artifacts to the system. Although ligand exit tunnels can give insights into ligand entrance tunnel, simulations of ligand entrance are required to understand how ligands can enter the active site of P450 and ligand specificity of P450.

8.1.3 Simulations of Interaction of P450 and CPR

Simulations of P450 and CPR show the possible binding interfaces in the two proteins. However, these simulations were performed with crystal structures of individual (unbound) proteins and only relatively small conformational changes can be sampled in simulations. Diffusional encounter complexes were generated in these simulations and long MD simulations are required to equilibrate these systems to generate bound complexes. Further simulations of the electron transfer

process are required to simulate the electron transfer process and validate models of the complexes of P450 and CPR generated in this thesis.

The simulation of a model of the membrane-bound complex of *T. brucei* CYP51 and human CPR provides a basis for the understanding interaction of the two proteins on a lipid bilayer. However, there is not much experimental data about the structure of the complex and it is difficult to validate the model. The procedure can also be used to study P450-P450 interactions and generate structures of P450-P450 complexes. Mutagenesis and crystallography experiments can be performed to validate the model and these experimental data will give insights into the interaction of the two proteins on the membrane.

8.1.4 webSDA

The web interface for SDA 7 was developed to improve the user-friendliness of SDA. However, not all parameters of SDA 7 can be defined and not all functions of SDA 7 are supported by webSDA. Further development of webSDA is required to support parameters and functions of SDA 7 that are not yet supported by webSDA.

Bibliography

[1] P. B. Danielson, "The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans," *Current Drug Metabolism*, vol. 3, pp. 1389–2002, Sept. 2002.

[2] X. Yu, V. Cojocaru, and R. C. Wade, "Conformational diversity and ligand tunnels of mammalian cytochrome P450s," *Biotechnology and Applied Biochemistry*, vol. 60, pp. 134–145, Jan. 2013.

[3] T. Omura and R. Sato, "The carbon monoxide-binding pigment of liver microsomes. I. evidence for its hemoprotein nature.," *The Journal of biological chemistry*, vol. 239, pp. 2370–8, July 1964.

[4] D. R. Nelson, D. C. Zeldin, S. M. G. Hoffman, L. J. Maltais, H. M. Wain, and D. W. Nebert, "Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants.," *Pharmacogenetics*, vol. 14, pp. 1–18, Jan. 2004.

[5] U. M. Zanger and M. Schwab, "Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation.," *Pharmacology & therapeutics*, vol. 138, pp. 103–41, Apr. 2013.

[6] G. I. Lepesheva and M. R. Waterman, "CYP51–the omnipotent P450.," *Molecular and cellular endocrinology*, vol. 215, pp. 165–70, Feb. 2004.

[7] G. I. Lepesheva, H.-W. Park, T. Y. Hargrove, B. Vanhollebeke, Z. Wawrzak, J. M. Harp, M. Sundaramoorthy, W. D. Nes, E. Pays, M. Chaudhuri, F. Villalta, and M. R. Waterman, "Crystal structures of Trypanosoma brucei sterol

14alpha-demethylase and implications for selective treatment of human infections.," *The Journal of biological chemistry*, vol. 285, pp. 1773–80, Jan. 2010.

[8] G. Mustafa, X. Yu, and R. C. Wade, "Structure and Dynamics of Human Drug-Metabolizing Cytochrome P450 Enzymes," in *Drug Metabolism Prediction* (K. Johannes, ed.), ch. 4, pp. 77–102, Weinheim: Wiley-VCH, 2014.

[9] E. F. Johnson and C. D. Stout, "Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases.," *Biochemical and biophysical research communications*, vol. 338, pp. 331–6, Dec. 2005.

[10] E. F. Johnson and C. D. Stout, "Structural diversity of eukaryotic membrane cytochrome p450s.," *The Journal of biological chemistry*, vol. 288, pp. 17082–90, June 2013.

[11] J. Skopalík, P. Anzenbacher, and M. Otyepka, "Flexibility of human cytochromes p450: molecular dynamics reveals differences between cyps 3a4, 2c9, and 2a6, which correlate with their substrate preferences.," *J. Phys. Chem. B*, vol. 112, pp. 8165–8173, Jul 2008.

[12] V. Cojocaru, P. J. Winn, and R. C. Wade, "The ins and outs of cytochrome P450s.," *Biochimica et biophysica acta*, vol. 1770, pp. 390–401, Mar. 2007.

[13] S. K. Lüdemann, O. Carugo, and R. C. Wade, "Substrate access to cytochrome p450cam: a comparison of a thermal motion pathway analysis with molecular dynamics simulation data," *J. Mol. Model.*, vol. 3, pp. 1–5, 1997.

[14] D. C. Haines, D. R. Tomchick, M. Machius, and J. A. Peterson, "Pivotal role of water in the mechanism of p450bm-3.," *Biochemistry*, vol. 40, pp. 13456–13465, Nov 2001.

[15] L. C. Bell-Parikh and F. P. Guengerich, "Kinetics of cytochrome p450 2e1-catalyzed oxidation of ethanol to acetic acid via acetaldehyde.," *J. Biol. Chem.*, vol. 274, pp. 23833–23840, Aug 1999.

[16] S. K. Lüdemann, 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*, vol. 303, pp. 797–811, Nov. 2000.

[17] S. K. Lüdemann, V. Lounnas, and R. C. Wade, "How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways.," *Journal of molecular biology*, vol. 303, pp. 813–30, Nov. 2000.

[18] T. Zhang, L. A. Liu, D. F. V. Lewis, and D.-Q. Wei, "Long-range effects of a peripheral mutation on the enzymatic activity of cytochrome P450 1A2.," *Journal of chemical information and modeling*, vol. 51, pp. 1336–46, June 2011.

[19] W. Li, J. Shen, G. Liu, Y. Tang, and T. Hoshino, "Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations.," *Proteins*, vol. 79, pp. 271–81, Jan. 2011.

[20] T. Hendrychova, K. Berka, V. Navratilova, P. Anzenbacher, and M. Otyepka, "Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations.," *Current drug metabolism*, vol. 13, pp. 177–89, Feb. 2012.

[21] W. Li, H. Liu, X. Luo, W. Zhu, and Y. Tang, "Possible pathway (s) of metyrapone egress from the active site of cytochrome P450 3A4: a molecular dynamics simulation," *Drug Metabolism and Disposition*, vol. 35, no. 4, pp. 689–696, 2007.

[22] D. Fishelovitch, S. Shaik, H. J. Wolfson, and R. Nussinov, "Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism.," *The journal of physical chemistry. B*, vol. 113, pp. 13018–25, Oct. 2009.

[23] N. Krishnamoorthy, P. Gajendrarao, S. Thangapandian, Y. Lee, and K. W. Lee, "Probing possible egress channels for multiple ligands in human CYP3A4: a molecular modeling study.," *Journal of molecular modeling*, vol. 16, pp. 607–14, Apr. 2010.

[24] I. G. Denisov, A. Y. Shih, and S. G. Sligar, "Structural differences between soluble and membrane bound cytochrome P450s.," *Journal of inorganic biochemistry*, vol. 108, pp. 150–8, Mar. 2012.

[25] P. Gajendrarao, N. Krishnamoorthy, S. Sakkiah, P. Lazar, and K. W. Lee, "Molecular modeling study on orphan human protein CYP4A22 for identification of potential ligand binding site.," *Journal of molecular graphics & modelling*, vol. 28, pp. 524–32, Feb. 2010.

[26] W. Li, H. Liu, E. E. Scott, F. Gräter, J. R. Halpert, X. Luo, J. Shen, and H. Jiang, "Possible pathway(s) of testosterone egress from the active site of cytochrome P450 2B1: a steered molecular dynamics simulation.," *Drug metabolism and disposition: the biological fate of chemicals*, vol. 33, pp. 910–9, July 2005.

[27] K. Schleinkofer, P. J. Winn, S. K. Lüdemann, and R. C. Wade, "Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling?," *EMBO reports*, vol. 6, pp. 584–589, June 2005.

[28] V. Cojocaru, P. J. Winn, and R. C. Wade, "Multiple, Ligand-dependent Routes from the Active Site of Cytochrome P450 2C9," *Current Drug Metabolism*, vol. 13, pp. 143–154, Feb. 2012.

[29] V. Cojocaru, K. Balali-Mood, M. S. P. Sansom, and R. C. Wade, "Structure and dynamics of the membrane-bound cytochrome P450 2C9.," *PLoS computational biology*, vol. 7, p. e1002152, Aug. 2011.

[30] K. Berka, T. Hendrychová, P. Anzenbacher, and M. Otyepka, "Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site.," *The journal of physical chemistry. A*, vol. 115, pp. 11248–55, Oct. 2011.

[31] Z. Shen, F. Cheng, Y. Xu, J. Fu, W. Xiao, J. Shen, G. Liu, W. Li, and Y. Tang, "Investigation of indazole unbinding pathways in CYP2E1 by molecular dynamics simulations.," *PloS one*, vol. 7, p. e33500, Jan. 2012.

[32] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor,

and J. Damborsky, "CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures.," *PLoS computational biology*, vol. 8, p. e1002708, Jan. 2012.

[33] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten, "Steered molecular dynamics," in *Computational Molecular Dynamics: Challenges, Methods, Ideas, Lecture Notes in Computational Science and Engineering* (J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel, eds.), pp. 39–65, Springer-Verlag,, 1998.

[34] P. J. Winn, S. K. Lüdemann, R. Gauges, V. Lounnas, and R. C. Wade, "Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 5361–6, Apr. 2002.

[35] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, pp. 2690–3, 1997.

[36] A. Zawaira, L. Coulson, M. Gallotta, O. Karimanzira, and J. Blackburn, "On the deduction and analysis of singlet and two-state gating-models from the static structures of mammalian cyp450.," *J. Struct. Biol.*, vol. 173, pp. 282–293, Feb 2011.

[37] D. F. Lewis, A. Wiseman, and M. H. Tarbit, "Molecular modelling of lanosterol 14 alpha-demethylase (cyp51) from saccharomyces cerevisiae via homology with cyp102, a unique bacterial cytochrome p450 isoform: quantitative structure-activity relationships (qsars) within two related series of antifungal azole derivatives.," *J Enzyme Inhib*, vol. 14, no. 3, pp. 175–192, 1999.

[38] B. Rupp, S. Raub, C. Marian, and H.-D. Höltje, "Molecular design of two sterol 14alpha-demethylase homology models and their interactions with the azole antifungals ketoconazole and bifonazole.," *J Comput Aided Mol Des*, vol. 19, pp. 149–163, Mar 2005.

[39] S. Li, K. Xu, Y. Liu, M. He, Y. Yuan, J. Yang, and Y. Yan, "Structural modeling of cyp51 from penicillium digitatum and the inhibitor coordination inter-

actions," in *5th International Conference on Bioinformatics and Biomedical Engineering, (iCBBE)*, pp. 1–4, IEEE, 2011.

[40] X. Yu, V. Cojocaru, G. Mustafa, O. M. H. Salo-Ahen, G. I. Lepesheva, and R. C. Wade, "Dynamics of cyp51: implications for function and inhibitor design.," *J Mol Recognit*, vol. 28, pp. 59–73, Feb 2015.

[41] A. R. Leach in *Molecular Modelling: Principles and Applications*, Essex, England: Pearson Education Limited, 2nd edition ed., 2001.

[42] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. a. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *Journal of the American Chemical Society*, vol. 117, pp. 5179–5197, May 1995.

[43] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "Charmm: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, pp. 187–217, 1983.

[44] W. R. P. Scott and W. F. van Gunsteren, "The gromos software package for biomolecular simulations," in *Methods and Techniques in Computational Chemistry: METECC-95* (E. Clementi and G. Corongiu, eds.), (Cagliari, Italy), pp. 397–434, 1995.

[45] O. Guvench and A. D. MacKerell, Jr, "Comparison of protein force fields for molecular dynamics simulations.," *Methods Mol Biol*, vol. 443, pp. 63–88, 2008.

[46] P. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot, and A. Pohorille in *Computer Simulations of Biomolecular Systems* (W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, eds.), vol. 3, Kluwer Academic Publishers, 1997.

[47] J. Wang, P. Cieplak, and P. A. Kollman, "How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules?," *J. Comput. Chem.*, vol. 21, pp. 1049–74, 2000.

[48] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.," *J. Comput. Chem.*, vol. 24, pp. 1999–2012, Dec 2003.

[49] D. A. Case, T. Darden, T. Cheatham, III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. C. Walker, W. Zhang, K. Merz, B. Reberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. Kollman, *Amber 11*. San Francisco: University of Califonia, 2010.

[50] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters.," *Proteins*, vol. 65, pp. 712–25, Nov. 2006.

[51] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the amber ff99sb protein force field.," *Proteins*, vol. 78, pp. 1950–1958, Jun 2010.

[52] E. J. Sorin and V. S. Pande, "Exploring the helix-coil transition via all-atom equilibrium ensemble simulations.," *Biophys J*, vol. 88, pp. 2472–2493, Apr 2005.

[53] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field.," *Journal of computational chemistry*, vol. 25, pp. 1157–74, July 2004.

[54] D. A. Case, T. Darden, T. Cheatham, III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. C. Walker, W. Zhang, K. Merz, B. Reberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Goetz, I. Kolossvary, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, C. Sagui, R. Salomon-Ferrer, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. Kollman, *Amber 12*. San Francisco: University of Califonia, 2012.

[55] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, "Scalable molecular dynamics with NAMD.," *Journal of computational chemistry*, vol. 26, pp. 1781–802, Dec. 2005.

[56] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, "A new force field for molecular mechanical simulation of nucleic acids and proteins," *J. Am. Chem. Soc.*, vol. 106, no. 3, pp. 765–84, 1984.

[57] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model," *The Journal of Physical Chemistry*, vol. 97, pp. 10269–10280, Oct. 1993.

[58] J. W. Ponder and D. A. Case, "Force fields for protein simulations.," *Adv. Protein Chem.*, vol. 66, pp. 27–85, 2003.

[59] P. P. Ewald, "Die berechnung optischer und elektrostatischer gitterpotentiale," *Ann. Phys.*, vol. 369, pp. 253–87, 1921.

