

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 B. Pharmacy

Ghulam Mustafa

Born in Larkana, Sindh, Pakistan

Oral-examination:

Modeling and Simulation of Membrane Proteins to Understand their Structure, Dynamics and Function

Advisors:

Prof. Dr. Rebecca C. Wade

Prof. Dr. Christian Klein

ACKNOWLEDGMENTS

I want to express my deep gratitude to my supervisors Prof. Rebecca Wade of Center for Molecular Biology (ZMBH) and Prof. Christian Klein of Institute of Pharmacy and Molecular Biotechnology, Heidelberg University. I am indebted to Prof. Rebecca Wade for her unrelenting mentorship, guidance and genuine support at personal and professional level to accomplish my career goals.

I am thankful to DAAD (Deutscher Akademischer Austauschdienst) and Klaus Tschira Stiftung gGmbH for providing me financial support to complete my studies in Germany, and giving me opportunity to explore and extend my academic, social and cultural experience. And my previous supervisor Prof. Zaheer-ul-Haq Qasmi and Prof. Dr. M. Iqbal Choudhary for their help in achieving DAAD scholarship and all the respectable teachers who have laid the foundations for my scientific career.

I would like to pay my thanks to all current and past members of molecular and cellular modeling (MCM) group at Heidelberg Institute for Theoretical Studies (HITS gGmbH), with whom I had chance to develop social and scientific interactions, who welcomed me and extended their help when required. Especially, I want to express my sincere thanks to Stefan Richter, for his helping hand and excellent expertise in solving the otherwise unsolvable problems of IT and to Imme Roggenbach, for her wonderful German translation skills. I am also thankful to my colleagues Xiaofeng Yu and Prajwal Nandekar for sharing their knowledge and scientific interest in the common project that involved cytochrome P450 membrane systems. I would not like to forget Neil Bruce for his interest in long and extended discussions, critical views, suggestions and help in developing my understanding in the field of molecular modeling and simulations.

I am grateful to my friends Dr. Umair Amin and Dr. Sahib Khan Mirani for their motivation and support throughout my academic journey, for standing beside me and guiding me in the time of bewilderment.

Last but never the least, achieving my aims in life would not have been possible without

a support from my family, especially my parents and my elder brother, who has faced all hardships in life but remained a shield for his siblings. I want to pay my thanks to my children whom I have been missing during my stay abroad and whose love and trust in their father, have been a driving force for achieving my objectives.

TABLE OF CONTENTS

LIST OF FIGURES.....	1
LIST OF TABLES.....	4
LIST OF ABBREVIATIONS.....	6
LIST OF PUBLICATIONS.....	9
POSTER PRESENTATIONS	10
ORAL PRESENTATIONS.....	10
ABSTRACT	11
ZUSAMMENFASSUNG	12
CHAPTER 1: INTRODUCTION.....	14
1.1 BIOLOGICAL MEMBRANES	15
1.1.1 Membrane dynamics and fluidity	17
1.1.2 Annular and non-annular lipids.....	20
1.2 MEMBRANE PROTEINS	21
1.2.1 Transmembrane alpha helix and amino acid distribution.....	23
1.2.2 Hydrophobic mismatch.....	24
1.3 BITOPIC MEMBRANE PROTEIN: CYTOCHROME P450.....	25
1.3.1 Three-dimensional structures of human CYPs.....	26
1.3.2 Substrate recognition sites	28
1.3.3 Dynamics of CYPs.....	29
1.3.4 Molecular dynamics simulations of CYPs in lipid bilayers	31
1.3.5 Active site access and egress pathways.....	32
1.4 CONCLUSION.....	35
CHAPTER 2: COMPUTATIONAL MODELING AND SIMULATION METHODOLOGY	36
2.1 COMPUTATIONAL MICROSCOPE	36
2.2 COMPUTATIONAL MODELING OF MEMBRANE PROTEINS.....	37
2.2.1 Homology modeling.....	39
2.2.2 Force-fields	40
2.2.3 MARTINI coarse-grained force-field	42

2.2.4	MARTINI Coarse-grained mapping.....	43
2.2.5	Bonded interactions in MARTINI.....	44
2.2.6	Non-bonded interactions in MARTINI.....	44
2.2.7	MARTINI protein and Elastic Network models.....	45
2.2.8	MARTINI coarse-grained water models	46
2.3	MOLECULAR DYNAMICS SIMULATION.....	46
2.3.1	Integration of equation of motion	47
2.3.2	Treatment of non-bonded interactions	48
2.3.3	Periodic boundary conditions	49
2.3.4	Simulation conditions	49
2.3.5	Pressure control for membrane protein simulation	50
2.4	MULTISCALE SIMULATIONS OF MEMBRANE-PROTEIN SYSTEM.....	51
2.4.1	All-atomic molecular dynamics simulation	51
2.4.2	Coarse-grained simulation.....	52
2.4.3	From coarse-grained to all-atom (CG2AA).....	54
 CHAPTER 3: CYTOCHROME P450 MEMBRANE ASSOCIATION: AN OPTIMIZATION OF THE MARTINI COARSE-GRAINED PROTOCOL.....		55
3.1	INTRODUCTION	55
3.2	MATERIAL AND METHOD	58
3.2.1	All-atom models of CYP3A4	58
3.2.2	Preparation of coarse-grained (CG) systems	58
3.2.3	Coarse-grained simulations using different parameters.....	59
3.2.4	Back conversion from coarse-grained to all-atom model.....	61
3.2.5	All-atom molecular dynamics simulation.....	62
3.3	DATA ANALYSIS	63
3.3.1	Angle and distance calculations.....	63
3.3.2	Analysis of membrane properties.....	64
3.4	RESULTS AND DISCUSSION.....	65
3.4.1	Effect of different simulation parameters on performance.....	65
3.4.2	Effect of different simulation parameters on convergence	66

3.4.3	Effect of different simulation parameters on the bilayer properties.....	70
3.4.4	All-atom molecular dynamics simulation of CYP3A4	72
3.4.5	All-atom molecular dynamics simulation area per lipid.....	75
3.4.6	Comparison of results with experimental and simulation data	77
3.5	CONCLUSION	79

CHAPTER 4: UNDERSTANDING THE MEMBRANE PROTEIN INTERACTIONS OF TWO ISOFORMS OF THE CYP2C SUBFAMILY: CYP2C9 AND CYP2C1980

4.1	INTRODUCTION	80
4.2	MATERIAL AND METHOD	81
4.2.1	Preparation of full length CYP2C9 and CYP2C19.....	81
4.2.2	Preparation of additional models of CYP2C9:	82
4.2.3	Mutants of CYP2C9 and CYP2C19	83
4.2.4	Preparation of coarse-grained systems	83
4.2.5	Coarse-grained simulations with different parameters:	84
4.2.6	Convergence of coarse-grained simulation and back conversion.....	87
4.2.7	All-atom molecular dynamics simulations of CYP2C9 AND CYP2C19.....	88
4.3	COARSE-GRAINED SIMULATION RESULTS.....	90
4.3.1	Distinct orientations adopted by CYP2C9 and CYP2C19:	90
4.3.2	Full length vs globular domain simulation results.....	93
4.3.3	Orientations of CYP2C9 modeled structures:	95
4.3.4	Orientations of mutant models of CY2C9 and CYP2C19	98
4.4	ALL-ATOM SIMULATION RESULTS	101
4.4.1	CYP2C9 orientations in the membrane.....	101
4.4.2	CYP2C19 orientation in the membrane	106
4.4.3	Effect of sequence on the membrane-protein interactions:	111
4.4.4	Structural differences result in different membrane-protein interactions.....	116
4.4.5	Different all-atom force-fields and simulations parameters affect dynamics of proteins	119
4.4.6	Structure and dynamics of the membrane is affected by choice of force-field ...	126
4.4.7	Area per lipid	126

4.4.8	Electron density and membrane thickness.....	130
4.4.9	Order parameters.....	133
4.4.10	Comparison with experiments and previous simulations.....	134
4.5	CONCLUSION.....	138
CHAPTER 5: INFLUENCE OF THE TRANSMEMBRANE HELIX-ANCHOR ON CYTOCHROME P450 17A1 MEMBRANE INTERACTIONS.....		140
5.1	ABSTRACT.....	140
5.2	INTRODUCTION.....	141
5.3	MATERIAL AND METHODS.....	144
5.3.1	Preparation of all-atom models.....	144
5.3.2	Preparation of coarse-grained systems and simulations.....	144
5.3.3	Trajectory analysis and back-conversion to all-atom models.....	146
5.3.4	All-atom molecular dynamics simulation.....	147
5.4	RESULTS AND DISCUSSION.....	148
5.4.1	Transmembrane helix membrane interactions.....	148
5.4.2	Convergence of coarse-grained simulations.....	150
5.4.3	Difference in the orientation and position of wtTM vs mtTM helix:.....	157
5.4.4	All-atom MD simulation results.....	162
5.5	CONCLUSION.....	170
CHAPTER 6: CONCLUSION.....		172
REFERENCES.....		179
APPENDIX I.....		197

LIST OF FIGURES

1.1.	Different phospholipid molecules varying in the lipid headgroup	16
1.2.	Schematic representation of different phases of lipid bilayer..	18
1.3.	The structure of a di(C18:1) PC (DOPC) bilayer..	20
1.4.	Level of protein structure.	22
1.5.	IMPs topology and membrane orientation.	23
1.6.	Crystal structure of CYP2E1 (PDB 3T3Z) in two different orientations	27
1.7.	Alignment of the six-major drug metabolizing CYPs.....	31
1.8.	The important tunnels defined in Wade's nomenclature	34
2.1.	Application ranges for molecular modeling at different resolutions	38
2.2.	Coarse-grained representation of all amino acid residues.....	43
2.3.	Different pressure coupling schemes and their effect on a membrane bilayer	51
3.1.	Representative initial and final configurations of CYP3A4 in a POPC bilayer...63	
3.2.	Convergence of the position of CYP3A4 with respect to the bilayer	69
3.3.	Angles and distances calculations.....	73
3.4.	Comparison of representative B-factor values.....	74
3.5.	2D-RMSD in simulations of CYP3A4.....	75
3.6.	Area per lipid calculations.....	77
4.1.	CG simulations results comparison between CYP2C9 and CYP2C19.....	91
4.2.	Snapshots from CG simulations of globular domains of CYP2C9 and CYP2C19.	94
4.3.	Angles and distance graphs of CYP2C9 and CYP2C19.....	95
4.4.	Angles (°) and distance (Å) values vs time (μs) for CG simulations of CYP2C9 M1 (left) and M3 (right)	97
4.5.	Angles (°) and distance (Å) values vs time (μs) for CG simulations of mtCYP2C9 (left) and mtCYP2C19 (right).....	100

4.6.	Last frame from CYP2C9:SIM1 all-atom simulation. Orientation of CYP in the membrane representation.	104
4.7.	Last frame from CYP2C9:SIM5 all-atom simulation.....	105
4.8.	Last frame from CYP2C9:SIM7 all-atom simulation.....	106
4.9.	Last frame from CYP2C19:SIM1 all-atom simulation.	109
4.10.	Last frame from CYP2C19:SIM4 all-atom simulation.	110
4.11.	Sequence alignment between CYP2C9 and 2C19.....	112
4.12.	Sequence alignment between human CYP2C family (2C9,2C19,2C18,2C8) and rabbit CYP2C5.....	113
4.13.	Residues in contact with lipid headgroup and tail region	115
4.14.	The difference in the CYP2C9 and CYP2C19 residues at the interface.....	118
4.15.	Root mean square deviation (RMSD) plots of CYP2C9 and CYP2C19.....	121
4.16.	Comparison of the average B-factor values of CYP2C9 and 2C19 with X-ray structure.....	122
4.17.	X-ray crystal structure B-factor values for the two CYPs.....	123
4.18.	Secondary structure analysis of CYP2C9.Comparison between two force-fields, GAFF with ff99SB vs LIPID14 with ff12SB	124
4.19.	Secondary structure analysis of 2C19.Comparison between two force-fields, GAFF lipid and ff99SB vs LIPID14 and ff12SB.....	125
4.20.	The area per lipid calculations. CYP2C9 simulations with LIPID14 compared with GAFF lipid.	129
4.21.	Electron density profile comparison between LIPID14 and GAFF lipid force-field calculated for total system	131
4.22.	Electron density profile Comparison between LIPID14 and GAFF lipid force-field	132
4.23.	Deuterium order parameters compared between LIPID14 and GAFF lipid. ...	134
4.24.	Orientations of CYP in the membrane suggested by AFM study	135
5.1.	Multiple sequence alignment for first 1-30 residues of wtCYP17, mtCYP17, rat, mouse, pig and bovine sequences.....	143

5.2.	Full length modeled structure of CYP17 superimposed on the x-ray crystal structure (PDB 3RUK chain C)	149
5.3.	Evaluation of protein position with respect to the bilayer during CG simulations	151
5.4.	Snapshots at 4 different time steps showing different orientations of the globular domain	156
5.5.	Position of TM-helices in the full-length mtCYP17 simulation (S1:3) compared with wtTM helix self-assembly simulations (S6: 1-5)	157
5.6.	Starting configuration used for AAMD simulations of wtCYP17 vs mtCYP17 superimposed on the x-ray crystal structure.....	159
5.7.	The TM helix orientation and position predicted by PPM server.....	160
5.8.	The different linker positions compared between mtCYP17 CG simulations S1-S3 (A-C) and wtCYP17 simulation S4.....	161
5.9.	Initial representative frame after back-conversion of mtCYP17 superimposed on X-ray crystal structure (PDB ID 3RUK).....	162
5.10.	Evaluation of the position of the protein with respect to the phospholipid bilayer during AAMD.....	164
5.11.	Last frame from mtCYP17: SIM3 and wtCYP17: SIM6	167
5.12.	Comparison between the initial and last frames and between mtCYP17 and wtCYP17	168
5.13:	RMSD and the average B-factor values for wtCYP17 and mtCYP17 simulations	169

LIST OF TABLES

2.1.	Different coarse-grained models used to study biomolecular systems.....	53
3.1.	CG Simulations performed for five CYPs.....	61
3.2.	The performance of CG simulations with two water models.....	68
3.3.	The means and standard deviations for angles and distances defining the position of the CYP3A4 globular domain with respect to the bilayer	70
3.4.	The area per lipid (APL) computed for CG simulations of the pure POPC and CYP3A4-bilayer systems.....	71
3.5.	APL computed and shown for first and last frames from AAMD trajectory analysis.....	76
3.6.	Comparison of the heme-tilt angle from experiments and simulation results for CYP3A4	78
4.1.	Coarse-grained simulation protocols used for CYP2C9 and CYP2C19.....	85
4.2.	AAMD Simulations performed for CYP2C9 and CYP2C19. Different starting structures, parameters and force-fields used.	89
4.3.	CG simulation results for CYP2C9 and CYP2C19.....	92
4.4.	CG simulation results of CYP2C9 models to show dependence of final orientation in the membrane on the initial conformation of protein.....	96
4.5.	Angles and distances values obtained from CYP2C9 M3 CG simulations.	98
4.6.	The angles and distances values of mtCYP2C9 and mtCYP2C19 calculated for the 5 orientations.....	99
4.7.	Mean and standard deviations of angles and distances calculated for all-atom simulations of CYP2C9	103
4.8.	Mean and standard deviations for angles and distances for all atom simulations of CYP2C19.....	108
4.9.	Area per lipid and box dimensions for the initial and final frames obtained from MD trajectories of CYP2C9 under different conditions.	128
4.10.	Heme-tilt angle comparison with previously available data.	137

5.1.	CG simulations of mutant and wildtype CYP17 models and CYP17 globular domain only.	146
5.2.	AAMD Simulations performed for CYP17.	148
5.3.	Mean and standard deviations for angles and distances characterizing positioning of the protein in the phospholipid bilayer calculated for wtCYP17 vs mtCYP17 and globular domain of CYP17 CG simulations.	155
5.4.	Mean and standard deviations calculated for angles and distances formed by mtCYP17 (SIM1-3) and wtCYP17 (SIM4-6) in AAMD simulations.....	165
A.1.	A list of in silico studies on human drug metabolizing CYPs is presented	197

LIST OF ABBREVIATIONS

AA	All-Atom
AAMD	All-Atom Molecular Dynamics
AFM	Atomic Force Microscopy
APL	Area per lipid
BD	Brownian dynamics
CG	Coarse-grained
CGMD	Coarse-Grained MD
CoM	Center of Mass
CYP	Cytochrome P450
DPD	Dissipative particle dynamics
DPPC	Dipalmitoylphosphatidylcholine
DSSP	Define Secondary Structure of Proteins
EN	Elastic network
FF	Force-Field
GAFF	Generalized Amber Force-Field
IMP	Integral Membrane Protein
LJ	Lennard-Jones
MC	Monte-Carlo
MD	Molecular dynamics
MEMSAT	Membrane protein Structure and Topology
MM	Molecular Mechanics
MP	Membrane protein
NMA	Normal mode analysis
NMR	Nuclear Magnetic Resonance
NPT	Constant Number Pressure Temperature
NPW	Non-polarizable water
NVE	Constant Number Volume Energy
NVT	Constant Number Volume Temperature
OPM	Orientation of Protein in Membranes
PBC	Periodic boundary condition
PC	Phosphatidylcholine
PDB	Protein Data Bank
PE	Phosphatidylethanolamine
PI	Phosphatidylinositol

PME	Particle mesh Ewald
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
PPM	Position of Protein in Membranes
PR	Parrinello-Rahman
PS	Phosphatidylserine
PW	Polarizable water
QM	Quantum Mechanics
RAMD	Random acceleration molecular dynamics
RF	Reaction field
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SMD	Steered molecular dynamics
SRS	Substrate recognition sites
SVM- MEMSAT	Support Vector Machines-MEMSAT
SWAXS	Small-and-Wide-Angle X-ray Scattering
TM	Transmembrane
VdW	Van der Waals
VTMC	Voronoi Tessellation Monti-Carlo

Amino Acids

3 Letters	1 Letter	
Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartate
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamate
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine

LIST OF PUBLICATIONS

1. Understanding the membrane protein interactions of two isoforms of CYP2C Subfamily: CYP2C9 vs CYP2C19 (To be submitted)
2. Influence of the transmembrane helix-anchor on Cytochrome 17a1 membrane interactions and enzyme catalysis (To be submitted)
3. X. Yu, P. Nandekar, **G. Mustafa**, V. Cojocar, G. I. Lepesheva, and R. C. Wade, "Ligand tunnels in *T. brucei* and human CYP51: Insights for parasite-specific drug design," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1860, no. 1, pp. 67–78, Jan **2016**.
4. **G. Mustafa**, P. P. Nandekar, X. Yu, and R. C. Wade, "On the application of the MARTINI coarse-grained model to immersion of a protein in a phospholipid bilayer," *J. Chem. Phys.*, vol. 143, p. 243139, Dec. **2015**.
5. X. Yu, V. Cojocar, **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**.
6. **G. Mustafa**, X. Yu, and R. C. Wade, "Structure and Dynamics of Human Drug-Metabolizing Cytochrome P450 Enzymes," in *Drug Metabolism Prediction*, J. Kirchmair, Ed. Weinheim: Wiley-VCH, pp. 75–102. Aug **2014**

Other publications:

7. K. M. Khan, M. A. Mesaik, O. M. Abdalla, F. Rahim, S. Soomro, S. A. Halim, **G. Mustafa**, N. Ambreen, A. S. Khalid, M. Taha, S. Perveen, M. T. Alam, A. Hameed, Z. Ul-Haq, H. Ullah, Z. U. Rehman, R. A. Siddiqui, and W. Voelter, "The immunomodulation potential of the synthetic derivatives of benzothiazoles: Implications in immune system disorders through in vitro and in silico studies," *Bioorg. Chem.*, vol. 64, pp. 21–28, Feb **2016**
8. R. Rafiullah, M. Aslamkhan, N. Paramasivam, C. Thiel, **G. Mustafa**, S. Wiemann, M. Schlesner, R. C. Wade, G. A. Rappold, and S. Berkel, "Homozygous missense mutation in the LMAN2L gene segregates with intellectual disability in a large consanguineous Pakistani family," *J. Med. Genet.*, vol. 53, pp. 138–144, Feb **2016**
9. Rafiullah Rafiullah, Rafiullah Rafiullah, Alyssa Long, Anna Ivanova, Hazrat Ali, Simone Berkel, **Ghulam Mustafa**, Nagarajan Paramasivam, Matthias Schlesner, Stefan Wiemann, Rebecca Wade, Eugen Boltshauser, Martin Blum, Richard Kahn, Tamara Caspary, and Gudrun Rappold, "A novel homozygous ARL13B variant in patients with Joubert syndrome impairs its guanine nucleotide-exchange factor activity", *European Journal of Human Genetics*, (accepted) **2017**

Poster Presentations

1. 20th International Conference on Cytochrome P450, Düsseldorf, Germany, Aug 27-31, 2017
(Poster)
2. 2015 Martini Coarse Graining Workshop in University of Groningen, Groningen, Netherlands, August 23-28, 2015 **(Poster)**
3. Free Energy Workshop (FEW2015) in University of Münster, Germany, March 09-11, 2015 **(Poster)**
4. EMBO Practical Course on Biomolecular Simulation in Paris, France. July 20–27, 2014 **(Poster)**
5. 2nd DKFZ-ZMBH Alliance Retreat at Kloster Schöntal, July 16-18, 2014 **(Poster)**
6. 20th International Symposium on Microsomes and Drug Oxidations in Stuttgart, May 18 - 22, 2014,
(Poster)
7. CCP5 Summer School on molecular dynamics simulations for PhD students in Manchester University, UK. July 22-30, 2013, **(Best poster prize)**

Oral Presentations

1. Molecular Docking: “A tool for Structure Based Drug Design “
Methods in Pharmacy & Molecular Biotechnology on 06.07.2016, Institute of Pharmacy and Molecular Biotechnology, Heidelberg University.
2. "Prediction of 3D structure of protein by homology modelling"
Methods in Pharmacy & Molecular Biotechnology on 18.11.2015, Institute of Pharmacy and Molecular Biotechnology, Heidelberg University.
3. “Understanding cytochrome P450-membrane interactions using multiscale simulations”
HITS-S³: Scientific Seminar Series, a presentation during PhD Scholarship, 15-5-2017, Heidelberg Institute for Theoretical Studies (HITS gGmbH), Heidelberg University.