[60] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald: An n log(n) method for ewald sums in large systems," *J. Chem. Phys.*, vol. 98, pp. 10089–92, 1993.

[61] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, pp. 327–341, Mar. 1977.

[62] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J. Chem. Phys.*, vol. 81, no. 3684, 1985.

[63] B. Andersson and J. M. Anderson, "Lateral heterogeneity in the distribution of chlorophyll-protein complexes of the thylakoid membranes of spinach chloroplasts.," *Biochim Biophys Acta*, vol. 593, pp. 427–440, Dec 1980.

[64] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–35, 1983.

[65] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction models for water in relation to protein hydration," in *Intermolecular Forces* (B. Pullman, ed.), no. 331-42, D. Reidel, 1981.

[66] S.-J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI force field: coarse grained model for biomolecular simulations.," *The journal of physical chemistry. B*, vol. 111, pp. 7812–24, July 2007.

[67] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, "The MARTINI Coarse-Grained Force Field: Extension to Proteins," *Journal of Chemical Theory and Computation*, vol. 4, pp. 819–834, May 2008.

[68] X. Periole, M. Cavalli, S.-J. Marrink, and M. A. Ceruso, "Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition," *Journal of Chemical Theory and Computation*, vol. 5, pp. 2531–2543, Sept. 2009.

[69] T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, and D. P. Tieleman, "Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models," *Journal of Chemical Theory and Computation*, pp. 676–90, Dec. 2013.

[70] R. Brown, "A brief account of microscopical observations," 1827.

[71] R. R. Gabdoulline and R. C. Wade, "Brownian dynamics simulation of protein-protein diffusional encounter.," *Methods (San Diego, Calif.)*, vol. 14, pp. 329–41, Mar. 1998.

[72] D. L. Ermak and J. A. McCammon, "Brownian dynamics with hydrodynamic interactions," *The Journal of Chemical Physics*, vol. 69, no. 4, p. 1352, 1978.

[73] R. R. Gabdoulline and R. C. Wade, "Simulation of the diffusional association of barnase and barstar.," *Biophysical journal*, vol. 72, pp. 1917–29, May 1997.

[74] R. R. Gabdoulline and R. C. Wade, "On the contributions of diffusion and thermal activation to electron transfer between Phormidium laminosum plastocyanin and cytochrome f: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer event," *Journal of the American Chemical Society*, vol. 131, pp. 9230–8, July 2009.

[75] R. R. Gabdoulline and R. C. Wade, "Effective Charges for Macromolecules in Solvent," *The Journal of Physical Chemistry*, vol. 100, pp. 3868–3878, Jan. 1996.

[76] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: application to microtubules and the ribosome.," *Proc. Natl. Acad. Sci.*, vol. 98, pp. 10037–10041, Aug 2001.

[77] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. Scott, and J. McCammon, "Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program," *Computer Physics Communications*, vol. 91, pp. 57–95, Sept. 1995.

[78] A. H. Elcock, R. R. Gabdoulline, R. C. Wade, and J. A. McCammon, "Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin.," *Journal of Molecular Biology*, vol. 291, pp. 149–162, Aug 1999.

[79] A. Karyakin, D. Motiejunas, R. C. Wade, and C. Jung, "FTIR studies of the redox partner interaction in cytochrome P450: the Pdx-P450cam couple.," *Biochimica et biophysica acta*, vol. 1770, pp. 420–31, Mar. 2007.

[80] G. V. Pachov, R. R. Gabdoulline, and R. C. Wade, "On the structure and dynamics of the complex of the nucleosome and the linker histone.," *Nucleic acids research*, vol. 39, pp. 5255–63, July 2011.

[81] A. Sandikci, F. Gloge, M. Martinez, M. P. Mayer, R. Wade, B. Bukau, and G. Kramer, "Dynamic enzyme docking to the ribosome coordinates N-terminal processing with polypeptide folding.," *Nature structural & molecular biology*, vol. 20, pp. 843–50, July 2013.

[82] L. Gombos, A. Neuner, M. Berynskyy, L. L. Fava, R. C. Wade, C. Sachse, and E. Schiebel, "GTP regulates the microtubule nucleation activity of γ-tubulin.," *Nature cell biology*, vol. 15, pp. 1317–27, Nov. 2013.

[83] D. Motiejunas, R. Gabdoulline, T. Wang, A. Feldman-Salit, T. Johann, P. J. Winn, and R. C. Wade, "Protein-protein docking by simulating the process of association subject to biochemical constraints.," *Proteins*, vol. 71, pp. 1955–69, June 2008.

[84] P. Mereghetti, M. Ganadu, E. Papaleo, P. Fantucci, and L. De Gioia, "Validation of protein models by a neural network approach," *BMC Bioinformatics*, vol. 9, no. 1, p. 66, 2008.

[85] P. Mereghetti, R. R. Gabdoulline, and R. C. Wade, "Brownian dynamics simulation of protein solutions: structural and dynamical properties.," *Biophysical journal*, vol. 99, pp. 3782–91, Dec. 2010.

[86] P. Mereghetti, D. Kokh, J. A. McCammon, and R. C. Wade, "Diffusion and association processes in biological systems: theory, computation and experiment.," *BMC biophysics*, vol. 4, p. 2, Jan. 2011.

[87] P. Mereghetti and R. C. Wade, "Atomic detail brownian dynamics simulations of concentrated protein solutions with a mean field treatment of hydrodynamic interactions.," *The journal of physical chemistry. B*, vol. 116, pp. 8523–33, July 2012.

[88] P. Mereghetti and R. C. Wade, "Brownian dynamics simulation of protein diffusion in crowded environments," in *AIP Conf. Proc.*, pp. 511–516, 2013.

[89] Y. Yoshida, Y. Aoyama, M. Noshiro, and O. Gotoh, "Sterol 14-demethylase P450 (CYP51) provides a breakthrough for the discussion on the evolution of cytochrome P450 gene superfamily.," *Biochemical and biophysical research communications*, vol. 273, pp. 799–804, July 2000.

[90] Y. Aoyama, M. Noshiro, O. Gotoh, S. Imaoka, and Y. Funae, "Sterol 14-Demethylase P450 (P45014DM *) Is One of the Most Ancient and Conserved P450 Species," *The Journal of Biochemistry*, vol. 933, pp. 926–933, 1996.

[91] D. R. Nelson, "Cytochrome P450 and the individuality of species.," *Archives of biochemistry and biophysics*, vol. 369, pp. 1–10, Sept. 1999.

[92] Y. Yoshida, "Cytochrome P450 of fungi: primary target for azole antifungal agents.," *Current topics in medical mycology*, vol. 2, pp. 388–418, Jan. 1988.

[93] G. I. Lepesheva, T. Y. Hargrove, Y. Kleshchenko, W. D. Nes, F. Villalta, and M. R. Waterman, "CYP51: A major drug target in the cytochrome P450 superfamily.," *Lipids*, vol. 43, pp. 1117–25, Dec. 2008.

[94] G. I. Lepesheva and M. R. Waterman, "Sterol 14alpha-Demethylase (CYP51) as a Therapeutic Target for Human Trypanosomiasis and Leishmaniasis," *Current Topics in Medicinal Chemistry*, vol. 11, pp. 2060–2071, Aug. 2011.

[95] F. Villalta, M. C. Dobish, P. N. Nde, Y. Y. Kleshchenko, T. Y. Hargrove, C. A. Johnson, M. R. Waterman, J. N. Johnston, and G. I. Lepesheva, "VNI cures acute and chronic experimental Chagas disease.," *The Journal of infectious diseases*, vol. 208, pp. 504–11, Aug. 2013.

[96] T. Y. Hargrove, Z. Wawrzak, J. Liu, M. R. Waterman, W. D. Nes, and G. I. Lepesheva, "Structural complex of sterol 14α-demethylase (CYP51) with 14α-methylenecyclopropyl-Δ7-24, 25-dihydrolanosterol.," *Journal of lipid research*, vol. 53, pp. 311–20, Feb. 2012.

[97] G. I. Lepesheva, T. Y. Hargrove, S. Anderson, Y. Kleshchenko, V. Furtak, Z. Wawrzak, F. Villalta, and M. R. Waterman, "Structural insights into inhibition of sterol 14alpha-demethylase in the human pathogen Trypanosoma cruzi.," *The Journal of biological chemistry*, vol. 285, pp. 25582–90, Aug. 2010.

[98] T. Y. Hargrove, K. Kim, M. de Nazaré Correia Soeiro, C. F. da Silva, D. da Gama Jaen Batista, M. M. Batista, E. M. Yazlovitskaya, M. R. Waterman, G. A. Sulikowski, and G. I. Lepesheva, "CYP51 structures and structure-based development of novel, pathogen-specific inhibitory scaffolds," *International Journal for Parasitology: Drugs and Drug Resistance*, vol. 2, pp. 178–186, Dec. 2012.

[99] G. Andriani, E. Amata, J. Beatty, Z. Clements, B. J. Coffey, G. Courtemanche, W. Devine, J. Erath, C. E. Juda, Z. Wawrzak, J. T. Wood, G. I. Lepesheva, A. Rodriguez, and M. P. Pollastri, "Antitrypanosomal lead discovery: identification of a ligand-efficient inhibitor of Trypanosoma cruzi CYP51 and parasite growth.," *Journal of medicinal chemistry*, vol. 56, pp. 2556–67, Mar. 2013.

[100] T. Y. Hargrove, Z. Wawrzak, J. Liu, W. D. Nes, M. R. Waterman, and G. I. Lepesheva, "Substrate preferences and catalytic parameters determined by structural characteristics of sterol 14alpha-demethylase (CYP51) from Leishmania infantum.," *The Journal of biological chemistry*, vol. 286, pp. 26838–48, July 2011.

[101] T. H. Bayburt and S. G. Sligar, "Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 6725–30, May 2002.

[102] B. Frey and D. David, "Determination of the Membrane Topology of the Phenobarbital-inducible Rat Liver Cytochrome P-450 Isoenzyme PB-4 Using Site-specific Antibodies," *The Journal of Cell Biology*, vol. 104, pp. 209–219, 1987.

[103] M. Headlam, "The F-G loop region of cytochrome P450scc (CYP11A1) interacts with the phospholipid membrane," *Biochimica et Biophysica Acta, Biomembranes*, vol. 1617, pp. 96–108, Oct. 2003.

[104] Y. Ohta, S. Kawato, H. Tagashira, S. Takemori, and S. Kominami, "Dynamic structures of adrenocortical cytochrome P-450 in proteoliposomes and microsomes: protein rotation study," *Biochemistry*, vol. 31, pp. 12680–12687, Dec. 1992.

[105] C. Ozalp, E. Szczesna-Skorupa, and B. Kemper, "Identification of membrane-contacting loops of the catalytic domain of cytochrome P450 2C2 by tryptophan fluorescence scanning.," *Biochemistry*, vol. 45, pp. 4629–37, Apr. 2006.

[106] I. A. Pikuleva, "Putative F-G loop is involved in association with the membrane in P450scc (P450 11A1).," *Molecular and cellular endocrinology*, vol. 215, pp. 161–4, Feb. 2004.

[107] S. D. Black, S. T. Martin, and C. A. Smith, "Membrane Topology of Liver Microsomal Cytochrome P450 2B4 Determined via Monoclonal Antibodies Directed to the Halt-Transfer Signal," *Biochemistry*, vol. 33, pp. 6945–6951, June 1994.

[108] K. Yamamoto, M. Gildenberg, S. Ahuja, S.-C. Im, P. Pearcy, L. Waskell, and A. Ramamoorthy, "Probing the transmembrane structure and topology of microsomal cytochrome-p450 by solid-state NMR on temperature-resistant bicelles.," *Scientific reports*, vol. 3, p. 2556, Aug. 2013.

[109] B. C. Monk, T. M. Tomasiak, M. V. Keniya, F. U. Huschmann, J. D. a. Tyndall, J. D. O'Connell, R. D. Cannon, J. G. McDonald, A. Rodriguez, J. S. Finer-Moore, and R. M. Stroud, "Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 3865–3870, Feb. 2014.

[110] J. Sgrignani and A. Magistrato, "Influence of the membrane lipophilic environment on the structure and on the substrate access/egress routes of the human aromatase enzyme. A computational study.," *Journal of chemical information and modeling*, vol. 52, pp. 1595–606, June 2012.

[111] J. L. Baylon, I. L. Lenov, S. G. Sligar, and E. Tajkhorshid, "Characterizing the membrane-bound state of cytochrome P450 3A4: structure, depth of insertion, and orientation.," *Journal of the American Chemical Society*, vol. 135, pp. 8542–51, June 2013.

[112] K. Berka, M. Paloncýová, P. Anzenbacher, and M. Otyepka, "Behavior of Human Cytochromes P450 on Lipid Membranes.," *The journal of physical chemistry. B*, vol. 117, pp. 11556–11564, Sept. 2013.

[113] S. Haider, B. Islam, V. D'Atri, M. Sgobba, C. Poojari, L. Sun, T. Yuen, M. Zaidi, and M. I. New, "Structure-phenotype correlations of human CYP21A2 mutations in congenital adrenal hyperplasia.," *Proceedings of the*

National Academy of Sciences of the United States of America, vol. 110, pp. 2605–10, Feb. 2013.

[114] Y. Miao and J. Baudry, "Active-site hydration and water diffusion in cytochrome P450cam: a highly dynamic process.," *Biophysical journal*, vol. 101, pp. 1493–503, Sept. 2011.

[115] Y. Miao, Z. Yi, C. Cantrell, D. C. Glass, J. Baudry, N. Jain, and J. C. Smith, "Coupled flexibility change in cytochrome P450cam substrate binding determined by neutron scattering, NMR, and molecular dynamics simulation.," *Biophysical journal*, vol. 103, pp. 2167–76, Nov. 2012.

[116] Schrodinger LLC, "The PyMOL Molecular Graphics System," 2010.

[117] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. a. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.," *Nucleic acids research*, vol. 32, pp. W665–7, July 2004.

[118] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, Feb. 1996.