Abstract

Human cytochrome P450 (CYP) enzymes play an important role in the metabolism of drugs, steroids, fatty acids and xenobiotics. CYPs also catalyze the conversion of some pro-drugs into active drugs. Only about a dozen human CYPs metabolize 70-80% of all drugs. A subset of CYPs is responsible for steroidogenesis, of these CYP17 is a major drug target for prostate cancer therapy. Human CYPs are anchored to the endoplasmic reticulum membrane by their N-terminal transmembrane (TM) helix. However, most crystal structures of CYPs have been resolved after truncating the TM-helix or mutating residues that form contacts with the membrane. Therefore, the structural basis for CYP-membrane interactions and orientation, and the mechanism of substrate entrance into the buried binding pocket and product release is not clearly understood.

In order to understand the interactions and orientations of CYPs and their degree of penetration into the membrane, I have optimized a multiscale modeling protocol that involves coarse-grained and all-atom molecular dynamics simulations. The protocol was validated by applying it to several drug-metabolizing CYPs (CYP1A1, 1A2, 2C9, 2C19, 3A4) and CYPs involved in steroidogenesis (CYP17, CYP19) in a lipid bilayer. The simulations revealed that the sequence and structural differences in the protein-membrane interface alter the interactions and orientations of CYPs in the membrane. Furthermore, mutations in the TM-helix of CYP17, particularly W2A and E3L, were seen to disrupt the CYP-membrane interactions and in some cases, obstruct the ligand tunnels between the active site and the membrane, which could lower enzyme turnover.

In conclusion, the optimized multiscale simulation protocol has been used to identify different interactions and orientations adopted by the globular domains of CYPs with the membrane that have implications for CYP function. This protocol is also suitable for studying protein-protein-membrane complexes and proteins in membranes with different lipid compositions.

Zusammenfassung

Humane Cytochrom P450 (CYP) Enzyme spielen eine wichtige Rolle bei der Verstoffwechslung von Steroiden, Fettsäuren, Arzneistoffen und anderen Xenobiotika. Ca. 70-80% aller Arzneistoffe werden durch die CYP450 Familie umgesetzt inkl. der Umwandlung von Prodrugs in aktive Wirkstoffe. Die Enzymfamilie dient zudem als Zielstruktur einiger Therapieansätze, z. B. das für die Steroidogenese wichtige CYP17 bei der Bekämpfung von Prostatakrebs. Humane CYPs sind durch eine N-terminale Transmembran (TM)-Helix mit der Membran des Endoplasmatischen Retikulums assoziiert. Kristallstrukturen der CYP Enzyme wurden meist jedoch nur für CYPs ohne TM-Helix oder mit eingeführten Punktmutationen an den Protein-Membran Kontaktstellen gelöst. Aus diesem Grund ist die strukturelle Basis der CYP-Membran-Interaktionen und der Mechanismus des Substratein- und austritts in die versteckte Bindetasche nur unzureichend beschrieben.

Um die Interaktionen und Orientierungen von CYPs in der Membran zu verstehen, beschreibe ich hier die Optimierung eines multiskalaren Modellierungsprotokolls, das „course-grained“ und „all-atom“ Molekulardynamik Simulationen nutzt. Das Protokoll wurde anhand von mehreren, funktionell variierenden CYPs validiert, inkl. CYP1A1, 1A2, 2C9, 2C19 und 3A4 für die Verstoffwechslung von Arzneistoffen und die an der Steroidogenese beteiligten Enzyme CYP17 und CYP19. Die multiskalaren Simulationen zeigten, dass strukturelle und Sequenzunterschiede an der Protein-Membran Schnittstelle die Interaktionen und Orientierungen der globulären CYP-Domänen in der Membran verändern. Zudem behinderten Mutationen in der TM-Helix von CYP17 (insbesondere W2A und E3L) die Ausbildung der Interaktionen und führten in einigen Fällen zur partiellen Blockade des Liganden-Tunnels zwischen der Membran und dem aktiven Zentrum. Aus diesem Grund könnten sich diese Punktmutationen negativ auf den enzymatischen Substratumsatz auswirken.

Zusammenfassend wurden mit der Anwendung des optimierten multiskalaren Modellierungsprotokolls Interaktionen und Orientierungen der globulären CYP-Domänen in der Membran identifiziert, die für die Funktionsweise der Enzyme von Bedeutung sein könnten. Verallgemeinert bietet das Protokoll zudem Möglichkeiten der Untersuchung von Protein-Protein-Membran Komplexen und von Proteinen in Membranen mit variierenden Lipidzusammensetzungen.

CHAPTER 1: INTRODUCTION

Membrane proteins (MPs) play important roles in cells as receptors, channels, transporters and enzymes. Around 26% of the human coding genome represents MPs (~5539), and the largest fraction of MPs are single transmembrane helix proteins(1, 2). Membrane proteins are an important target for pharmaceutical companies, accounting for 50% of total drugs on the market(3). Membrane proteins are structurally divided into two main classes, integral membrane proteins (IMPs) and peripheral membrane proteins (PMPs), depending on their localization in the lipid bilayer. IMPs have some part inserted into the membrane forming a transmembrane domain (TMD) which has either an α -helical or a β -sheet topology. The TM regions of IMPs contain hydrophobic amino acid residues, due to which they are embedded into the central 30Å thick regions region of the lipid membrane. An α -helical IMP can interact with membranes by a single α -helical transmembrane anchor which spans the membrane bilayer and is therefore called a bitopic IMP (like Cytochrome P450 or CYPs). Alternatively, α -helical IMPs form a more complex helical bundle spanning the membrane, and are known as polytopic IMPs (examples are: GPCR, ion channels and transporters). There have been tremendous advances in the structure determination of IMPs, at the time of writing this the number of TM proteins has reached to 3099 out of which 2723 are α -helical and 364 are with β -sheet topology (<http://pdbtm.enzim.hu/>)(4). These structures do not contain the lipid bilayer atoms and therefore lack the crucial information of how lipids modulate the function of IMPs. For example, the crystal structures of CYPs are resolved by truncating the N-terminal α -helical transmembrane anchor and mutating residues in the globular domain which develop interactions with the membrane, thereby losing significant information on protein-membrane interactions and function. However, there is only one full length structure of *Saccharomyces cerevisiae* lanosterol 14-alpha demethylase (CYP51) reported to date which contains single TM helix domain(5). Therefore, there is increased need for computational tools, such as molecular modeling,

docking and simulation to provide insights on structure function relationships, and the dynamic interplay between protein, ligand and membrane systems.

I will discuss briefly biochemical/physiological aspect of biological membranes, membrane proteins and the interactions between membrane proteins. In the last part of the chapter, I will give overview of what has been learnt about the structure and dynamics of human cytochrome P450s (CYPs) which are proteins with single α -helical membrane anchor. I will discuss what experimental and computational approaches have been employed to understand structure-function relationship of human CYPs.

1.1 Biological membranes

Biological membranes are selectively permeable to various molecules and play important roles in maintaining the difference between the internal (cytosolic) and external environment of the cell. Inside the cells of eukaryotes, various internal organelles are surrounded by membranes that separate the internal content of these organelles from the cytosol. These membranes are formed by lipid molecules which constitute 50% of the membrane. Lipids are amphipathic in nature i.e, they have both hydrophobic and hydrophilic characteristics, causing them to self-assemble in aqueous environment, forming lipid bilayer. Proteins are the second most abundant constituent of membranes. The protein-lipid interactions result in an array of specialized cellular functions, including cell-cell communications, receptor and signal transduction, the passage of polar substances into and out of the cell and its compartments, energy transduction and enzymatic activity.

The composition of lipid bilayers depends on cell type and cellular compartment. The most abundant lipids in the membrane are phospholipids, which are further classified based on the nature of the lipid head groups into phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), and phosphatidylinositol (PI) (Figure 1.1). The first two phospholipid head groups are zwitterions with net zero charge, whereas the later contain a negative charge on the head groups. Mammalian

bilayers also contain sphingolipid, glycolipids and cholesterol. The varying concentrations of cholesterol in membranes can influence the membrane fluidity, and can result in the formation of microdomains. The outer plasma membrane (PM) composition differs from that of internal organelles. For example, the endoplasmic reticulum (ER) contains more neutral PC (~40%) compared to the PM (~10-15%) while the PM contains more anionic lipid on the inner leaflet (PS and PI) which results in a negatively charged surface on the cytosolic side of the PM.