[119] D. L. Harris, J.-Y. Park, L. Gruenke, and L. Waskell, "Theoretical study of the ligand-CYP2B4 complexes: effect of structure on binding free energies and heme spin state.," *Proteins*, vol. 55, pp. 895–914, June 2004.

[120] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak, and F.-Y. Dupradeau, "R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments.," *Nucleic acids research*, vol. 39, pp. W511–7, July 2011.

[121] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, and P. Cieplak, "The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building.," *Physical chemistry chemical physics : PCCP*, vol. 12, pp. 7821–39, July 2010.

[122] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations.," *Journal of molecular graphics & modelling*, vol. 25, pp. 247–60, Oct. 2006.

[123] W. L. Jorgensen and J. D. Madura, "Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water," *Journal of the American Chemical Society*, vol. 105, pp. 1407–1413, Mar. 1983.

[124] J. Seelig and N. Waespe-Sarcevic, "Molecular order in cis and trans unsaturated phospholipid bilayers.," *Biochemistry*, vol. 17, pp. 3310–5, Aug. 1978.

[125] B. Jójárt and T. A. Martinek, "Performance of the general amber force field in modeling aqueous POPC membrane bilayers.," *Journal of computational chemistry*, vol. 28, pp. 2051–8, Sept. 2007.

[126] S. W. I. Siu, R. Vácha, P. Jungwirth, and R. A. Böckmann, "Biomolecular simulations of membranes: physical properties from different force fields.," *The Journal of Chemical Physics*, vol. 128, p. 125103, Mar. 2008.

[127] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints.," *Journal of molecular biology*, vol. 234, pp. 779–815, Dec. 1993.

[128] S.-J. Marrink, A. H. de Vries, and A. E. Mark, "Coarse Grained Model for Semiquantitative Lipid Simulations," *The Journal of Physical Chemistry B*, vol. 108, pp. 750–760, Jan. 2004.

[129] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation," *Journal of Chemical Theory and Computation*, vol. 4, pp. 435–447, Mar. 2008.

[130] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: fast, flexible, and free.," *Journal of computational chemistry*, vol. 26, pp. 1701–18, Dec. 2005.

[131] H. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation," *Computer Physics Communications*, vol. 91, pp. 43–56, Sept. 1995.

[132] E. Lindahl, B. Hess, and D. van der Spoel, "GROMACS 3.0: a package for molecular simulation and trajectory analysis," *Journal of molecular modeling*, pp. 306–317, 2001.

[133] A. J. Rzepiela, L. V. Schäfer, N. Goga, H. J. Risselada, A. H. De Vries, and S. J. Marrink, "Reconstruction of atomistic details from coarse-grained structures.," *Journal of computational chemistry*, vol. 31, pp. 1333–43, Apr. 2010.

[134] J. Pei, B.-H. Kim, and N. V. Grishin, "PROMALS3D: a tool for multiple protein sequence and structure alignments.," *Nucleic acids research*, vol. 36, pp. 2295–300, Apr. 2008.

[135] W. J. Allen, J. A. Lemkul, and D. R. Bevan, "GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics.," *Journal of computational chemistry*, vol. 30, pp. 1952–8, Sept. 2009.

[136] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.," *Bioinformatics (Oxford, England)*, vol. 29, pp. 845–54, Apr. 2013.

[137] M. a. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, "OPM: orientations of proteins in membranes database.," *Bioinformatics (Oxford, England)*, vol. 22, pp. 623–5, Mar. 2006.

[138] O. Gotoh, "Substrate Recognition Sites in Cytochrome P450 Family 2 (CYP2) Proteins Inferred from Comparative Analyses of Amino Acid and Coding Nucleotide Sequences," *Journal of Biological Chemistry*, vol. 267, no. 1, pp. 83–90, 1992.

[139] Z. Chen, T. W. B. Ost, and J. P. M. Schelvis, "Phe393 mutants of cytochrome P450 BM3 with modified heme redox potentials have altered heme vinyl and propionate conformations.," *Biochemistry*, vol. 43, pp. 1798–808, Feb. 2004.

[140] E. Stjernschantz, B. M. a. van Vugt-Lussenburg, A. Bonifacio, S. B. a. de Beer, G. van der Zwan, C. Gooijer, J. N. M. Commandeur, N. P. E. Ver-

meulen, and C. Oostenbrink, "Structural rationalization of novel drug metabolizing mutants of cytochrome P450 BM3.," *Proteins*, vol. 71, pp. 336–52, Apr. 2008.

[141] C. A. Hasemann, R. G. Kurumbail, S. S. Boddupalli, J. A. Peterson, and J. Deisenhofer, "Structure and function of cytochromes P450:a comparative analysis of three crystal structures," *Structure*, vol. 3, pp. 41–62, Jan. 1995.

[142] V. Guallar, M.-h. Baik, S. J. Lippard, and R. A. Friesner, "Peripheral heme substituents control the hydrogen-atom abstraction chemistry in cytochromes P450.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp. 6998–7002, June 2003.

[143] R. C. Wade, P. J. Winn, I. Schlichting, and Sudarko, "A survey of active site access channels in cytochromes P450.," *Journal of inorganic biochemistry*, vol. 98, pp. 1175–82, July 2004.

[144] G. I. Lepesheva, W. D. Nes, W. Zhou, G. C. Hill, and M. R. Waterman, "CYP51 from Trypanosoma brucei is obtusifoliol-specific.," *Biochemistry*, vol. 43, pp. 10789–99, Aug. 2004.

[145] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09 Revision A.1," 2009.

[146] Schrodinger LLC, "Maestro, version 9.3," 2012.

[147] R. C. Wade, D. Motiejunas, K. Schleinkofer, Sudarko, P. J. Winn, A. Banerjee, A. Kariakin, and C. Jung, "Multiple molecular recognition mechanisms. Cytochrome P450–a case study.," *Biochimica et biophysica acta*, vol. 1754, pp. 239–44, Dec. 2005.

[148] A. N. Eddine, J. P. von Kries, M. V. Podust, T. Warrier, S. H. E. Kaufmann, and L. M. Podust, "X-ray structure of 4,4'-dihydroxybenzophenone mimicking sterol substrate in the active site of sterol 14alpha-demethylase (CYP51).," *The Journal of biological chemistry*, vol. 283, pp. 15152–9, May 2008.

[149] S.-C. Im and L. Waskell, "The interaction of microsomal cytochrome P450 2B4 with its redox partners, cytochrome P450 reductase and cytochrome b(5).," *Archives of biochemistry and biophysics*, vol. 507, pp. 144–53, Mar. 2011.

[150] I. F. Sevrioukova, H. Li, H. Zhang, J. A. Peterson, and T. L. Poulos, "Structure of a cytochrome P450-redox partner electron-transfer complex," *Proceedings of the National Academy of Sciences*, vol. 96, pp. 1863–1868, Mar. 1999.

[151] N. Strushkevich, F. MacKenzie, T. Cherkesova, I. Grabovec, S. Usanov, and H.-W. Park, "Structural basis for pregnenolone biosynthesis by the mitochondrial monooxygenase system," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 25, pp. 10139–43, 2011.

[152] S. Tripathi, H. Li, and T. L. Poulos, "Structural basis for effector control and redox partner recognition in cytochrome P450.," *Science (New York, N.Y.)*, vol. 340, pp. 1227–30, June 2013.

[153] Y. Hiruma, M. a.S. Hass, Y. Kikui, W.-M. Liu, B. Ölmez, S. P. Skinner, A. Blok, A. Kloosterman, H. Koteishi, F. Löhr, H. Schwalbe, M. Nojiri, and M. Ubbink, "The Structure of the Cytochrome P450cam-Putidaredoxin Complex Determined by Paramagnetic NMR Spectroscopy and Crystallography," *Journal of Molecular Biology*, July 2013.

[154] A. W. Munro, D. G. Leys, K. J. McLean, K. R. Marshall, T. W. B. Ost, S. Daff, C. S. Miles, S. K. Chapman, D. A. Lysek, C. C. Moser, C. C. Page, and P. L.

Dutton, "P450 bm3: the very model of a modern flavocytochrome.," *Trends Biochem Sci*, vol. 27, pp. 250–257, May 2002.

[155] C. Xia, S. P. Panda, C. C. Marohnic, P. Martásek, B. S. Masters, and J.-J. P. Kim, "Structural basis for human nadph-cytochrome p450 oxidoreductase deficiency.," *Proc Natl Acad Sci U S A*, vol. 108, pp. 13486–13491, Aug 2011.

[156] D. Hamdane, C. Xia, S.-C. Im, H. Zhang, J.-J. P. Kim, and L. Waskell, "Structure and function of an NADPH-cytochrome P450 oxidoreductase in an open conformation capable of reducing cytochrome P450.," *The Journal of biological chemistry*, vol. 284, pp. 11374–84, Apr. 2009.

[157] C. Kenaan, H. Zhang, E. V. Shea, and P. F. Hollenberg, "Uncovering the role of hydrophobic residues in cytochrome p450-cytochrome p450 reductase interactions.," *Biochemistry*, vol. 50, pp. 3957–3967, May 2011.

[158] H.-H. Jang, A. P. Jamakhandi, S. Z. Sullivan, C.-H. Yun, P. F. Hollenberg, and G. P. Miller, "Beta sheet 2-alpha helix c loop of cytochrome p450 reductase serves as a docking site for redox partners.," *Biochim Biophys Acta*, vol. 1804, pp. 1285–1293, Jun 2010.

[159] R. W. Kelley, D. Cheng, and W. L. Backes, "Heteromeric complex formation between CYP2E1 and CYP1A2: evidence for the involvement of electrostatic interactions.," *Biochemistry*, vol. 45, pp. 15807–16, Dec. 2006.

[160] J. R. Reed, R. W. Kelley, and W. L. Backes, "An evaluation of methods for the reconstitution of cytochromes P450 and NADPH P450 reductase into lipid vesicles.," *Drug metabolism and disposition: the biological fate of chemicals*, vol. 34, pp. 660–6, Apr. 2006.

[161] J. R. Reed and W. L. Backes, "Formation of p450 · p450 complexes and their effect on p450 function.," *Pharmacol Ther*, vol. 133, pp. 299–310, Mar 2012.

[162] J. R. Reed, G. F. Cawley, and W. L. Backes, "Interactions between cytochromes p450 2b4 (cyp2b4) and 1a2 (cyp1a2) lead to alterations in toluene disposition and p450 uncoupling.," *Biochemistry*, vol. 52, pp. 4003–4013, Jun 2013.

[163] J. R. Reed, J. P. Connick, D. Cheng, G. F. Cawley, and W. L. Backes, "Effect of homomeric p450-p450 complexes on p450 function.," *Biochem J*, vol. 446, pp. 489–497, Sep 2012.

[164] E. Hazai and D. Kupfer, "Interactions between CYP2C9 and CYP2C19 in reconstituted binary systems influence their catalytic activity: possible rationale for the inability of CYP2C19 to catalyze methoxychlor demethylation in human liver microsomes.," *Drug metabolism and disposition: the biological fate of chemicals*, vol. 33, pp. 157–64, Jan. 2005.

[165] A. Sündermann and C. Oostenbrink, "Molecular dynamics simulations give insight into the conformational change, complex formation, and electron transfer pathway for cytochrome P450 reductase.," *Protein science : a publication of the Protein Society*, vol. 22, pp. 1183–95, Sept. 2013.

[166] A. Zawaira, M. Gallotta, N. Beeton-Kempen, L. Coulson, P. Marais, M. Kuttel, and J. Blackburn, "Exhaustive computational search of ionic-charge clusters that mediate interactions between mammalian cytochrome p450 (cyp) and p450-oxidoreductase (por) proteins.," *Comput Biol Chem*, vol. 34, pp. 42–52, Feb 2010.

[167] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker, "Pdb2pqr: expanding and upgrading automated preparation of biomolecular structures for molecular simulations.," *Nucleic Acids Res*, vol. 35, pp. W522–W525, Jul 2007.

[168] V. Tsui and D. A. Case, "Theory and applications of the generalized born solvation model in macromolecular simulations," *Biopolymers*, vol. 56, pp. 275–291, 2001.

[169] Y. Oguchi, E. Kummer, F. Seyffer, M. Berynskyy, B. Anstett, R. Zahn, R. C. Wade, A. Mogk, and B. Bukau, "A tightly regulated molecular toggle controls AAA+ disaggregase.," *Nature structural & molecular biology*, vol. 19, pp. 1338–46, Dec. 2012.

[170] Y. Madrona, S. A. Hollingsworth, S. Tripathi, J. B. Fields, J.-C. N. Rwigema, D. J. Tobias, and T. L. Poulos, "Crystal structure of cindoxin, the p450cin redox partner.," *Biochemistry*, vol. 53, pp. 1435–1446, Mar 2014.

[171] R. Méndez, R. Leplae, M. F. Lensink, and S. J. Wodak, "Assessment of capri predictions in rounds 3-5 shows progress in docking procedures.," *Proteins*, vol. 60, pp. 150–169, Aug 2005.

[172] a. Bridges, "Identification of the Binding Site on Cytochrome P450 2B4 for Cytochrome b5 and Cytochrome P450 Reductase," *Journal of Biological Chemistry*, vol. 273, pp. 17036–17049, July 1998.

[173] D. R. Davydov, N. Y. Davydova, E. V. Sineva, I. Kufareva, and J. R. Halpert, "Pivotal role of P450-P450 interactions in CYP3A4 allostery: the case of α-naphthoflavone.," *The Biochemical journal*, vol. 453, pp. 219–30, July 2013.

[174] D. R. Davydov, N. Y. Davydova, E. V. Sineva, and J. R. Halpert, "Interactions among cytochromes p450 in microsomal membranes: Oligomerization of cytochromes p450 3a4, 3a5, and 2e1 and its functional consequences.," *J Biol Chem*, vol. 290, pp. 3850–3864, Feb 2015.

[175] R. Verma, U. Schwaneberg, and D. Roccatano, "Insight into the redox partner interaction mechanism in cytochrome p450bm-3 using molecular dynamics simulations.," *Biopolymers*, vol. 101, pp. 197–209, Mar 2014.