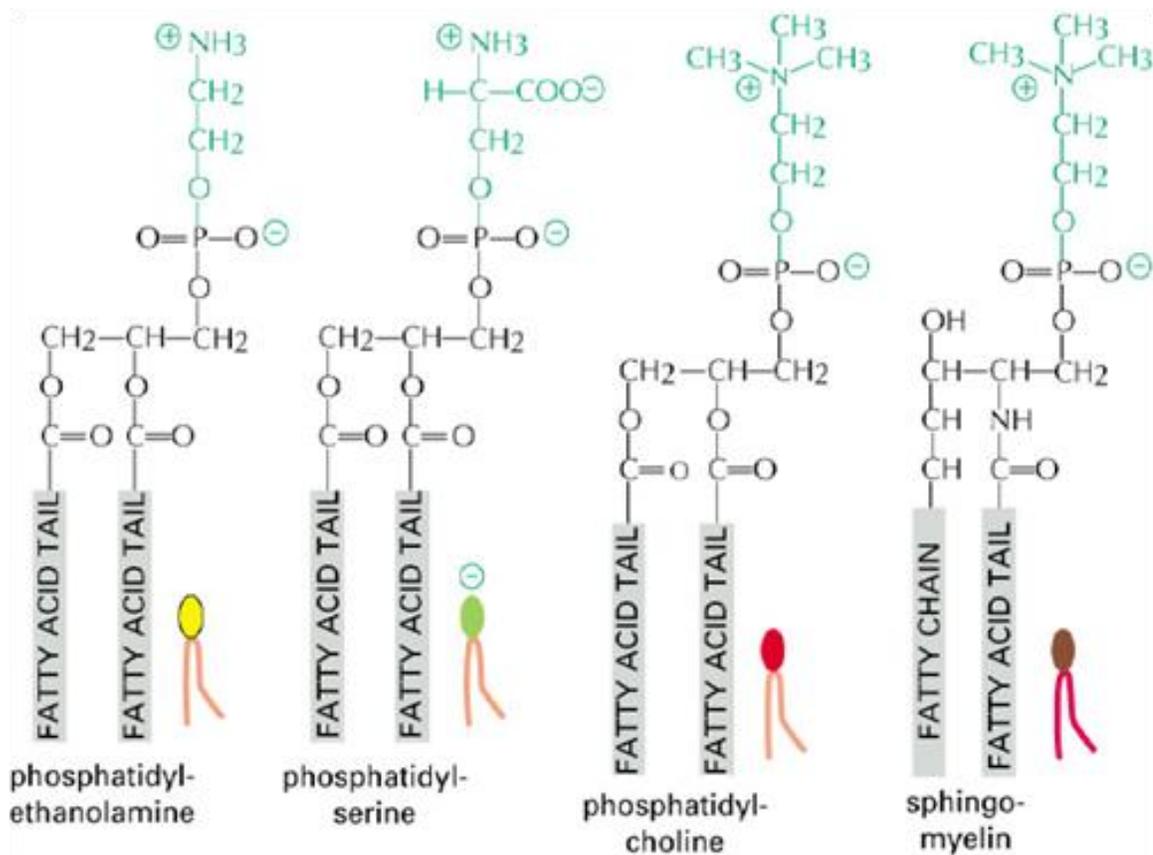


Figure 1.1: Different phospholipid molecules varying in the lipid headgroup (shown in green). Reprinted with permission from the book "Molecular Biology of the Cell", 4th Edition.

Membrane lipids have characteristic interfacial hydrophilic head group and a hydrophobic tail groups formed by two fatty acid chains that vary in length from 14 to 20 carbon atoms. In the case of unsaturated fatty acids, one of the tails contains a cis

double bond which results in a kink in that region. The chain length and saturation of the fatty acids influence the fluidity of lipid bilayer. A shorter chain length and a kink influences the tight packing allowing lipids to remain fluid at lower temperatures.

1.1.1 Membrane dynamics and fluidity

Biological membranes are dynamic in nature and the orientation and position of lipids in the membrane change by lateral diffusion in the membrane plane and transbilayer lipid diffusion (flip-flop movement or transverse diffusion). The transbilayer movement is slower compared to a lateral diffusion and is assisted by an enzyme known as flipase. Lipid molecules can change from trans to gauche conformations which can influence lipid packing and order (6). Lipid diffusion is influenced by the chemical composition of the lipid bilayer, presence of cholesterol, temperature, pH or ionic strength (Figure 1.2).

The gel to liquid phase transition is a unique property of lipid bilayers which affects membrane associated phenomena such as enzymatic activity, energy coupling reactions and the transport of solutes across the membrane. In the gel phase, a lipid molecule will likely have poor van der Waals contacts with the rough surface of an IMP, compared to in the liquid crystalline phase. This can influence the activity of the protein in the membrane.

Although there are no big molecular rearrangements during phase transitions, the molecular packing and fluidity of the lipid bilayer change. In the gel state, fatty acyl chains remain fully extended and are packed tightly in a hexagonal array which results in a small cross-sectional area per lipid, and maximum bilayer thickness and restricted molecular motion. In the liquid crystalline state, long range order, created by electrostatic interactions between the polar head groups and water, and hydrophobic interactions among acyl chains, are maintained. On the other hand, short range disorder is created due to flapping of acyl chains and formation of a hinge at the glycerol backbone.

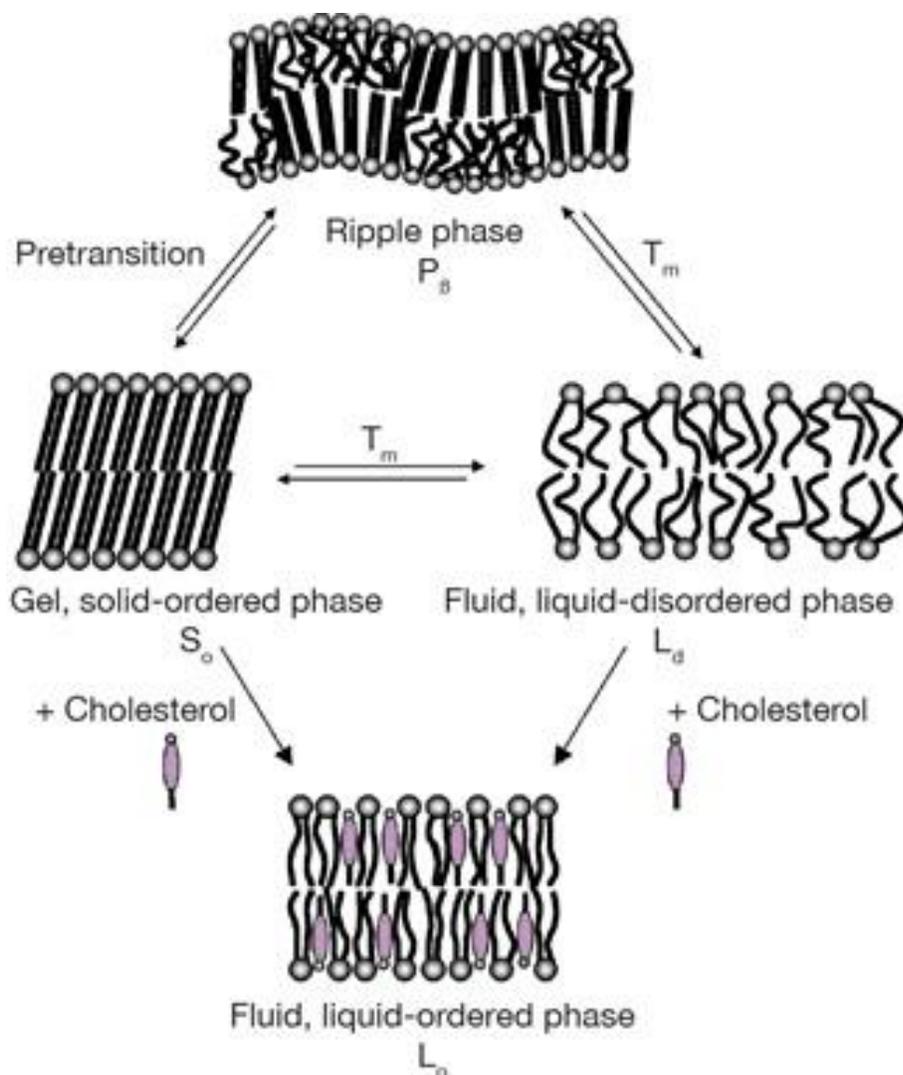


Figure 1.2: Schematic representation of different phases of lipid bilayer. An example of gel to fluid transition in the absence (to and right arrow) and in the presence of cholesterol (bottom arrows). Reprinted with permission from (6).

There is minimum fatty acyl chain motion at the carbonyl carbons of fatty acyl chains which increases progressively to the terminal methyl carbon (hydrocarbon core of bilayer) thus creating gradient fluidity, increased cross-sectional area per lipid and decreased bilayer thickness. For example, a dipalmitoyl phosphatidylcholine (DPPC) bilayer in the gel phase has a cross-sectional area per lipid of 40-45 Å² and bilayer thickness of about 50-55 Å, which changes in liquid crystalline state to cross-sectional area of 60-70 Å² and bilayer thickness to 40-45 Å (7).

The high-resolution X-ray and electron microscope (EM) structures of IMPs contain small numbers of lipid molecules which are highly ordered and do not represent the typical bulk lipid properties. The X-ray and neutron diffraction method is used to study the structure of the bilayer of dioleoylphosphatidylcholine [di(C18:1) PC] (DOPC) in the liquid crystalline phase at low hydration reveals a number of fragments formed by the lipid bilayer which are described by Gaussian distribution. The Gaussian distribution suggests the possible location of the fragments and range of the thermal motion depending on the width of the Gaussian, in the direction of the bilayer normal (8) (Figure 1.3). Accordingly, the Gaussian distribution is the narrowest and rigid region is glycerol backbone of the lipid molecule. While choline region of the headgroup and hydrocarbon tail towards the terminus methyl group shows high thermal motion and thus high flexibility. The hydrocarbon core thickness in the lipid bilayer is estimated at 32 Å. However, the thickness is measured in low hydration condition which can decrease slightly in fully hydrated conditions. The hydrocarbon core is the region which a transmembrane α -helix is required to span(8). According to the distance of 1.5 Å/residue in the perfect α -helix, it requires 20-21 residues for transmembrane α -helix to span the hydrocarbon core of 32 Å thickness in an un-tilted orientation. However, the number of residues can increase 10% if the helix form a tilt of 20°(9).