[176] R. R. Gabdoulline and R. C. Wade, "Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations.," *J Mol Biol*, vol. 306, pp. 1139–1155, Mar 2001.

[177] K. M. Elsawy, L. S. D. Caves, and R. Twarock, "The impact of viral rna on the association rates of capsid protein assembly: bacteriophage ms2 as a case study.," *J Mol Biol*, vol. 400, pp. 935–947, Jul 2010.

[178] J. Balbo, P. Mereghetti, D.-P. Herten, and R. C. Wade, "The shape of protein crowders is a major determinant of protein diffusion.," *Biophysical journal*, vol. 104, pp. 1576–84, Apr. 2013.

[179] S. J. de Vries, M. van Dijk, and A. M. J. J. Bonvin, "The haddock web server for data-driven biomolecular docking.," *Nat Protoc*, vol. 5, pp. 883–897, May 2010.

[180] S. R. Comeau, D. W. Gatchell, S. Vajda, and C. J. Camacho, "Cluspro: a fully automated algorithm for protein-protein docking.," *Nucleic Acids Res*, vol. 32, pp. W96–W99, Jul 2004.

[181] G. Macindoe, L. Mavridis, V. Venkatraman, M.-D. Devignes, and D. W. Ritchie, "HexServer: an FFT-based protein docking server powered by graphics processors.," *Nucleic acids research*, vol. 38, pp. W445–9, July 2010.

[182] B. G. Pierce, K. Wiehe, H. Hwang, B.-H. Kim, T. Vreven, and Z. Weng, "ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers.," *Bioinformatics (Oxford, England)*, Mar. 2014.

[183] M. Torchala, I. H. Moal, R. A. G. Chaleil, J. Fernandez-Recio, and P. A. Bates, "Swarmdock: a server for flexible protein-protein docking.," *Bioinformatics*, vol. 29, pp. 807–809, Mar 2013.

[184] B. Jiménez-García, C. Pons, and J. Fernández-Recio, "pydockweb: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring.," *Bioinformatics*, vol. 29, pp. 1698–1699, Jul 2013.

[185] S. Qin, X. Pang, and H.-X. Zhou, "Automated prediction of protein association rate constants.," *Structure (London, England : 1993)*, vol. 19, pp. 1744–51, Dec. 2011.

List of Figures

List of Tables

9

Appendix

The lib file for parameters of ligand VNI:

```
!!index array str
 "VNI"
!entry.VNI.unit.atoms table  str name  str type  int typex  int resx  int flags  int seq  int elmnt  dbl chg
 "O1" "o" 0 1 131072 1 8 -0.539900
 "C1" "c" 0 1 131072 2 6 0.580500
 "N1" "n" 0 1 131072 3 7 -0.523100
 "H1" "hn" 0 1 131072 4 1 0.247100
 "C2" "c3" 0 1 131072 5 6 0.122200
 "H2" "h1" 0 1 131072 6 1 0.119600
 "C3" "c3" 0 1 131072 7 6 -0.047300
 "H32" "h1" 0 1 131072 8 1 0.075300
 "H33" "h1" 0 1 131072 9 1 0.075300
 "N2" "na" 0 1 131072 10 7 0.018900
 "C4" "cc" 0 1 131072 11 6 -0.374800
 "H4" "h4" 0 1 131072 12 1 0.215900
 "C5" "cd" 0 1 131072 13 6 0.100400
 "H5" "h4" 0 1 131072 14 1 0.133400
 "N3" "nd" 0 1 131072 15 7 -0.567300
 "C6" "cc" 0 1 131072 16 6 0.250500
 "H6" "h5" 0 1 131072 17 1 0.084900
 "C7" "ca" 0 1 131072 18 6 0.085200
 "C8" "ca" 0 1 131072 19 6 -0.167900
 "H8" "ha" 0 1 131072 20 1 0.175200
 "C9" "ca" 0 1 131072 21 6 -0.124500
 "H9" "ha" 0 1 131072 22 1 0.152300
 "C10" "ca" 0 1 131072 23 6 -0.008500
 "CL1" "cl" 0 1 131072 24 6 -0.086800
 "C11" "ca" 0 1 131072 25 6 -0.018400
 "H11" "ha" 0 1 131072 26 1 0.133500
 "C12" "ca" 0 1 131072 27 6 -0.036000
 "CL2" "cl" 0 1 131072 28 6 -0.087200
 "C13" "ca" 0 1 131072 29 6 -0.050400
 "C14" "ca" 0 1 131072 30 6 -0.107600
 "H14" "ha" 0 1 131072 31 1 0.145900
 "C15" "ca" 0 1 131072 32 6 -0.096500
 "H15" "ha" 0 1 131072 33 1 0.123200
 "C16" "ca" 0 1 131072 34 6 -0.156900
 "C17" "ca" 0 1 131072 35 6 -0.096500
 "H17" "ha" 0 1 131072 36 1 0.123200
 "C18" "ca" 0 1 131072 37 6 -0.107600
 "H18" "ha" 0 1 131072 38 1 0.145900
 "C19" "cc" 0 1 131072 39 6 0.638400
 "N4" "nd" 0 1 131072 40 7 -0.404300
 "N5" "nd" 0 1 131072 41 7 -0.359000
 "C20" "cc" 0 1 131072 42 6 0.589900
 "O2" "os" 0 1 131072 43 8 -0.335700
 "C21" "ca" 0 1 131072 44 6 -0.148300
 "C22" "ca" 0 1 131072 45 6 -0.073600
 "H22" "ha" 0 1 131072 46 1 0.109500
 "C23" "ca" 0 1 131072 47 6 -0.117700
 "H23" "ha" 0 1 131072 48 1 0.140000
 "C24" "ca" 0 1 131072 49 6 -0.160900
 "H24" "ha" 0 1 131072 50 1 0.151900
```

"C25" "ca" 0 1 131072 51 6 -0.117700
"H25" "ha" 0 1 131072 52 1 0.140000
"C26" "ca" 0 1 131072 53 6 -0.073600
"H26" "ha" 0 1 131072 54 1 0.109500
!entry.VNI.unit.atomspertinfo table str pname str ptype int ptypex int pelmnt dbl pchg
"O1" "o" 0 -1 0.0
"C1" "c" 0 -1 0.0
"N1" "n" 0 -1 0.0
"H1" "hn" 0 -1 0.0
"C2" "c3" 0 -1 0.0
"H2" "h1" 0 -1 0.0
"C3" "c3" 0 -1 0.0
"H32" "h1" 0 -1 0.0
"H33" "h1" 0 -1 0.0
"N2" "na" 0 -1 0.0
"C4" "cc" 0 -1 0.0
"H4" "h4" 0 -1 0.0
"C5" "cd" 0 -1 0.0
"H5" "h4" 0 -1 0.0
"N3" "nd" 0 -1 0.0
"C6" "cc" 0 -1 0.0
"H6" "h5" 0 -1 0.0
"C7" "ca" 0 -1 0.0
"C8" "ca" 0 -1 0.0
"H8" "ha" 0 -1 0.0
"C9" "ca" 0 -1 0.0
"H9" "ha" 0 -1 0.0
"C10" "ca" 0 -1 0.0
"CL1" "cl" 0 -1 0.0
"C11" "ca" 0 -1 0.0
"H11" "ha" 0 -1 0.0
"C12" "ca" 0 -1 0.0
"CL2" "cl" 0 -1 0.0
"C13" "ca" 0 -1 0.0
"C14" "ca" 0 -1 0.0
"H14" "ha" 0 -1 0.0
"C15" "ca" 0 -1 0.0
"H15" "ha" 0 -1 0.0
"C16" "ca" 0 -1 0.0
"C17" "ca" 0 -1 0.0
"H17" "ha" 0 -1 0.0
"C18" "ca" 0 -1 0.0
"H18" "ha" 0 -1 0.0
"C19" "cc" 0 -1 0.0
"N4" "nd" 0 -1 0.0
"N5" "nd" 0 -1 0.0
"C20" "cc" 0 -1 0.0
"O2" "os" 0 -1 0.0
"C21" "ca" 0 -1 0.0
"C22" "ca" 0 -1 0.0
"H22" "ha" 0 -1 0.0
"C23" "ca" 0 -1 0.0
"H23" "ha" 0 -1 0.0
"C24" "ca" 0 -1 0.0
"H24" "ha" 0 -1 0.0
"C25" "ca" 0 -1 0.0

"H25" "ha" 0 -1 0.0
"C26" "ca" 0 -1 0.0
"H26" "ha" 0 -1 0.0
!entry.VNI.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.VNI.unit.childsequence single int
2
!entry.VNI.unit.connect array int
0
0
!entry.VNI.unit.connectivity table int atom1x int atom2x int flags
1 2 2
2 3 1
2 29 1
3 4 1
3 5 1
5 6 1
5 7 1
5 18 1
7 8 1
7 9 1
7 10 1
10 11 4
10 16 4
11 12 1
11 13 4
13 14 1
13 15 4
15 16 4
16 17 1
18 19 4
18 27 4
19 20 1
19 21 4
21 22 1
21 23 4
23 24 1
23 25 4
25 26 1
25 27 4
27 28 1
29 30 4
29 37 4
30 31 1
30 32 4
32 33 1
32 34 4
34 35 4
34 39 1
35 36 1
35 37 4
37 38 1

39 40 4
39 43 4
40 41 4
41 42 4
42 43 4
42 44 1
44 45 4
44 53 4
45 46 1
45 47 4
47 48 1
47 49 4
49 50 1
49 51 4
51 52 1
51 53 4
53 54 1
!entry.VNI.unit.hierarchy table str abovetype int abovex str belowtype int belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
"R" 1 "A" 18
"R" 1 "A" 19
"R" 1 "A" 20
"R" 1 "A" 21
"R" 1 "A" 22
"R" 1 "A" 23
"R" 1 "A" 24
"R" 1 "A" 25
"R" 1 "A" 26
"R" 1 "A" 27
"R" 1 "A" 28
"R" 1 "A" 29
"R" 1 "A" 30
"R" 1 "A" 31
"R" 1 "A" 32
"R" 1 "A" 33
"R" 1 "A" 34
"R" 1 "A" 35
"R" 1 "A" 36
"R" 1 "A" 37

"R" 1 "A" 38
"R" 1 "A" 39
"R" 1 "A" 40
"R" 1 "A" 41
"R" 1 "A" 42
"R" 1 "A" 43
"R" 1 "A" 44
"R" 1 "A" 45
"R" 1 "A" 46
"R" 1 "A" 47
"R" 1 "A" 48
"R" 1 "A" 49
"R" 1 "A" 50
"R" 1 "A" 51
"R" 1 "A" 52
"R" 1 "A" 53
"R" 1 "A" 54
!entry.VNI.unit.name single str
"VNI"
!entry.VNI.unit.positions table dbl x dbl y dbl z
-0.922000 2.279000 0.821000
-0.726000 1.185000 0.359000
-1.732000 0.424000 -0.131000
-1.525000 -0.444000 -0.565000
-3.126000 0.830000 -0.006000
-3.236000 1.285000 0.963000
-3.466000 1.908000 -1.053000
-2.722000 2.687000 -0.981000
-3.426000 1.504000 -2.055000
-4.772000 2.491000 -0.847000
-5.106000 3.436000 0.094000
-4.383000 3.843000 0.768000
-6.426000 3.668000 -0.073000
-7.052000 4.344000 0.471000
-6.921000 2.896000 -1.091000
-5.912000 2.217000 -1.518000
-5.938000 1.506000 -2.320000
-4.017000 -0.399000 -0.097000
-4.085000 -1.117000 -1.290000
-3.518000 -0.780000 -2.140000
-4.859000 -2.251000 -1.431000
-4.891000 -2.782000 -2.363000
-5.599000 -2.690000 -0.349000
-6.587000 -4.111000 -0.497000
-5.560000 -2.013000 0.852000
-6.134000 -2.356000 1.690000
-4.769000 -0.881000 0.969000
-4.756000 -0.100000 2.526000
0.659000 0.609000 0.276000
0.916000 -0.759000 0.267000
0.116000 -1.471000 0.348000
2.213000 -1.227000 0.199000
2.415000 -2.281000 0.202000
3.272000 -0.327000 0.134000
3.021000 1.039000 0.154000
3.836000 1.737000 0.112000

```
 1.721000 1.502000 0.233000
 1.518000 2.555000 0.265000
 4.642000 -0.843000 0.051000
 5.028000 -2.050000 0.003000
 6.400000 -2.023000 -0.068000
 6.742000 -0.799000 -0.057000
 5.675000 0.013000 0.017000
 8.089000 -0.224000 -0.113000
 8.279000 1.152000 -0.085000
 7.434000 1.811000 -0.021000
 9.561000 1.672000 -0.139000
 9.706000 2.737000 -0.117000
 10.652000 0.823000 -0.221000
 11.647000 1.229000 -0.263000
 10.463000 -0.551000 -0.249000
 11.309000 -1.211000 -0.313000
 9.186000 -1.077000 -0.196000
 9.028000 -2.139000 -0.217000
!entry.VNI.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int c5x  int c6x
 0 0 0 0 0 0
!entry.VNI.unit.residues table  str name  int seq  int childseq  int startatomx  str restype  int imagingx
 "VNI" 1 55 1 "?" 0
!entry.VNI.unit.residuesPdbSequenceNumber array int
 0
!entry.VNI.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.VNI.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
```

0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

The frcmod file for parameters of ligand VNI is:

Force field parameter for VNI ligand
MASS

BOND

ANGLE

DIHE

ca-ca-cc-nd	1	2.550	180.000	2.000	same as X -c2-ca-X
ca-ca-cc-os	1	2.550	180.000	2.000	same as X -c2-ca-X
ca-cc-os-cc	1	1.050	180.000	2.000	same as X -c2-os-X
cc-os-cc-nd	1	1.050	180.000	2.000	same as X -c2-os-X

IMPROPER

ca-n -c -o	10.5	180.0	2.0	General improper torsional angle (2 general atom types)
c3-cc-na-cc	1.1	180.0	2.0	Using default value
cd-h4-cc-na	1.1	180.0	2.0	Using default value
cc-h4-cd-nd	1.1	180.0	2.0	Using default value
h5-na-cc-nd	1.1	180.0	2.0	Using default value
ca-ca-ca-ha	1.1	180.0	2.0	General improper torsional angle (2 general atom types)
c -ca-ca-ca	1.1	180.0	2.0	Using default value
ca-ca-ca-cc	1.1	180.0	2.0	Using default value
ca-nd-cc-os	1.1	180.0	2.0	Using default value

NONBON

The lib file for parameters of ligand MCP(LNP in PDB files) is:

```
!!index array str
 "LNP"
 "lnp"
!entry.LNP.unit.atoms table  str name  str type  int typex  int resx  int flags  int seq  int elmnt  dbl chg
 "CAA" "c3" 0 1 131072 1 6 -0.079100
 "HAA2" "hc" 0 1 131072 2 1 0.050600
 "HAA3" "hc" 0 1 131072 3 1 0.050600
 "CAB" "c3" 0 1 131072 4 6 -0.039700
 "HAB" "hc" 0 1 131072 5 1 0.101300
 "CAC" "c2" 0 1 131072 6 6 -0.114100
 "CAD" "c3" 0 1 131072 7 6 0.004400
 "CAE" "c3" 0 1 131072 8 6 0.088100
 "CAF" "c3" 0 1 131072 9 6 -0.193100
 "HAF2" "hc" 0 1 131072 10 1 0.068600
 "HAF3" "hc" 0 1 131072 11 1 0.068600
 "CAQ" "c3" 0 1 131072 12 6 0.019700
 "HAQ" "hc" 0 1 131072 13 1 0.030900
 "CAP" "c3" 0 1 131072 14 6 -0.201200
 "HAP2" "hc" 0 1 131072 15 1 0.061100
 "HAP3" "hc" 0 1 131072 16 1 0.061100
 "CAO" "c3" 0 1 131072 17 6 -0.051400
 "HAO2" "hc" 0 1 131072 18 1 0.062300
 "HAO3" "hc" 0 1 131072 19 1 0.062300
 "CAV" "c3" 0 1 131072 20 6 0.090200
 "HAV" "hc" 0 1 131072 21 1 0.007700
 "CAY" "c3" 0 1 131072 22 6 -0.155100
 "HAY1" "hc" 0 1 131072 23 1 0.032000
 "HAY2" "hc" 0 1 131072 24 1 0.032000
 "HAY3" "hc" 0 1 131072 25 1 0.032000
 "CAZ" "c3" 0 1 131072 26 6 -0.028300
 "HAZ2" "hc" 0 1 131072 27 1 0.003800
 "HAZ3" "hc" 0 1 131072 28 1 0.003800
 "CBA" "c3" 0 1 131072 29 6 0.004200
 "HBA2" "hc" 0 1 131072 30 1 0.005800
 "HBA3" "hc" 0 1 131072 31 1 0.005800
 "CBB" "c3" 0 1 131072 32 6 -0.103500
 "HBB2" "hc" 0 1 131072 33 1 0.025000
 "HBB3" "hc" 0 1 131072 34 1 0.025000
 "CBC" "c3" 0 1 131072 35 6 0.214500
 "HBC" "hc" 0 1 131072 36 1 -0.013800
 "CBD" "c3" 0 1 131072 37 6 -0.241100
 "HBD1" "hc" 0 1 131072 38 1 0.053300
 "HBD2" "hc" 0 1 131072 39 1 0.053300
 "HBD3" "hc" 0 1 131072 40 1 0.053300
 "CBE" "c3" 0 1 131072 41 6 -0.241100
 "HBE1" "hc" 0 1 131072 42 1 0.053300
 "HBE2" "hc" 0 1 131072 43 1 0.053300
 "HBE3" "hc" 0 1 131072 44 1 0.053300
 "CAS" "c3" 0 1 131072 45 6 -0.087300
 "HAS1" "hc" 0 1 131072 46 1 0.019800
```

"HAS2" "hc" 0 1 131072 47 1 0.019800
"HAS3" "hc" 0 1 131072 48 1 0.019800
"CAR" "c2" 0 1 131072 49 6 -0.310200
"HAR" "ha" 0 1 131072 50 1 0.147900
"CBF" "cu" 0 1 131072 51 6 0.180600
"CBG" "cx" 0 1 131072 52 6 -0.301700
"HBG2" "hc" 0 1 131072 53 1 0.132400
"HBG3" "hc" 0 1 131072 54 1 0.132400
"CBH" "cx" 0 1 131072 55 6 -0.301700
"HBH2" "hc" 0 1 131072 56 1 0.132400
"HBH3" "hc" 0 1 131072 57 1 0.132400
"CAJ" "c2" 0 1 131072 58 6 -0.141500
"HAJ" "ha" 0 1 131072 59 1 0.079600
"CAI" "c3" 0 1 131072 60 6 -0.061100
"HAI2" "hc" 0 1 131072 61 1 0.032000
"HAI3" "hc" 0 1 131072 62 1 0.032000
"CAH" "c3" 0 1 131072 63 6 0.012300
"HAH" "hc" 0 1 131072 64 1 0.049400
"CAG" "c3" 0 1 131072 65 6 0.038500
"CAK" "c3" 0 1 131072 66 6 -0.044500
"HAK2" "hc" 0 1 131072 67 1 0.028400
"HAK3" "hc" 0 1 131072 68 1 0.028400
"CAL" "c3" 0 1 131072 69 6 -0.006400
"HAL2" "hc" 0 1 131072 70 1 0.018700
"HAL3" "hc" 0 1 131072 71 1 0.018700
"CAM" "c3" 0 1 131072 72 6 0.084100
"HAM" "h1" 0 1 131072 73 1 0.064600
"CAN" "c3" 0 1 131072 74 6 0.178500
"CAW" "c3" 0 1 131072 75 6 -0.235700
"HAW1" "hc" 0 1 131072 76 1 0.062800
"HAW2" "hc" 0 1 131072 77 1 0.062800
"HAW3" "hc" 0 1 131072 78 1 0.062800
"CAX" "c3" 0 1 131072 79 6 -0.235700
"HAX1" "hc" 0 1 131072 80 1 0.062800
"HAX2" "hc" 0 1 131072 81 1 0.062800
"HAX3" "hc" 0 1 131072 82 1 0.062800
"OBH" "oh" 0 1 131072 83 8 -0.667800
"HOBH" "ho" 0 1 131072 84 1 0.432300
"CAT" "c3" 0 1 131072 85 6 -0.285700
"HAT1" "hc" 0 1 131072 86 1 0.072500
"HAT2" "hc" 0 1 131072 87 1 0.072500
"HAT3" "hc" 0 1 131072 88 1 0.072500
!entry.LNP.unit.atomspertinfo table str pname str ptype int ptypex int pelmnt dbl pchg
"CAA" "c3" 0 -1 0.0
"HAA2" "hc" 0 -1 0.0
"HAA3" "hc" 0 -1 0.0
"CAB" "c3" 0 -1 0.0
"HAB" "hc" 0 -1 0.0
"CAC" "c2" 0 -1 0.0
"CAD" "c3" 0 -1 0.0
"CAE" "c3" 0 -1 0.0
"CAF" "c3" 0 -1 0.0
"HAF2" "hc" 0 -1 0.0
"HAF3" "hc" 0 -1 0.0
"CAQ" "c3" 0 -1 0.0
"HAQ" "hc" 0 -1 0.0

"CAP" "c3" 0 -1 0.0
"HAP2" "hc" 0 -1 0.0
"HAP3" "hc" 0 -1 0.0
"CAO" "c3" 0 -1 0.0
"HAO2" "hc" 0 -1 0.0
"HAO3" "hc" 0 -1 0.0
"CAV" "c3" 0 -1 0.0
"HAV" "hc" 0 -1 0.0
"CAY" "c3" 0 -1 0.0
"HAY1" "hc" 0 -1 0.0
"HAY2" "hc" 0 -1 0.0
"HAY3" "hc" 0 -1 0.0
"CAZ" "c3" 0 -1 0.0
"HAZ2" "hc" 0 -1 0.0
"HAZ3" "hc" 0 -1 0.0
"CBA" "c3" 0 -1 0.0
"HBA2" "hc" 0 -1 0.0
"HBA3" "hc" 0 -1 0.0
"CBB" "c3" 0 -1 0.0
"HBB2" "hc" 0 -1 0.0
"HBB3" "hc" 0 -1 0.0
"CBC" "c3" 0 -1 0.0
"HBC" "hc" 0 -1 0.0
"CBD" "c3" 0 -1 0.0
"HBD1" "hc" 0 -1 0.0
"HBD2" "hc" 0 -1 0.0
"HBD3" "hc" 0 -1 0.0
"CBE" "c3" 0 -1 0.0
"HBE1" "hc" 0 -1 0.0
"HBE2" "hc" 0 -1 0.0
"HBE3" "hc" 0 -1 0.0
"CAS" "c3" 0 -1 0.0
"HAS1" "hc" 0 -1 0.0
"HAS2" "hc" 0 -1 0.0
"HAS3" "hc" 0 -1 0.0
"CAR" "c2" 0 -1 0.0
"HAR" "ha" 0 -1 0.0
"CBF" "cu" 0 -1 0.0
"CBG" "cx" 0 -1 0.0
"HBG2" "hc" 0 -1 0.0
"HBG3" "hc" 0 -1 0.0
"CBH" "cx" 0 -1 0.0
"HBH2" "hc" 0 -1 0.0
"HBH3" "hc" 0 -1 0.0
"CAJ" "c2" 0 -1 0.0
"HAJ" "ha" 0 -1 0.0
"CAI" "c3" 0 -1 0.0
"HAI2" "hc" 0 -1 0.0
"HAI3" "hc" 0 -1 0.0
"CAH" "c3" 0 -1 0.0
"HAH" "hc" 0 -1 0.0
"CAG" "c3" 0 -1 0.0
"CAK" "c3" 0 -1 0.0
"HAK2" "hc" 0 -1 0.0
"HAK3" "hc" 0 -1 0.0
"CAL" "c3" 0 -1 0.0

"HAL2" "hc" 0 -1 0.0
"HAL3" "hc" 0 -1 0.0
"CAM" "c3" 0 -1 0.0
"HAM" "h1" 0 -1 0.0
"CAN" "c3" 0 -1 0.0
"CAW" "c3" 0 -1 0.0
"HAW1" "hc" 0 -1 0.0
"HAW2" "hc" 0 -1 0.0
"HAW3" "hc" 0 -1 0.0
"CAX" "c3" 0 -1 0.0
"HAX1" "hc" 0 -1 0.0
"HAX2" "hc" 0 -1 0.0
"HAX3" "hc" 0 -1 0.0
"OBH" "oh" 0 -1 0.0
"HOBH" "ho" 0 -1 0.0
"CAT" "c3" 0 -1 0.0
"HAT1" "hc" 0 -1 0.0
"HAT2" "hc" 0 -1 0.0
"HAT3" "hc" 0 -1 0.0
!entry.LNP.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.LNP.unit.childsequence single int
2
!entry.LNP.unit.connect array int
0
0
!entry.LNP.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
1 4 1
1 9 1
4 5 1
4 6 1
4 65 1
6 7 1
6 58 2
7 8 1
7 17 1
7 49 1
8 9 1
8 12 1
8 45 1
9 10 1
9 11 1
12 13 1
12 14 1
12 20 1
14 15 1
14 16 1
14 17 1
17 18 1
17 19 1

20 21 1
20 22 1
20 26 1
22 23 1
22 24 1
22 25 1
26 27 1
26 28 1
26 29 1
29 30 1
29 31 1
29 32 1
32 33 1
32 34 1
32 35 1
35 36 1
35 37 1
35 41 1
37 38 1
37 39 1
37 40 1
41 42 1
41 43 1
41 44 1
45 46 1
45 47 1
45 48 1
49 50 1
49 51 2
51 52 1
51 55 1
52 53 1
52 54 1
52 55 1
55 56 1
55 57 1
58 59 1
58 60 1
60 61 1
60 62 1
60 63 1
63 64 1
63 65 1
63 74 1
65 66 1
65 85 1
66 67 1
66 68 1
66 69 1
69 70 1
69 71 1
69 72 1
72 73 1
72 74 1
72 83 1
74 75 1

74 79 1
75 76 1
75 77 1
75 78 1
79 80 1
79 81 1
79 82 1
83 84 1
85 86 1
85 87 1
85 88 1
!entry.LNP.unit.hierarchy table str abovetype int abovex str belowtype int belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
"R" 1 "A" 18
"R" 1 "A" 19
"R" 1 "A" 20
"R" 1 "A" 21
"R" 1 "A" 22
"R" 1 "A" 23
"R" 1 "A" 24
"R" 1 "A" 25
"R" 1 "A" 26
"R" 1 "A" 27
"R" 1 "A" 28
"R" 1 "A" 29
"R" 1 "A" 30
"R" 1 "A" 31
"R" 1 "A" 32
"R" 1 "A" 33
"R" 1 "A" 34
"R" 1 "A" 35
"R" 1 "A" 36
"R" 1 "A" 37
"R" 1 "A" 38
"R" 1 "A" 39
"R" 1 "A" 40
"R" 1 "A" 41
"R" 1 "A" 42
"R" 1 "A" 43

```
 "R" 1 "A" 44
 "R" 1 "A" 45
 "R" 1 "A" 46
 "R" 1 "A" 47
 "R" 1 "A" 48
 "R" 1 "A" 49
 "R" 1 "A" 50
 "R" 1 "A" 51
 "R" 1 "A" 52
 "R" 1 "A" 53
 "R" 1 "A" 54
 "R" 1 "A" 55
 "R" 1 "A" 56
 "R" 1 "A" 57
 "R" 1 "A" 58
 "R" 1 "A" 59
 "R" 1 "A" 60
 "R" 1 "A" 61
 "R" 1 "A" 62
 "R" 1 "A" 63
 "R" 1 "A" 64
 "R" 1 "A" 65
 "R" 1 "A" 66
 "R" 1 "A" 67
 "R" 1 "A" 68
 "R" 1 "A" 69
 "R" 1 "A" 70
 "R" 1 "A" 71
 "R" 1 "A" 72
 "R" 1 "A" 73
 "R" 1 "A" 74
 "R" 1 "A" 75
 "R" 1 "A" 76
 "R" 1 "A" 77
 "R" 1 "A" 78
 "R" 1 "A" 79
 "R" 1 "A" 80
 "R" 1 "A" 81
 "R" 1 "A" 82
 "R" 1 "A" 83
 "R" 1 "A" 84
 "R" 1 "A" 85
 "R" 1 "A" 86
 "R" 1 "A" 87
 "R" 1 "A" 88
!entry.LNP.unit.name single str
 "LNP"
!entry.LNP.unit.positions table  dbl x  dbl y  dbl z
 1.037000 -1.069000 -1.605000
 1.377000 -1.879000 -2.235000
 1.179000 -0.163000 -2.180000
 1.872000 -0.987000 -0.309000
 1.679000 -1.913000 0.229000
 1.336000 0.147000 0.562000
 -0.094000 0.604000 0.292000
 -1.005000 -0.618000 -0.050000
```