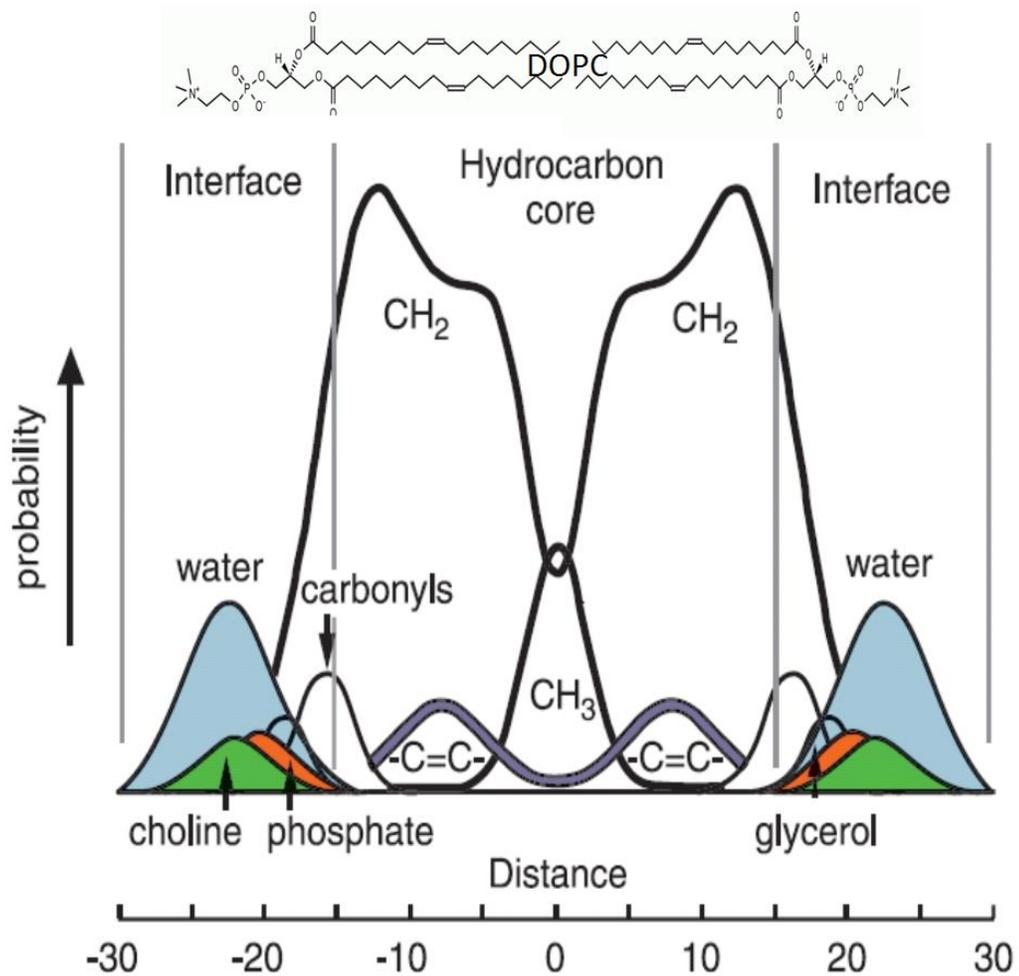


Figure 1.3: The structure of a di(C18:1) PC (DOPC) bilayer. The figure shows projections of the time-averaged transbilayer distributions of the principal structural groups along the bilayer normal. Reprinted with permission from A.G. Lee / *Biochimica et Biophysica Acta* 1612 (9).

1.1.2 Annular and non-annular lipids

The lipid molecules surrounding IMPs act as solvent for the proteins and show non-specific interaction with proteins. The solvent lipids surrounding these proteins are called annular lipids, or boundary lipids, and form annular shells around proteins. More specific interactions are observed between proteins and a small number of lipid molecules which can influence the activity of these proteins, such lipids act as cofactors. The cofactor lipids are bound within the grooves between α -helices, or between

protein-protein interfaces, and are known as non-annular lipids. Lipid molecules having no contact with proteins are referred to as bulk lipids, and retain the properties of a simple lipid bilayer in contrast to annular or cofactor lipids whose properties are influenced by the presence of membrane proteins (9). The bulk lipid molecules diffuse rapidly in the membrane plane, with diffusion constant of 10^{-7} - 10^{-10} cm²/s(6). The presence of proteins in the membrane restricts the diffusion of annular lipids. However, the rate of exchange between bulk and annular lipids is fast, indicating that the interactions between protein and annular lipids are non-specific and non-sticky.

1.2 Membrane proteins

Proteins participate in almost every process taking place in living organisms. They are the most abundant macromolecules and highly diverse in nature. Depending on their cellular environment, proteins are further classified into soluble and membrane proteins. Proteins are linear polymers formed by 20 different naturally occurring amino acids, which are joined together by peptide bonds. The amino acids share a common structure of a carboxyl group (COOH) and amino group (NH₃) connected to a central carbon atom (COOH-CRH-NH₂) but differ from each other in their side chain (R). The amino acids are broadly classified based on their side chains into hydrophobic (Ala (A), Val (V), Leu (L), Ile (I), Phe (F), Pro (P) and Met (M)), charged (Asp (D), Glu (E), Lys (K) and Arg (R)) and polar (Ser (S), Thr (T), Cys (C), Asn (N), Gln (Q), His (H), Tyr (Y) and Trp (W)) groups. The side chains of amino acids are responsible for the various conformations proteins can adopt. It is because of the different length and arrangement of these amino acids that such diversity in structure and functions of the proteins is observed.

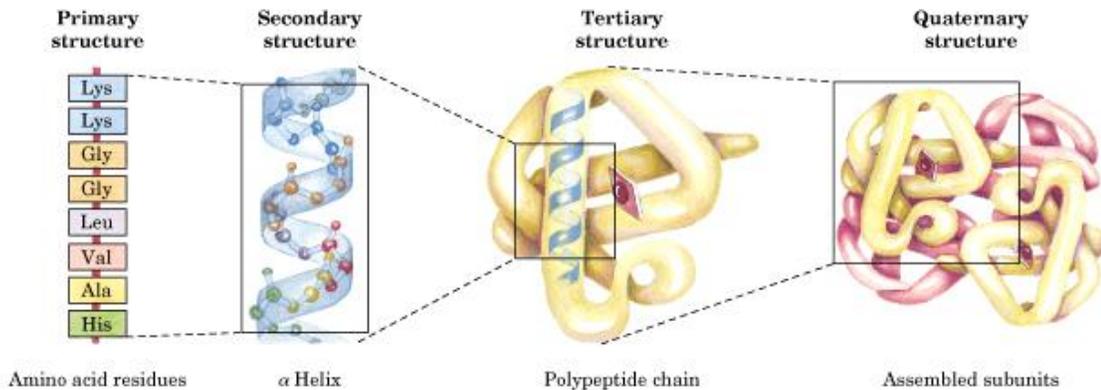


Figure 1.4: Level of protein structure. Reprinted with the permission from: “Lehninger principles of biochemistry”, 4th edition (2005).

Proteins can vary in length from 50-3000 amino acids. The folding of water soluble proteins is favored by the burial of hydrophobic residues inside the core of the protein, and polar residues exposed on the surface to the aqueous solvent. The protein backbone (-NH-CA-C=O) is highly polar with one hydrogen bond donor (NH) and one hydrogen bond acceptor (C=O) and hydrogen bonding within the backbone atoms neutralizes the protein backbone leading to the formation of secondary structure. The secondary structures such as α -helix or beta-strands are connected by loops and turns to make different motifs and domains in the tertiary structure (Figure 1.4).

On average, membrane proteins constitute 50 % of the mass of the membrane but this can vary from 25% to 75% depending on the cell type. Like lipid bilayer, membrane proteins are also amphipathic in nature i.e., they have polar surfaces which interact with water and lipid headgroup region, and a hydrophobic region which resides in the hydrophobic core of the lipid bilayer. Depending on the number of times a protein segment spans the membrane, IMPs are divided into single-pass (bitopic) or multipass (multitopic) integral membrane proteins (IMPs). The N-terminal side in most single-pass IMPs is found on the exoplasmic face (luminal side of endoplasmic reticulum) while the C-terminal end is on the cytoplasmic side, with a few exceptions where the reverse orientation is found (Figure 1.5).

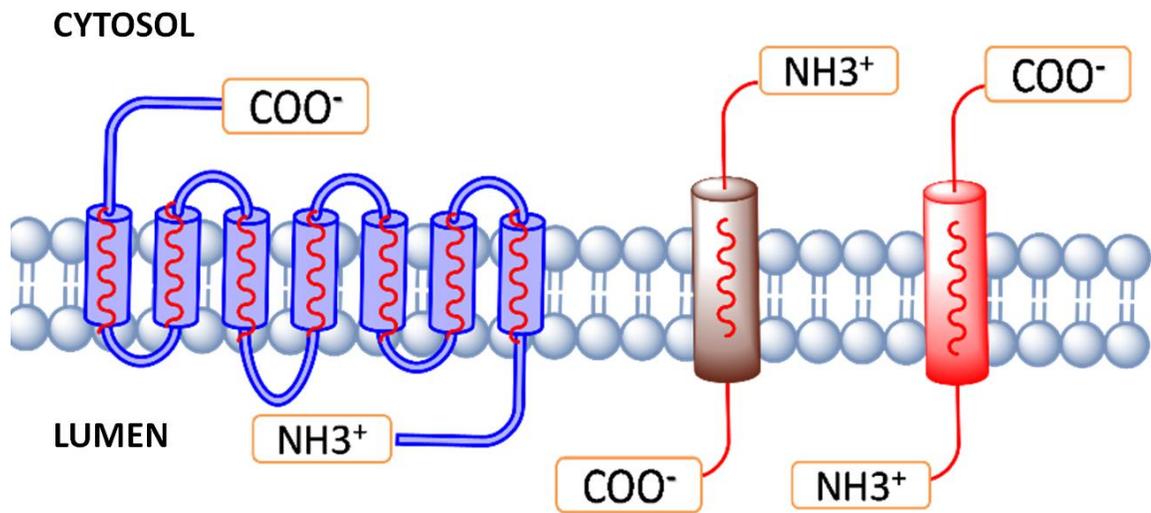


Figure 1.5: IMPs topology and membrane orientation. Based on the number of spanning regions, multi-pass and mono-pass transmembrane helices are shown with their N and C-termini.