-0.472000 -1.296000 -1.342000
-1.024000 -0.955000 -2.212000
-0.647000 -2.365000 -1.283000
-2.411000 0.059000 -0.109000
-2.466000 0.614000 -1.041000
-2.347000 1.093000 1.058000
-2.769000 2.042000 0.751000
-2.932000 0.756000 1.909000
-0.861000 1.253000 1.457000
-0.590000 2.289000 1.608000
-0.651000 0.732000 2.382000
-3.680000 -0.825000 -0.065000
-3.700000 -1.352000 0.887000
-3.713000 -1.870000 -1.187000
-2.874000 -2.551000 -1.137000
-3.694000 -1.393000 -2.164000
-4.612000 -2.474000 -1.136000
-4.941000 0.065000 -0.124000
-4.850000 0.864000 0.604000
-4.984000 0.550000 -1.099000
-6.267000 -0.657000 0.140000
-6.448000 -1.400000 -0.629000
-6.199000 -1.198000 1.083000
-7.450000 0.314000 0.202000
-7.225000 1.087000 0.935000
-7.547000 0.826000 -0.755000
-8.802000 -0.319000 0.569000
-8.675000 -0.852000 1.510000
-9.857000 0.770000 0.788000
-9.550000 1.468000 1.562000
-10.809000 0.341000 1.088000
-10.024000 1.339000 -0.123000
-9.289000 -1.325000 -0.480000
-10.259000 -1.728000 -0.205000
-8.610000 -2.163000 -0.590000
-9.393000 -0.850000 -1.453000
-1.004000 -1.648000 1.105000
-1.488000 -1.268000 1.996000
-0.008000 -1.954000 1.391000
-1.543000 -2.541000 0.809000
-0.083000 1.605000 -0.884000
-0.530000 1.294000 -1.811000
0.424000 2.813000 -0.856000
0.694000 4.047000 -1.596000
1.461000 4.042000 -2.354000
-0.122000 4.721000 -1.800000
1.128000 3.878000 -0.144000
0.600000 4.445000 0.606000
2.179000 3.756000 0.059000
2.095000 0.694000 1.488000
1.702000 1.478000 2.114000
3.539000 0.368000 1.711000
3.689000 0.162000 2.766000
4.095000 1.282000 1.523000
4.126000 -0.811000 0.892000
3.902000 -1.715000 1.454000

```
 3.428000 -0.973000 -0.505000
 3.894000 0.171000 -1.428000
 3.493000 0.034000 -2.427000
 3.493000 1.113000 -1.065000
 5.421000 0.270000 -1.520000
 5.836000 -0.592000 -2.031000
 5.686000 1.134000 -2.129000
 6.094000 0.415000 -0.156000
 5.793000 1.374000 0.262000
 5.701000 -0.687000 0.852000
 6.244000 -0.298000 2.247000
 7.325000 -0.262000 2.217000
 5.954000 -1.037000 2.989000
 5.901000 0.670000 2.588000
 6.416000 -2.018000 0.512000
 7.430000 -1.995000 0.889000
 5.908000 -2.857000 0.976000
 6.497000 -2.216000 -0.545000
 7.498000 0.404000 -0.285000
 7.767000 1.124000 -0.840000
 3.779000 -2.337000 -1.140000
 3.224000 -2.493000 -2.055000
 4.817000 -2.434000 -1.401000
 3.535000 -3.151000 -0.464000
!entry.LNP.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int c5x  int c6x
 0 0 0 0 0 0
!entry.LNP.unit.residues table  str name  int seq  int childseq  int startatomx  str restype  int imagingx
 "LNP" 1 89 1 "?" 0
!entry.LNP.unit.residuesPdbSequenceNumber array int
 0
!entry.LNP.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.LNP.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
```

```
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
```

0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

The frcmod file for parameters of ligand MCP (LNP in PDB files) is:

remark goes here
MASS

BOND

ANGLE
c3-c2-cu 64.330 123.420 same as c2-c2-c3
ha-c2-cu 50.040 120.940 same as c2-c2-ha
cx-cu-cx 62.700 116.520 same as c3-c2-c3

DIHE
c3-c2-cu-cx 1 6.650 180.000 2.000 same as X -c2-c2-X
c2-cu-cx-hc 1 0.380 180.000 -3.000 same as hc-c3-c2-c2
c2-cu-cx-hc 1 1.150 0.000 1.000 same as hc-c3-c2-c2
c2-cu-cx-cx 1 0.000 0.000 2.000 same as X -c2-c3-X
ha-c2-cu-cx 1 6.650 180.000 2.000 same as X -c2-c2-X
cx-cu-cx-cx 1 0.000 0.000 2.000 same as X -c2-c3-X
cx-cu-cx-hc 1 0.000 0.000 2.000 same as X -c2-c3-X

IMPROPER
c2-c3-c2-c3 1.1 180.0 2.0 Using default value
c3-cu-c2-ha 1.1 180.0 2.0 Using default value
c2-c3-c2-ha 1.1 180.0 2.0 Using default value

NONBON

The lib file for parameters of ligand OBT (substrate of *T. brucei* CYP51) is:

!!index array str
"OBT"
!entry.OBT.unit.atoms table str name str type int typex int resx int flags int seq int elmnt dbl chg
"O1" "oh" 0 1 131072 1 8 -0.684400
"H41" "ho" 0 1 131072 2 1 0.425200
"C20" "c3" 0 1 131072 3 6 0.177900
"H28" "h1" 0 1 131072 4 1 0.030300
"C17" "c3" 0 1 131072 5 6 0.067600
"H22" "hc" 0 1 131072 6 1 0.070700
"C10" "c3" 0 1 131072 7 6 -0.012700
"H8" "hc" 0 1 131072 8 1 0.012400

"C5" "c3" 0 1 131072 9 6 0.018800
"C8" "c2" 0 1 131072 10 6 -0.024000
"C4" "c2" 0 1 131072 11 6 -0.010400
"C2" "c3" 0 1 131072 12 6 0.083700
"C1" "c3" 0 1 131072 13 6 0.132700
"C3" "c3" 0 1 131072 14 6 0.089200
"H1" "hc" 0 1 131072 15 1 -0.011000
"C9" "c3" 0 1 131072 16 6 -0.037000
"H6" "hc" 0 1 131072 17 1 0.016100
"H7" "hc" 0 1 131072 18 1 0.016100
"C6" "c3" 0 1 131072 19 6 -0.086200
"H2" "hc" 0 1 131072 20 1 0.018100
"H3" "hc" 0 1 131072 21 1 0.018100
"C13" "c3" 0 1 131072 22 6 0.044500
"H13" "hc" 0 1 131072 23 1 -0.004500
"C22" "c3" 0 1 131072 24 6 -0.025600
"H31" "hc" 0 1 131072 25 1 -1.000000E-04
"H32" "hc" 0 1 131072 26 1 -1.000000E-04
"C25" "c3" 0 1 131072 27 6 -0.020800
"H39" "hc" 0 1 131072 28 1 0.027600
"H40" "hc" 0 1 131072 29 1 0.027600
"C26" "c2" 0 1 131072 30 6 -0.018800
"C27" "c3" 0 1 131072 31 6 0.190600
"H42" "hc" 0 1 131072 32 1 0.025900
"C29" "c3" 0 1 131072 33 6 -0.207000
"H45" "hc" 0 1 131072 34 1 0.049100
"H46" "hc" 0 1 131072 35 1 0.049100
"H47" "hc" 0 1 131072 36 1 0.049100
"C30" "c3" 0 1 131072 37 6 -0.207000
"H48" "hc" 0 1 131072 38 1 0.049100
"H49" "hc" 0 1 131072 39 1 0.049100
"H50" "hc" 0 1 131072 40 1 0.049100
"C28" "c2" 0 1 131072 41 6 -0.458400
"H43" "ha" 0 1 131072 42 1 0.166200
"H44" "ha" 0 1 131072 43 1 0.166200
"C23" "c3" 0 1 131072 44 6 -0.072900
"H33" "hc" 0 1 131072 45 1 0.009900
"H34" "hc" 0 1 131072 46 1 0.009900
"H35" "hc" 0 1 131072 47 1 0.009900
"C7" "c3" 0 1 131072 48 6 -0.093300
"H4" "hc" 0 1 131072 49 1 0.014400
"H5" "hc" 0 1 131072 50 1 0.014400
"C11" "c3" 0 1 131072 51 6 -0.098300
"H9" "hc" 0 1 131072 52 1 0.061400
"H10" "hc" 0 1 131072 53 1 0.061400
"C15" "c3" 0 1 131072 54 6 -0.062100
"H16" "hc" 0 1 131072 55 1 0.002700
"H17" "hc" 0 1 131072 56 1 0.002700
"H18" "hc" 0 1 131072 57 1 0.002700
"C16" "c3" 0 1 131072 58 6 -0.201500
"H19" "hc" 0 1 131072 59 1 0.044500
"H20" "hc" 0 1 131072 60 1 0.044500
"H21" "hc" 0 1 131072 61 1 0.044500
"C12" "c3" 0 1 131072 62 6 -0.080700
"H11" "hc" 0 1 131072 63 1 0.047700
"H12" "hc" 0 1 131072 64 1 0.047700

"C14" "c3" 0 1 131072 65 6 -0.088700
"H14" "hc" 0 1 131072 66 1 0.028400
"H15" "hc" 0 1 131072 67 1 0.028400
"C18" "c3" 0 1 131072 68 6 -0.033900
"H23" "hc" 0 1 131072 69 1 0.006900
"H24" "hc" 0 1 131072 70 1 0.006900
"C21" "c3" 0 1 131072 71 6 -0.071200
"H29" "hc" 0 1 131072 72 1 0.029800
"H30" "hc" 0 1 131072 73 1 0.029800
"C19" "c3" 0 1 131072 74 6 -0.135700
"H25" "hc" 0 1 131072 75 1 0.037200
"H26" "hc" 0 1 131072 76 1 0.037200
"H27" "hc" 0 1 131072 77 1 0.037200
"C24" "c3" 0 1 131072 78 6 -0.141600
"H36" "hc" 0 1 131072 79 1 0.035900
"H37" "hc" 0 1 131072 80 1 0.035900
"H38" "hc" 0 1 131072 81 1 0.035900
!entry.OBT.unit.atomspertinfo table str pname str ptype int ptypex int pelmnt dbl pchg
"O1" "oh" 0 -1 0.0
"H41" "ho" 0 -1 0.0
"C20" "c3" 0 -1 0.0
"H28" "h1" 0 -1 0.0
"C17" "c3" 0 -1 0.0
"H22" "hc" 0 -1 0.0
"C10" "c3" 0 -1 0.0
"H8" "hc" 0 -1 0.0
"C5" "c3" 0 -1 0.0
"C8" "c2" 0 -1 0.0
"C4" "c2" 0 -1 0.0
"C2" "c3" 0 -1 0.0
"C1" "c3" 0 -1 0.0
"C3" "c3" 0 -1 0.0
"H1" "hc" 0 -1 0.0
"C9" "c3" 0 -1 0.0
"H6" "hc" 0 -1 0.0
"H7" "hc" 0 -1 0.0
"C6" "c3" 0 -1 0.0
"H2" "hc" 0 -1 0.0
"H3" "hc" 0 -1 0.0
"C13" "c3" 0 -1 0.0
"H13" "hc" 0 -1 0.0
"C22" "c3" 0 -1 0.0
"H31" "hc" 0 -1 0.0
"H32" "hc" 0 -1 0.0
"C25" "c3" 0 -1 0.0
"H39" "hc" 0 -1 0.0
"H40" "hc" 0 -1 0.0
"C26" "c2" 0 -1 0.0
"C27" "c3" 0 -1 0.0
"H42" "hc" 0 -1 0.0
"C29" "c3" 0 -1 0.0
"H45" "hc" 0 -1 0.0
"H46" "hc" 0 -1 0.0
"H47" "hc" 0 -1 0.0
"C30" "c3" 0 -1 0.0
"H48" "hc" 0 -1 0.0

"H49" "hc" 0 -1 0.0
"H50" "hc" 0 -1 0.0
"C28" "c2" 0 -1 0.0
"H43" "ha" 0 -1 0.0
"H44" "ha" 0 -1 0.0
"C23" "c3" 0 -1 0.0
"H33" "hc" 0 -1 0.0
"H34" "hc" 0 -1 0.0
"H35" "hc" 0 -1 0.0
"C7" "c3" 0 -1 0.0
"H4" "hc" 0 -1 0.0
"H5" "hc" 0 -1 0.0
"C11" "c3" 0 -1 0.0
"H9" "hc" 0 -1 0.0
"H10" "hc" 0 -1 0.0
"C15" "c3" 0 -1 0.0
"H16" "hc" 0 -1 0.0
"H17" "hc" 0 -1 0.0
"H18" "hc" 0 -1 0.0
"C16" "c3" 0 -1 0.0
"H19" "hc" 0 -1 0.0
"H20" "hc" 0 -1 0.0
"H21" "hc" 0 -1 0.0
"C12" "c3" 0 -1 0.0
"H11" "hc" 0 -1 0.0
"H12" "hc" 0 -1 0.0
"C14" "c3" 0 -1 0.0
"H14" "hc" 0 -1 0.0
"H15" "hc" 0 -1 0.0
"C18" "c3" 0 -1 0.0
"H23" "hc" 0 -1 0.0
"H24" "hc" 0 -1 0.0
"C21" "c3" 0 -1 0.0
"H29" "hc" 0 -1 0.0
"H30" "hc" 0 -1 0.0
"C19" "c3" 0 -1 0.0
"H25" "hc" 0 -1 0.0
"H26" "hc" 0 -1 0.0
"H27" "hc" 0 -1 0.0
"C24" "c3" 0 -1 0.0
"H36" "hc" 0 -1 0.0
"H37" "hc" 0 -1 0.0
"H38" "hc" 0 -1 0.0
!entry.OBT.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.OBT.unit.childsequence single int
2
!entry.OBT.unit.connect array int
0
0
!entry.OBT.unit.connectivity table int atom1x int atom2x int flags
1 2 1

1 3 1
3 4 1
3 5 1
3 71 1
5 6 1
5 7 1
5 78 1
7 8 1
7 9 1
7 65 1
9 10 1
9 68 1
9 74 1
10 11 2
10 51 1
11 12 1
11 62 1
12 13 1
12 19 1
12 58 1
13 14 1
13 48 1
13 54 1
14 15 1
14 16 1
14 22 1
16 17 1
16 18 1
16 19 1
19 20 1
19 21 1
22 23 1
22 24 1
22 44 1
24 25 1
24 26 1
24 27 1
27 28 1
27 29 1
27 30 1
30 31 1
30 41 2
31 32 1
31 33 1
31 37 1
33 34 1
33 35 1
33 36 1
37 38 1
37 39 1
37 40 1
41 42 1
41 43 1
44 45 1
44 46 1
44 47 1

48 49 1
48 50 1
48 51 1
51 52 1
51 53 1
54 55 1
54 56 1
54 57 1
58 59 1
58 60 1
58 61 1
62 63 1
62 64 1
62 65 1
65 66 1
65 67 1
68 69 1
68 70 1
68 71 1
71 72 1
71 73 1
74 75 1
74 76 1
74 77 1
78 79 1
78 80 1
78 81 1
!entry.OBT.unit.hierarchy table str abovetype int abovex str belowtype int belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
"R" 1 "A" 18
"R" 1 "A" 19
"R" 1 "A" 20
"R" 1 "A" 21
"R" 1 "A" 22
"R" 1 "A" 23
"R" 1 "A" 24
"R" 1 "A" 25
"R" 1 "A" 26
"R" 1 "A" 27

"R" 1 "A" 28
"R" 1 "A" 29
"R" 1 "A" 30
"R" 1 "A" 31
"R" 1 "A" 32
"R" 1 "A" 33
"R" 1 "A" 34
"R" 1 "A" 35
"R" 1 "A" 36
"R" 1 "A" 37
"R" 1 "A" 38
"R" 1 "A" 39
"R" 1 "A" 40
"R" 1 "A" 41
"R" 1 "A" 42
"R" 1 "A" 43
"R" 1 "A" 44
"R" 1 "A" 45
"R" 1 "A" 46
"R" 1 "A" 47
"R" 1 "A" 48
"R" 1 "A" 49
"R" 1 "A" 50
"R" 1 "A" 51
"R" 1 "A" 52
"R" 1 "A" 53
"R" 1 "A" 54
"R" 1 "A" 55
"R" 1 "A" 56
"R" 1 "A" 57
"R" 1 "A" 58
"R" 1 "A" 59
"R" 1 "A" 60
"R" 1 "A" 61
"R" 1 "A" 62
"R" 1 "A" 63
"R" 1 "A" 64
"R" 1 "A" 65
"R" 1 "A" 66
"R" 1 "A" 67
"R" 1 "A" 68
"R" 1 "A" 69
"R" 1 "A" 70
"R" 1 "A" 71
"R" 1 "A" 72
"R" 1 "A" 73
"R" 1 "A" 74
"R" 1 "A" 75
"R" 1 "A" 76
"R" 1 "A" 77
"R" 1 "A" 78
"R" 1 "A" 79
"R" 1 "A" 80
"R" 1 "A" 81
!entry.OBT.unit.name single str
"OBT"

!entry.OBT.unit.positions table dbl x dbl y dbl z
 7.880000 -0.680000 -0.890000
 8.196000 -1.339000 -1.492000
 6.484000 -0.546000 -1.034000
 6.253000 -0.260000 -2.062000
 6.021000 0.588000 -0.116000
 6.350000 0.339000 0.890000
 4.478000 0.690000 -0.155000
 4.213000 0.876000 -1.198000
 3.750000 -0.630000 0.235000
 2.225000 -0.436000 0.084000
 1.650000 0.761000 0.166000
 0.145000 0.898000 -0.050000
 -0.627000 -0.315000 0.564000
 -2.109000 0.115000 0.348000
 -2.327000 -0.028000 -0.705000
 -2.066000 1.646000 0.644000
 -2.492000 1.860000 1.618000
 -2.662000 2.197000 -0.075000
 -0.582000 2.088000 0.604000
 -0.447000 3.011000 0.050000
 -0.219000 2.272000 1.607000
 -3.236000 -0.604000 1.126000
 -3.052000 -0.466000 2.189000
 -4.621000 0.035000 0.867000
 -4.613000 1.073000 1.179000
 -5.334000 -0.463000 1.519000
 -5.142000 -0.023000 -0.580000
 -5.115000 -1.047000 -0.942000
 -4.490000 0.551000 -1.229000
 -6.557000 0.515000 -0.696000
 -7.674000 -0.469000 -0.369000
 -7.297000 -1.133000 0.407000
 -7.996000 -1.339000 -1.595000
 -7.113000 -1.837000 -1.982000
 -8.723000 -2.106000 -1.341000
 -8.411000 -0.732000 -2.393000
 -8.949000 0.180000 0.174000
 -9.443000 0.788000 -0.577000
 -9.655000 -0.585000 0.484000
 -8.738000 0.811000 1.031000
 -6.773000 1.765000 -1.071000
 -7.756000 2.193000 -1.137000
 -5.957000 2.419000 -1.330000
 -3.294000 -2.117000 0.867000
 -2.465000 -2.637000 1.330000
 -3.274000 -2.351000 -0.193000
 -4.206000 -2.539000 1.280000
 -0.135000 -1.576000 -0.146000
 -0.454000 -1.559000 -1.181000
 -0.571000 -2.470000 0.283000
 1.404000 -1.714000 -0.074000
 1.729000 -2.245000 -0.962000
 1.654000 -2.371000 0.756000
 -0.351000 -0.458000 2.079000
 -0.816000 -1.362000 2.457000

```
 0.707000 -0.523000 2.294000
 -0.749000 0.368000 2.655000
 -0.048000 1.050000 -1.584000
 0.318000 0.194000 -2.135000
 -1.080000 1.221000 -1.869000
 0.518000 1.911000 -1.924000
 2.426000 2.047000 0.352000
 2.317000 2.649000 -0.549000
 1.973000 2.632000 1.146000
 3.912000 1.858000 0.649000
 4.427000 2.780000 0.410000
 4.068000 1.691000 1.711000
 4.263000 -1.734000 -0.720000
 3.925000 -1.509000 -1.730000
 3.837000 -2.695000 -0.458000
 5.786000 -1.862000 -0.720000
 6.146000 -2.218000 0.240000
 6.082000 -2.612000 -1.453000
 4.030000 -1.053000 1.698000
 3.575000 -2.016000 1.906000
 5.087000 -1.150000 1.910000
 3.617000 -0.343000 2.402000
 6.711000 1.897000 -0.520000
 7.774000 1.735000 -0.643000
 6.316000 2.274000 -1.461000
 6.581000 2.671000 0.228000
!entry.OBT.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int c5x  int c6x
 0 0 0 0 0 0
!entry.OBT.unit.residues table  str name  int seq  int childseq  int startatomx  str restype  int imagingx
 "OBT" 1 82 1 "?" 0
!entry.OBT.unit.residuesPdbSequenceNumber array int
 0
!entry.OBT.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.OBT.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
```

```
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
```

0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

The frcmod file for parameters of ligand OBT (substrate) is:

Obtusifoliol (natural substrate of both TB and human CYP51)
MASS

BOND

ANGLE

DIHE

IMPROPER
c2-c3-c2-c3 1.1 180.0 2.0 Using default value
c2-ha-c2-ha 1.1 180.0 2.0 Using default value

NONBON

The lib file for parameters of ligand OBT_DM (ODM, product of *T. brucei* CYP51) is:

!!index array str
"ODM"
!entry.ODM.unit.atoms table str name str type int typex int resx int flags int seq int elmnt dbl chg
"O1" "oh" 0 1 131072 1 8 -0.704400
"H35" "ho" 0 1 131072 2 1 0.439700
"C14" "c3" 0 1 131072 3 6 0.208500
"H14" "h1" 0 1 131072 4 1 0.020500
"C7" "c3" 0 1 131072 5 6 0.019300
"H3" "hc" 0 1 131072 6 1 0.063500
"C2" "c3" 0 1 131072 7 6 0.019100
"H1" "hc" 0 1 131072 8 1 0.017900
"C1" "c3" 0 1 131072 9 6 0.033600
"C4" "c2" 0 1 131072 10 6 -0.031700
"C6" "ce" 0 1 131072 11 6 -0.026000
"C8" "ce" 0 1 131072 12 6 -0.087300
"C3" "c3" 0 1 131072 13 6 0.071000
"C5" "c3" 0 1 131072 14 6 0.028600
"H2" "hc" 0 1 131072 15 1 0.019500
"C16" "c3" 0 1 131072 16 6 -0.023600
"H17" "hc" 0 1 131072 17 1 0.042100
"H18" "hc" 0 1 131072 18 1 0.042100
"C19" "c2" 0 1 131072 19 6 -0.220800
"H23" "ha" 0 1 131072 20 1 0.118700
"C18" "c3" 0 1 131072 21 6 0.052600
"H22" "hc" 0 1 131072 22 1 0.006900
"C22" "c3" 0 1 131072 23 6 -0.024500
"H30" "hc" 0 1 131072 24 1 -0.003100

"H31" "hc" 0 1 131072 25 1 -0.003100
"C24" "c3" 0 1 131072 26 6 -0.008100
"H36" "hc" 0 1 131072 27 1 0.025500
"H37" "hc" 0 1 131072 28 1 0.025500
"C25" "c2" 0 1 131072 29 6 -0.019700
"C26" "c3" 0 1 131072 30 6 0.202200
"H38" "hc" 0 1 131072 31 1 0.024500
"C28" "c3" 0 1 131072 32 6 -0.219000
"H41" "hc" 0 1 131072 33 1 0.051700
"H42" "hc" 0 1 131072 34 1 0.051700
"H43" "hc" 0 1 131072 35 1 0.051700
"C29" "c3" 0 1 131072 36 6 -0.219000
"H44" "hc" 0 1 131072 37 1 0.051700
"H45" "hc" 0 1 131072 38 1 0.051700
"H46" "hc" 0 1 131072 39 1 0.051700
"C27" "c2" 0 1 131072 40 6 -0.459200
"H39" "ha" 0 1 131072 41 1 0.166100
"H40" "ha" 0 1 131072 42 1 0.166100
"C23" "c3" 0 1 131072 43 6 -0.102100
"H32" "hc" 0 1 131072 44 1 0.020000
"H33" "hc" 0 1 131072 45 1 0.020000
"H34" "hc" 0 1 131072 46 1 0.020000
"C11" "c3" 0 1 131072 47 6 -0.068500
"H8" "hc" 0 1 131072 48 1 0.020200
"H9" "hc" 0 1 131072 49 1 0.020200
"C12" "c3" 0 1 131072 50 6 -0.043200
"H10" "hc" 0 1 131072 51 1 0.047900
"H11" "hc" 0 1 131072 52 1 0.047900
"C20" "c3" 0 1 131072 53 6 -0.056800
"H24" "hc" 0 1 131072 54 1 0.016200
"H25" "hc" 0 1 131072 55 1 0.016200
"H26" "hc" 0 1 131072 56 1 0.016200
"C13" "c3" 0 1 131072 57 6 -0.024100
"H12" "hc" 0 1 131072 58 1 0.032300
"H13" "hc" 0 1 131072 59 1 0.032300
"C10" "c3" 0 1 131072 60 6 -0.069900
"H6" "hc" 0 1 131072 61 1 0.031400
"H7" "hc" 0 1 131072 62 1 0.031400
"C9" "c3" 0 1 131072 63 6 -0.061800
"H4" "hc" 0 1 131072 64 1 0.018700
"H5" "hc" 0 1 131072 65 1 0.018700
"C15" "c3" 0 1 131072 66 6 -0.089200
"H15" "hc" 0 1 131072 67 1 0.060300
"H16" "hc" 0 1 131072 68 1 0.060300
"C17" "c3" 0 1 131072 69 6 -0.134000
"H19" "hc" 0 1 131072 70 1 0.038300
"H20" "hc" 0 1 131072 71 1 0.038300
"H21" "hc" 0 1 131072 72 1 0.038300
"C21" "c3" 0 1 131072 73 6 -0.207400
"H27" "hc" 0 1 131072 74 1 0.045900
"H28" "hc" 0 1 131072 75 1 0.045900
"H29" "hc" 0 1 131072 76 1 0.045900
!entry.ODM.unit.atomspertinfo table str pname str ptype int ptypex int pelmnt dbl pchg
"O1" "oh" 0 -1 0.0
"H35" "ho" 0 -1 0.0
"C14" "c3" 0 -1 0.0