1.2.1 Transmembrane alpha helix and amino acid distribution

Most IMPs traverse the membrane hydrophobic region as transmembrane (TM) α -helices and belong to the α -helical bundle class. The TM helix is inserted co-translationally and folded in eukaryotes in the endoplasmic reticulum (ER) membrane by the Sec61 translocon (10). In contrast to globular proteins, where α -helices contain hydrophobic residues facing interior of protein and hydrophilic residues pointing to the outer surface of the protein, TM α -helices are built from predominantly hydrophobic residues. A TM α -helix requires a minimum of 20 residues to span the hydrophobic core, although the full length of a TM helix can be longer to cross through the glycerol backbone and lipid headgroup region of the bilayer. The prediction of TM helix segments from sequence is possible by using a hydrophobicity index calculated for each residue position in the sequence. Based on hydrophobicity indices, a hydrophathy plot of possible transmembrane helices with high positive values can be drawn. The hydrophobicity scale has been determined experimentally by computing free energy of transfer of small model peptides from lipid to water (11). These results show that aromatic residues are favored at interface while charged residues and peptide bonds

are disfavored. In in vitro studies, the apparent free energy (ΔG_{app}) of amino acid residues in TM segment suggest that Ile, Leu, Phe and Val have $\Delta G_{app} < 0$ and promote membrane insertion, while Cys, Met and Ala have $\Delta G_{app} \approx 0$, and all charged residues have $\Delta G_{app} > 0$. It has been found that Leu and Phe behave similarly when placed in the middle of the TM helix, while Trp and Tyr behave differently and reduce membrane insertion when present in the center of the TM segment. However, when placed farther from the center they show less unfavorable behavior (10). The polar aromatic residues Trp and Tyr and charged residues Arg and Lys are located at the terminus of TM helices and are known as flanking residues (9). The energetic cost of burying charged residues in the hydrophobic core of lipid bilayer is so high (about 37 KJ/mol) that these residues are usually found at the end of TM-helices and in the lipid-water interface region (12). However, two residues with opposite charges can be found in the middle of hydrophobic core which will neutralize the charge effect. The favored location of Trp residues in TM helices is unclear, as it contains largest non-polar surface (two aromatic rings) and an NH group which is capable of hydrogen bonding. Similarly, Tyr contains an aromatic ring and an OH group, resulting in amphipathic properties. Therefore, Trp and Tyr have preference for being positioned closer to the lipid headgroup as well as in hydrophobic core. Yet in most cases, Trp and Tyr are found at the end of hydrophobic region of a TM α -helix, behaving as floating residues close to the lipid headgroups (9, 13). Arg and Lys residues contain long flexible hydrophobic side chains which can be located in the hydrophobic core while the charge terminus snorkels up to the membrane-solvent interface. Several computational studies have been performed for measuring partitioning free energy of amino acids between water and hydrophobic solvents and used as basis for parametrization of molecular dynamics (MD) force-fields (see Chapter 2).

1.2.2 Hydrophobic mismatch

The energetic cost of exposing hydrophobic amino acids or lipid fatty acyl chains to water is so high that it leads lipid molecules or proteins to adopt conformations such

that exposure to aqueous solvent is avoided. If the length of the hydrophobic region of a protein and the lipid hydrophobic region does not match, it is so-called a hydrophobic mismatch. The mismatch is described as positive when the protein hydrophobic segment exceeds that of the membrane in length and the mismatch is negative when TM hydrophobic segment is shorter than the lipid hydrophobic region. In case of a positive mismatch the protein may tilt, aggregate, or distort the bilayer leading to increased bilayer thickness. In negative mismatch, the shorter hydrophobic peptide or protein can aggregate or change conformation. If the TM hydrophobic region of a protein is too short or the cost of incorporating the protein in the bilayer is too high, it may be pulled out and remain at the surface.

1.3 Bitopic membrane protein: Cytochrome P450

Cytochrome P450 (CYP or P450) comprises a large superfamily of heme-thiolate monooxygenases found in all domains of life. Its name derives from its spectral absorbance which has a maximum at 450 nm when carbon-monoxide is bound to the ferrous heme iron. CYPs are highly versatile enzymes with diverse substrate specificities. They carry out stereo- and regio-selective oxidation of numerous physiologically and biotechnologically important molecules (14). Mammalian CYPs play an important role in the metabolism of both endogenous and xenobiotic compounds, including steroids, prostaglandins, fatty acids, environmental pollutants, agrochemicals, plant allelochemicals, and pharmaceuticals. CYPs also catalyze the conversion of some prodrugs into active drugs, and of procarcinogens and promutagens into highly toxic compounds (15).

There are around 57 genes encoding CYPs in humans. These genes have been classified into 18 families and 44 subfamilies by sequence analysis (16). Based on their cellular location, CYPs can be divided into three main categories: mitochondrial CYPs which are responsible for steroid metabolism, microsomal CYPs which are bound to the endoplasmic reticulum (ER), and soluble cytoplasmic CYPs which are mainly found in

prokaryotes. Only about a dozen human CYPs are responsible for the phase I metabolism of most foreign substances, including 70-80% of all drugs (16, 17). These CYPs belong to the CYP1, CYP2 and CYP3 families and show partially overlapping substrate specificity. Most drugs are metabolized by CYPs 3A4, 2D6, 2C9, 1A2 and 2C19 (18).

1.3.1 Three-dimensional structures of human CYPs

The first crystal structure of a human microsomal CYP, that of CYP2C9, was solved by Williams et al. in 2003 (19). Currently, there are 94 crystal structures of almost all the major human drug metabolizing CYPs reported in the Protein Data Bank (PDB). Most of these are in ligand-bound, closed conformations. Since human microsomal CYPs are membrane-bound proteins with a single N-terminal transmembrane helix anchored in the ER-membrane, crystallizing the native sequence is not trivial. To facilitate crystallization, the original sequence is modified by removing the residues of the N-terminal transmembrane domain. Additional substitution mutations are often introduced into the globular part of the protein to replace hydrophobic residues that bind to the membrane. For example, for the CYP2C9 structure (PDB 1OG5), Williams et al. introduced seven substitutions to increase the solubility of the protein (19).

Although there is a low level of sequence identity (<40%) between different families of CYPs, the overall secondary structure and folding pattern are conserved in both prokaryotes and eukaryotes. The crystal structures show that all CYPs contain 12 α -helices, designated by letters A-L, and 4 β -sheets (pairs of strands), labeled 1-4 as shown in Figure 1.6 (20, 21). In mammalian CYPs, several additional helices, and in some cases β -strands (e.g. $\beta 3'$ in CYP1A2), are present, which are indicated by prime or double prime symbols and follow the name of the neighboring helices (22–24). For example, many mammalian CYPs have F' and G' helices between the F and G helices. The F and G helices, the F-G loop and the B-C loop have been shown to play an important role in substrate access to the active site (14). The conserved I helix, the longest helix in CYPs,

is seen to run through the catalytic domain and bends or kinks next to the heme prosthetic group (25).

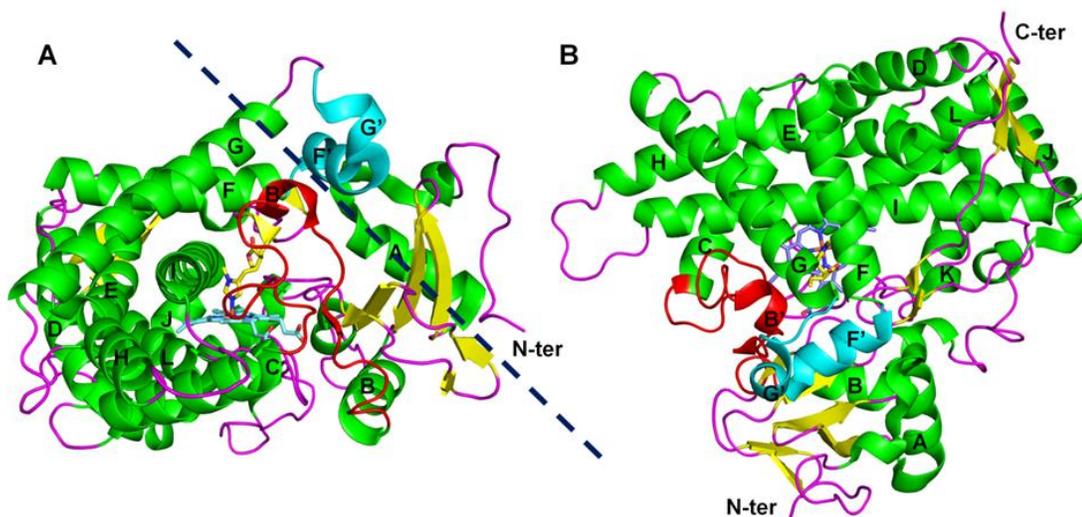


Figure 1.6: Crystal structure of CYP2E1 (PDB 3T3Z) in two different orientations. The main α -helices are shown in green, the B' helix in red, and the F' and G' helices in cyan. β -strands are shown in yellow and the loops in magenta. The heme is shown in light blue in the core of enzyme and the ligand pilocarpine in yellow above the heme co-factor. (A), looking along the I-helix. The membrane surface is indicated by a dashed line showing how the N-terminal region (truncated in the crystal structure), and part of catalytic domain embed in the membrane. (B), perpendicular view of CYP2E1, looking down on the heme from the distal side which shows the prism shape of the globular domain (26).

The heme is accommodated between the distal and proximal domains of the globular protein structure surrounded by the most conserved regions of the protein, including helices E, I, J, K and L and a portion of β -sheet 1 (21). The heme iron is axially coordinated by a cysteine residue on the proximal side. The meander on the proximal side contains positively charged residues and is important for binding the heme and the electron transfer protein which binds to the proximal face of CYPs. The substrate binding site is located on the distal side of the heme. Less conserved regions in CYPs are found at the outer surfaces of the substrate binding cavity and in regions such as the B-C and F-G helices and their intervening loops. The B-C loop adopts a range of conformations in

different CYPs and some structures contain a B' helix, e.g. CYP2A6 and 2C8, whereas in others it is less structured, e.g. CYP2C9 and 3A4 (21).

The N-terminal transmembrane helix in microsomal CYPs is connected to the catalytic domain by a linker region containing a proline-rich segment (21, 25). Microsomal CYPs have been seen to retain association with the membrane on removal of the N-terminal helix, indicating additional lipophilic interactions between the catalytic domain and the membrane. It was shown by epitope analysis that, along with the N-terminal domain, the helix A, β -strands and F-G loops are also buried inside the membrane and inaccessible to antibodies (27).