"H14" "h1" 0 -1 0.0
"C7" "c3" 0 -1 0.0
"H3" "hc" 0 -1 0.0
"C2" "c3" 0 -1 0.0
"H1" "hc" 0 -1 0.0
"C1" "c3" 0 -1 0.0
"C4" "c2" 0 -1 0.0
"C6" "ce" 0 -1 0.0
"C8" "ce" 0 -1 0.0
"C3" "c3" 0 -1 0.0
"C5" "c3" 0 -1 0.0
"H2" "hc" 0 -1 0.0
"C16" "c3" 0 -1 0.0
"H17" "hc" 0 -1 0.0
"H18" "hc" 0 -1 0.0
"C19" "c2" 0 -1 0.0
"H23" "ha" 0 -1 0.0
"C18" "c3" 0 -1 0.0
"H22" "hc" 0 -1 0.0
"C22" "c3" 0 -1 0.0
"H30" "hc" 0 -1 0.0
"H31" "hc" 0 -1 0.0
"C24" "c3" 0 -1 0.0
"H36" "hc" 0 -1 0.0
"H37" "hc" 0 -1 0.0
"C25" "c2" 0 -1 0.0
"C26" "c3" 0 -1 0.0
"H38" "hc" 0 -1 0.0
"C28" "c3" 0 -1 0.0
"H41" "hc" 0 -1 0.0
"H42" "hc" 0 -1 0.0
"H43" "hc" 0 -1 0.0
"C29" "c3" 0 -1 0.0
"H44" "hc" 0 -1 0.0
"H45" "hc" 0 -1 0.0
"H46" "hc" 0 -1 0.0
"C27" "c2" 0 -1 0.0
"H39" "ha" 0 -1 0.0
"H40" "ha" 0 -1 0.0
"C23" "c3" 0 -1 0.0
"H32" "hc" 0 -1 0.0
"H33" "hc" 0 -1 0.0
"H34" "hc" 0 -1 0.0
"C11" "c3" 0 -1 0.0
"H8" "hc" 0 -1 0.0
"H9" "hc" 0 -1 0.0
"C12" "c3" 0 -1 0.0
"H10" "hc" 0 -1 0.0
"H11" "hc" 0 -1 0.0
"C20" "c3" 0 -1 0.0
"H24" "hc" 0 -1 0.0
"H25" "hc" 0 -1 0.0
"H26" "hc" 0 -1 0.0
"C13" "c3" 0 -1 0.0
"H12" "hc" 0 -1 0.0
"H13" "hc" 0 -1 0.0

"C10" "c3" 0 -1 0.0
"H6" "hc" 0 -1 0.0
"H7" "hc" 0 -1 0.0
"C9" "c3" 0 -1 0.0
"H4" "hc" 0 -1 0.0
"H5" "hc" 0 -1 0.0
"C15" "c3" 0 -1 0.0
"H15" "hc" 0 -1 0.0
"H16" "hc" 0 -1 0.0
"C17" "c3" 0 -1 0.0
"H19" "hc" 0 -1 0.0
"H20" "hc" 0 -1 0.0
"H21" "hc" 0 -1 0.0
"C21" "c3" 0 -1 0.0
"H27" "hc" 0 -1 0.0
"H28" "hc" 0 -1 0.0
"H29" "hc" 0 -1 0.0
!entry.ODM.unit.boundbox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.ODM.unit.childsequence single int
2
!entry.ODM.unit.connect array int
0
0
!entry.ODM.unit.connectivity table int atom1x int atom2x int flags
1 2 1
1 3 1
3 4 1
3 5 1
3 66 1
5 6 1
5 7 1
5 73 1
7 8 1
7 9 1
7 60 1
9 10 1
9 63 1
9 69 1
10 11 2
10 50 1
11 12 1
11 57 1
12 13 1
12 19 2
13 14 1
13 47 1
13 53 1
14 15 1
14 16 1
14 21 1
16 17 1

16 18 1
16 19 1
19 20 1
21 22 1
21 23 1
21 43 1
23 24 1
23 25 1
23 26 1
26 27 1
26 28 1
26 29 1
29 30 1
29 40 2
30 31 1
30 32 1
30 36 1
32 33 1
32 34 1
32 35 1
36 37 1
36 38 1
36 39 1
40 41 1
40 42 1
43 44 1
43 45 1
43 46 1
47 48 1
47 49 1
47 50 1
50 51 1
50 52 1
53 54 1
53 55 1
53 56 1
57 58 1
57 59 1
57 60 1
60 61 1
60 62 1
63 64 1
63 65 1
63 66 1
66 67 1
66 68 1
69 70 1
69 71 1
69 72 1
73 74 1
73 75 1
73 76 1
!entry.ODM.unit.hierarchy table str abovetype int abovex str belowtype int belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2

"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
"R" 1 "A" 18
"R" 1 "A" 19
"R" 1 "A" 20
"R" 1 "A" 21
"R" 1 "A" 22
"R" 1 "A" 23
"R" 1 "A" 24
"R" 1 "A" 25
"R" 1 "A" 26
"R" 1 "A" 27
"R" 1 "A" 28
"R" 1 "A" 29
"R" 1 "A" 30
"R" 1 "A" 31
"R" 1 "A" 32
"R" 1 "A" 33
"R" 1 "A" 34
"R" 1 "A" 35
"R" 1 "A" 36
"R" 1 "A" 37
"R" 1 "A" 38
"R" 1 "A" 39
"R" 1 "A" 40
"R" 1 "A" 41
"R" 1 "A" 42
"R" 1 "A" 43
"R" 1 "A" 44
"R" 1 "A" 45
"R" 1 "A" 46
"R" 1 "A" 47
"R" 1 "A" 48
"R" 1 "A" 49
"R" 1 "A" 50
"R" 1 "A" 51
"R" 1 "A" 52
"R" 1 "A" 53
"R" 1 "A" 54
"R" 1 "A" 55
"R" 1 "A" 56
"R" 1 "A" 57
"R" 1 "A" 58

"R" 1 "A" 59
"R" 1 "A" 60
"R" 1 "A" 61
"R" 1 "A" 62
"R" 1 "A" 63
"R" 1 "A" 64
"R" 1 "A" 65
"R" 1 "A" 66
"R" 1 "A" 67
"R" 1 "A" 68
"R" 1 "A" 69
"R" 1 "A" 70
"R" 1 "A" 71
"R" 1 "A" 72
"R" 1 "A" 73
"R" 1 "A" 74
"R" 1 "A" 75
"R" 1 "A" 76
!entry.ODM.unit.name single str
"ODM"
!entry.ODM.unit.positions table dbl x dbl y dbl z
7.681000 -0.489000 -1.571000
8.068000 0.235000 -2.041000
6.290000 -0.302000 -1.453000
5.888000 0.038000 -2.409000
5.928000 0.743000 -0.386000
6.395000 0.427000 0.543000
4.391000 0.789000 -0.212000
3.981000 1.020000 -1.198000
3.773000 -0.584000 0.188000
2.238000 -0.456000 0.278000
1.637000 0.712000 0.525000
0.170000 0.791000 0.686000
-0.652000 -0.472000 0.901000
-2.089000 0.018000 0.520000
-2.146000 -0.057000 -0.563000
-2.051000 1.529000 0.849000
-2.394000 1.733000 1.864000
-2.688000 2.110000 0.189000
-0.594000 1.872000 0.690000
-0.250000 2.886000 0.597000
-3.307000 -0.724000 1.110000
-3.220000 -0.696000 2.194000
-4.643000 -0.016000 0.792000
-4.635000 0.986000 1.207000
-5.428000 -0.549000 1.323000
-5.026000 0.084000 -0.696000
-5.019000 -0.905000 -1.145000
-4.285000 0.670000 -1.227000
-6.393000 0.716000 -0.891000
-7.587000 -0.222000 -0.750000
-7.322000 -0.968000 -0.004000
-7.840000 -0.968000 -2.071000
-6.952000 -1.486000 -2.419000
-8.627000 -1.707000 -1.950000
-8.144000 -0.273000 -2.847000

-8.871000 0.459000 -0.269000
-9.258000 1.155000 -1.007000
-9.643000 -0.284000 -0.090000
-8.708000 1.005000 0.654000
-6.505000 2.002000 -1.176000
-7.454000 2.493000 -1.293000
-5.635000 2.623000 -1.304000
-3.370000 -2.201000 0.696000
-2.567000 -2.779000 1.136000
-3.305000 -2.323000 -0.381000
-4.304000 -2.650000 1.023000
-0.077000 -1.543000 -0.032000
-0.251000 -1.231000 -1.058000
-0.575000 -2.499000 0.094000
1.429000 -1.741000 0.167000
1.787000 -2.328000 -0.668000
1.610000 -2.353000 1.049000
-0.536000 -0.897000 2.378000
-1.033000 -1.842000 2.568000
0.503000 -1.007000 2.669000
-0.973000 -0.150000 3.032000
2.379000 2.025000 0.635000
2.127000 2.633000 -0.233000
2.008000 2.572000 1.497000
3.895000 1.888000 0.727000
4.338000 2.843000 0.477000
4.195000 1.671000 1.748000
4.190000 -1.609000 -0.893000
3.691000 -1.357000 -1.827000
3.857000 -2.604000 -0.623000
5.699000 -1.663000 -1.132000
6.215000 -2.060000 -0.264000
5.917000 -2.340000 -1.952000
4.265000 -1.082000 1.570000
3.859000 -2.066000 1.778000
5.343000 -1.165000 1.625000
3.942000 -0.426000 2.368000
6.522000 2.109000 -0.756000
7.584000 2.025000 -0.956000
6.042000 2.524000 -1.640000
6.420000 2.829000 0.046000
!entry.ODM.unit.residueconnect table int c1x int c2x int c3x int c4x int c5x int c6x
0 0 0 0 0 0
!entry.ODM.unit.residues table str name int seq int childseq int startatomx str restype int imagingx
"ODM" 1 77 1 "?" 0
!entry.ODM.unit.residuesPdbSequenceNumber array int
0
!entry.ODM.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.ODM.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
0.0 0.0 0.0

```
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
```

0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

The frcmod file of parameters of ligand ODM is:

OBT product (OBT_DM) after demethylation by CYP51
MASS

BOND

ANGLE

DIHE

c2-ce-c3-hc	1	0.380	180.000	-3.000	same as hc-c3-c2-c2
c2-ce-c3-hc	1	1.150	0.000	1.000	same as hc-c3-c2-c2
c2-ce-c3-c3	1	0.000	0.000	2.000	same as X -c2-c3-X
ce-ce-c3-c3	1	0.000	0.000	2.000	same as c2-ce-c3-c3
ce-ce-c3-hc	1	0.380	180.000	-3.000	same as c2-ce-c3-hc
ce-ce-c3-hc	1	1.150	0.000	1.000	same as c2-ce-c3-hc

IMPROPER

c3-c3-c2-ce	1.1	180.0	2.0	Using default value
c2-c3-ce-ce	1.1	180.0	2.0	Using default value
c3-ce-c2-ha	1.1	180.0	2.0	Using default value
c2-c3-c2-c3	1.1	180.0	2.0	Using default value
c2-ha-c2-ha	1.1	180.0	2.0	Using default value

NONBON

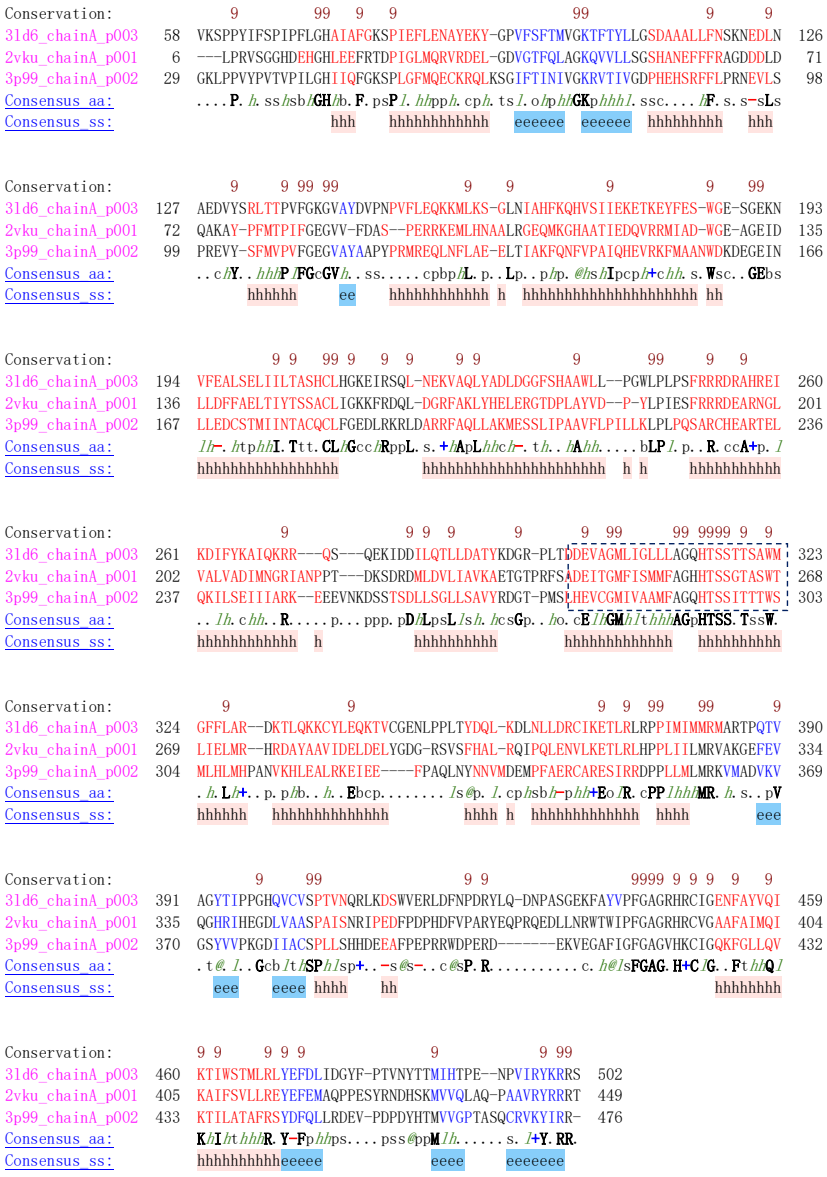

Figure 9.1: Sequence alignment of *M. tuberculosis*, *T. brucei* and human CYP51. The I helix region is marked by the box with dashed lines.

Figure 9.2: Sequence alignment of human, rat and *T. brucei* CPR. Interface residues that were identified by mutagenesis experiments as discussed in Chapter 5 are marked by the box with dashed lines.