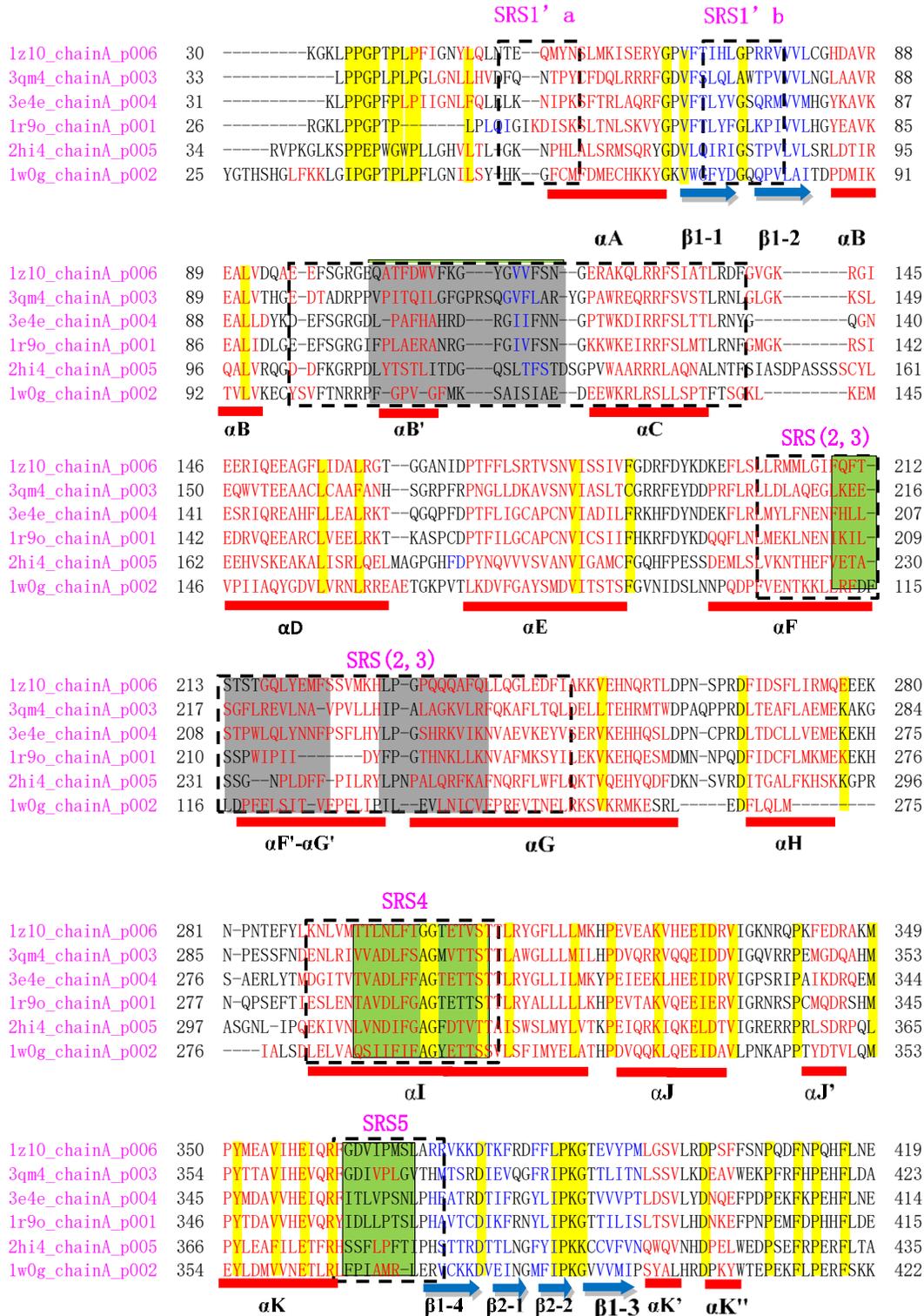
1.3.2 Substrate recognition sites

Gotoh categorized the most important residues for substrate binding by using a manual alignment of the mammalian CYP2 family with the structure of a bacterial CYP-substrate complex. He identified six regions that line the active site which he named Substrate Recognition Sites (SRS)(28). These regions accounted for 16% of the residues. SRS1 includes the B' helix and flanking sequences, SRS2 the C-terminal end of the F helix, SRS3 the N-terminal end of helix G, SRS4 the N-terminal half of helix I, SRS5 the loop between the K helix and strand 3 of β -sheet 1, and SRS6 the turn in β -sheet 4 (24, 28). However, as defined, these six SRS do not fully represent the active site residues of mammalian CYPs that have large active sites, like CYP3A4. Recently, a new SRS map based on mammalian crystal structures and docking of 868 known substrates of 10 mammalian CYP isoforms has been developed by Zawaira et al. (29). According to the new SRS map, two new regions, named SRS1'a and SRS1'b, were identified N-terminal to SRS1. Further, SRS1 in the new map is much longer (49 residues) than in the Gotoh map (28 residues). In the new map, SRS2 and SRS3 are merged into a single large site named SRS (2, 3), which covers 60 aligned sequence positions compared to 19 positions in the Gotoh map. SRS 4, 5 and 6 show extensions of a few residues compared to the Gotoh map. In the new SRS map, about 33% of the residues in a CYP are in the SRS. The

differences between the Gotoh and Zawaira SRS definitions are shown in the sequence alignment in Figure 1.7, where the dashed box represents the new SRS map and the Gotoh map is colored green.

1.3.3 Dynamics of CYPs

X-ray crystallography is one of the most valuable experimental techniques to obtain 3D structures of biomolecules, yet it does not provide direct information on the dynamics of membrane proteins or how these dynamics are affected by the presence of the lipid bilayer. Spectroscopic methods such nuclear magnetic resonance (NMR) spectroscopy can reveal the dynamics of CYPs. However, to obtain atomic-detail understanding of the dynamics of CYPs and how they influence function, a range of different computational approaches have been used. In this section, I will discuss computational studies to understand the dynamics of human microsomal CYPs, their interactions with phospholipid bilayer, their influence on protein flexibility and opening and closing of access and egress tunnels. A list of *in silico* studies on human drug metabolizing CYPs is presented in Appendix I which is taken from book chapter (26).



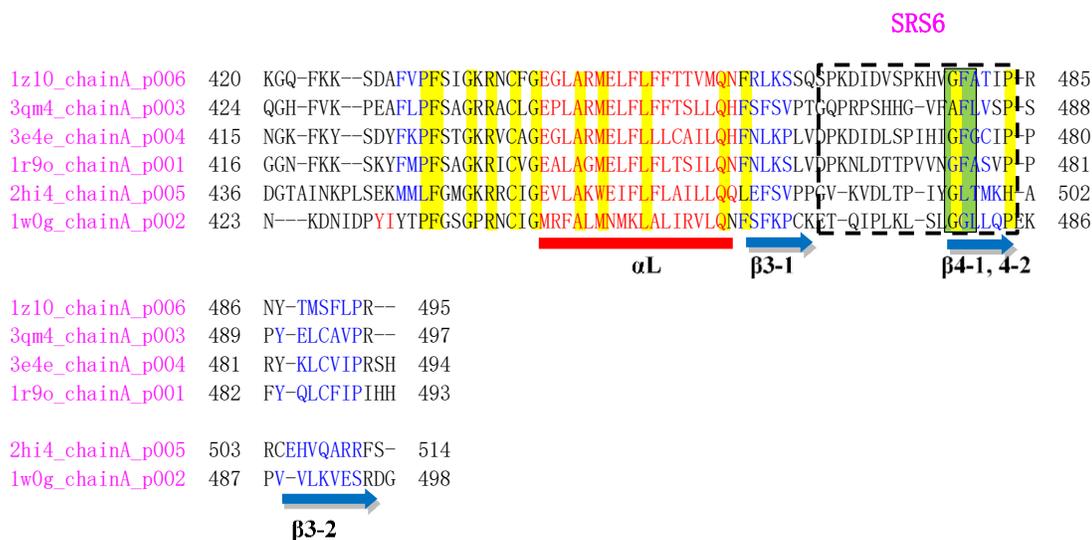


Figure 1.7: Alignment of the six-major drug metabolizing CYPs. The alignment was derived by structural alignment (chain A in the PDB files) using the PROMALS3D online server. The CYPs structures used are: 1A2 (PDB: 2HI4), 2A6 (PDB: 1Z10), 2C9 (PDB: 1R9O), 2D6 (PDB: 3QM4), 2E1 (PDB: 3E4E), 3A4 (PDB: 1WOG). The conserved secondary structures are represented by red and blue colors, with thick lines and arrows, for α -helices and β -strands, respectively. Boxes with dashed lines represent the new SRS map which differs in size and location from the SRS map first defined by Gotoh et al., 1992 which is shown in green colored blocks.

1.3.4 Molecular dynamics simulations of CYPs in lipid bilayers

Removal of the N-terminal helix for crystal structure determination limits our knowledge of the orientation and position of CYPs in the membrane. Experiments such as epitope analysis, mutagenesis, and tryptophan fluorescence scanning provide indirect information regarding the CYP-membrane topology, regions of the protein interacting with the membrane (27, 30). MD simulations of CYP2C9 in a phospholipid bilayer were performed by Cojocaru et al. (31) and Berka et al. (32) using the structures described in PDB 1R9O (in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)) and 1OG5 (in dioleoylphosphatidylcholine (DOPC)), respectively. Cojocaru et al. used an *ab initio* modeling and simulation procedure for predicting the missing F'-G' helix/loop conformations and further employed coarse grained and atomic-detail MD simulations to immerse the protein in the bilayer and then compared the result with data from

epitope analysis and tryptophan fluorescence scanning. On membrane binding, the CYPs were seen to exhibit decreased flexibility in the B-C loop, the F-G loop and the β 1 sheet whereas the G-H and H-I loops did not show any change in flexibility. The opening of ligand tunnels for lipid soluble drugs was also observed with membrane binding although the overall protein structure remained in a closed or nearly closed conformation (31). The membrane-bound model of CYP2C9 shows the tunnels 2a and 2f opening into the membrane whereas the 2b, 2ac, 2c, 2d and S tunnels lead to the aqueous solution (31, 33). The opening of the 2c and S tunnels showed correlation with the closure of the 2a and/or 2b tunnels. It was suggested that membrane-protein interactions favor the opening of the 2a tunnel for the access of hydrophobic substrates to the active site and the release of more polar products into the cytosol by either the 2c or the S tunnels (31, 32). MD simulations performed for CYP3A4 in a bilayer revealed different conformations of CYP3A4 than in aqueous solution. Whereas the 2b and 2e tunnels were open in the absence of the membrane, only the 2b tunnel was open in the simulations of the membrane-bound protein due to structural rearrangement of the F and G helices and the beta domain, which interact with the membrane, and the B-C loop (34).

1.3.5 Active site access and egress pathways

In an analysis of CYP crystal structures by Wade et al., the conformations were classified as closed, holey or wide open based on the solvent accessibility of the active site (35). Computational methods to analyze ligand access and egress pathways include classical MD simulation, Random Acceleration Molecular Dynamics (RAMD) simulation (36) and Steered Molecular Dynamics (SMD) simulations (37). as well as tools to compute tunnels such as: CAVER (<http://loschmidt.chemi.muni.cz/caver/>) and MOL (<http://mole.upol.cz/online/>). These methods have been applied to both microbial and human microsomal CYPs to understand the routes of entry and exit adopted by the ligands and the important residues lining these pathways. The naming scheme for access/egress/ channels first identified in microbial CYPs was further extended and

modified for human microsomal CYPs (35, 36, 38–41). Cojocaru et al. studied the access/egress tunnels from the active sites to the protein surface in all then (2007) available crystal structures of CYPs using CAVER (42). These access/egress tunnels were suggested to play an important role in substrate selectivity and specificity as some of the SRS are found along these tunnels. The ligand pathways have been named according to the secondary structure elements lining them and their spatial orientation with respect to the CYP 3D-fold (42). Here, I discuss briefly the most common pathways and their location as described by Cojocaru et al.,(42)

Pathway 2a (pw2a) lies between the F-G loop, the B' helix, the B-B' loop/B-C loop and the β 1 sheet, where the B-C loop corresponds to SRS1 (Gotoh, 1992), the C terminus of the F helix constitutes SRS2, and the N-terminus of the G helix forms SRS3. Pathway 2b lies between the B-B' loop and β 1 and β 3 (SRS5 is in the β 3 sheet). Pathway 2c lies between the G and I helices and the B' helix/B-C loop, with the I helix forming SRS4. Pathway 2ac lies between pw2c and pw2a and egresses between the tip of the B' helix/B-C loop and the G helix. Pathway 2e egresses through the B-C loop. Pathways 2f and 2d are closer to pw2a, pw2d leaving between the N-terminus and helices A' and A, whereas pw2f runs between the F' helix/F-G loop and the β 5 sheet. The solvent channel which was reported to be open in almost half of the CYPs studied by Cojocaru et al., lies between the F, E and I helices and the β 5 sheet. As discussed earlier, the solvent (S) channel was originally found open in ligand-free CYP2D6 (PDB 2F9Q) (43), whereas it remained closed when CYP2D6 was complexed with a ligand (PDB 3QM4) (44). The water (W) channel runs between the B-C loop and the C terminus of the B helix. From analysis of the mechanism of tunnel gating, it has been noted that the relative movement of two secondary structure elements, the B-C loop and the F-G loop, leads to open/closed conformations of the 2c, 2ac, 2a, 2b and 2f tunnels, whereas opening of tunnels 2d, 2e and 4 results from conformational changes in the loops through which they thread.

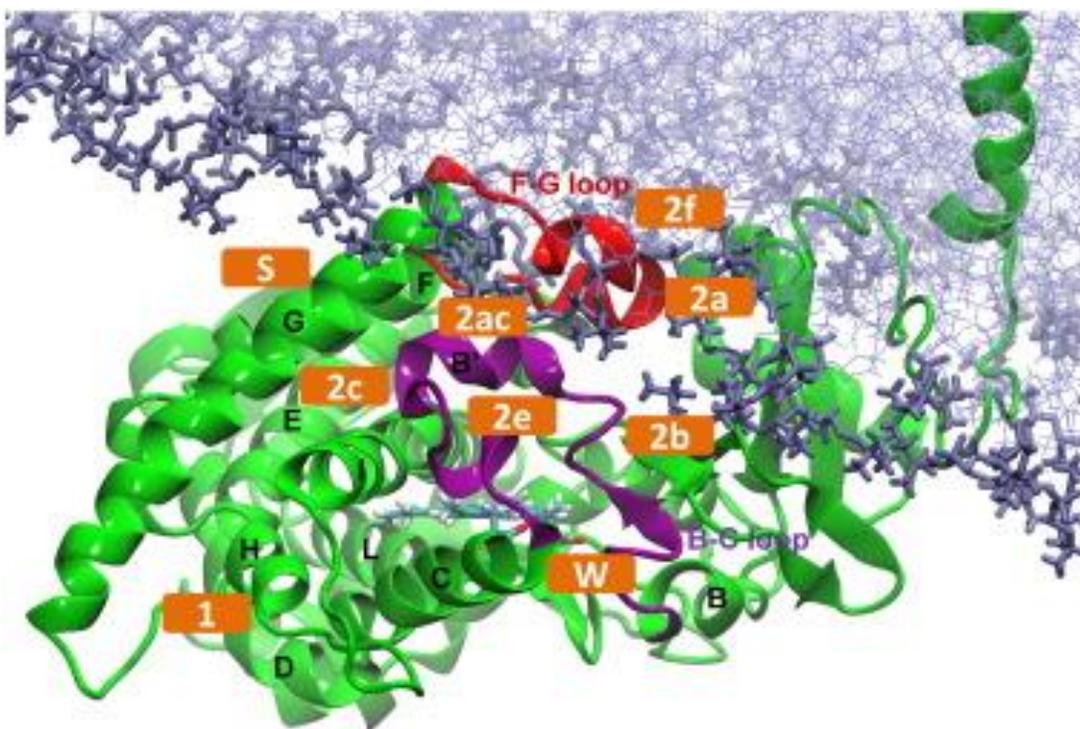


Figure 1.8: The important tunnels defined in Wade's nomenclature are labeled (orange). The protein is shown in cartoon representation with the B-C loop colored purple, the F-G loop red and the helices labelled. The heme (cyan) and the POPC bilayer (purple) are shown in stick representation(45).

Although classical MD simulation provides information on tunnel flexibility, the conformational space sampled in standard MD simulations is not generally sufficient to fully understand ligand access and egress from CYPs. To study phenomena taking place on longer time scales, enhanced sampling techniques such as RAMD(38, 46), in which an artificial force with random direction is applied to the center of mass of the ligand, and SMD (37, 39, 47, 48), in which a force is applied to the ligand in specified direction, are used. In RAMD simulations of mammalian CYP2C5 (41), pw2c was found to be the predominant product egress route, this was in contrast to the main egress route, pw2a, found in soluble bacterial CYPs (38). Since the 2a tunnel opens into the membrane in mammalian CYPs, a two route mechanism for lipophilic substrate entry (via pw2a) and hydrophilic product egress (via pw2c) was proposed by Schleinkofer et al. (41). Cojocar et al. performed RAMD simulations of CYP2C9 in the presence of the substrates, flurbiprofen and warfarin, and their products (49). Depending on the physicochemical

properties of the ligands, different tunnels were preferred. More non-polar ligands tended to exit through the 2a tunnel to the membrane whereas the hydroxylated product of flurbiprofen exited through the 2ac and 2e tunnels, leading to solvent. RAMD simulations of membrane bound *T.brucei* CYP51 by Yu et al found tunnel 2f as predominant ligand egress tunnel which leads to the membrane(45).

1.4 Conclusion

Given the important role of MPs in regulating cellular function, there is increasing focus on studying structure-function relationship of membrane proteins. For this purpose, several experimental techniques are used which are in the process of continuous development, aiming at a high-throughput pipeline for structure determination of MPs. Advances in experimental techniques like electron microscopy, NMR and X-ray crystallography have improved our understanding of the 3D-folds of MPs. However, there are practical difficulties with overexpression and solving high resolution crystal structures of α -helical MPs. Although, our knowledge in the field of structural biology, especially MPs, has increased over the past few years, the dynamic behavior of MPs in their physiological environment is difficult to understand through experimental techniques. Thus, computational tools such as modeling and simulations at various time and length scales have played a significant role in understanding the dynamics and mechanistic behavior of membrane protein interactions. Structural biology together with computational methods offer greater opportunity and better insights into MP which can be utilized to manipulate and target MP function and lead to novel drug discovery and design. A detail discussion about computational techniques employed to study membrane, protein and ligand interactions is given in Chapter 2.

CHAPTER 2: COMPUTATIONAL MODELING AND SIMULATION

METHODOLOGY

2.1 Computational microscope

The experimental difficulty in studying membrane proteins (MPs) arises because of the inherent property of MPs to fold and function in a complex heterogeneous and dynamic membrane environment. Although, improved experimental techniques have resulted in an increased number of atomic-resolution models of MPs, which have offered invaluable insight into secondary and tertiary folds of MPs, these models present static pictures of the conformational states biomolecules can adopt. MPs are however highly flexible and can adopt different conformational states which result in various cellular responses such as cell-signaling, ion-conductance, transportation across the cell-membrane and enzyme catalysis reactions. The conformational changes can occur due to ligand binding, protein-protein interaction, a change in potential difference or *pH* across the membrane. The presence of lipid molecules also modulates the structure and function of MPs (50). This effect is difficult to observe through experimental procedures due to spatial and temporal limitations (51). MPs, in turn, affect membrane properties such as lipid diffusion (52), membrane thickness (53), and membrane curvature or distortion (54). Such effects of MPs on shaping the membrane are exemplified by molecular dynamics (MD) simulations of the voltage-gated potassium channels (KvAP K⁺ channel) S4 transmembrane helix. The positively charged residues in the S4 (TM-helix) reduce the effective thickness of the bilayer core from 27 Å to 10 Å. (55).

The interplay between membrane and protein can be studied by employing various computational modeling approaches. Computational modeling has been recently termed “Computational Microscopy” by Schulten and co-workers (56) owing to its ability to study natural processes occurring at different spatial and temporal resolutions. For example, zooming in to highest resolution by quantum mechanics

simulations, which include electrons and nuclei, to study any chemical reaction i.e., breaking/formation of chemical bonds or charge transfer reactions. And gradually zooming out to study MPs interactions with membrane, water and ions at atomic resolution and on a time scale as small as femtoseconds, or further zooming out to supra-atomic (coarse-grained) level where groups of atoms are represented by beads and the timescale is increased by 2 to 3 orders of magnitude (57), to supra-molecular (supra-coarse grained) or mesoscopic level simulations (58). In a recent review by Ingolfsson et al (59), the power of computational microscopy to study the lateral organization of plasma membrane model is compared with traditional microscopy-based methods such as electron microscopy, atomic-force microscopy (AFM) as well as supra-resolution microscopy. Computational microscopy is useful to study dynamics, interactions and conformational changes similar to experimentally observed by fluorescence microscopy techniques such as fluorescence correlation spectroscopy (FCS), single-molecular fluorescence resonance energy transfer (smFRET), single particle tracking (SPT) and fluorescence recovery after photobleaching (FRAP) (59).

2.2 Computational modeling of membrane proteins

Biological processes occurring at different time and length scales and various computational modeling approaches are shown in Figure 2.1. It is observed that biologically relevant events, such as protein-folding, domain motion, protein-protein interactions, aggregation or assembly of biomolecules take place at time scales ranging from milliseconds to seconds, which require adjusting the computational microscope at appropriate resolution, which means modeling biomolecules (from fine-grained to coarse-grained levels and from higher resolution to lower resolutions) to obtain longer length and timescales. However, zooming out requires the trade-off of losing atomic detail interactions. In large-scale CG level simulations of 63 different lipids, the lipids were found to distribute asymmetrically in two leaflets, resembling closely a mammalian plasma membrane (60).

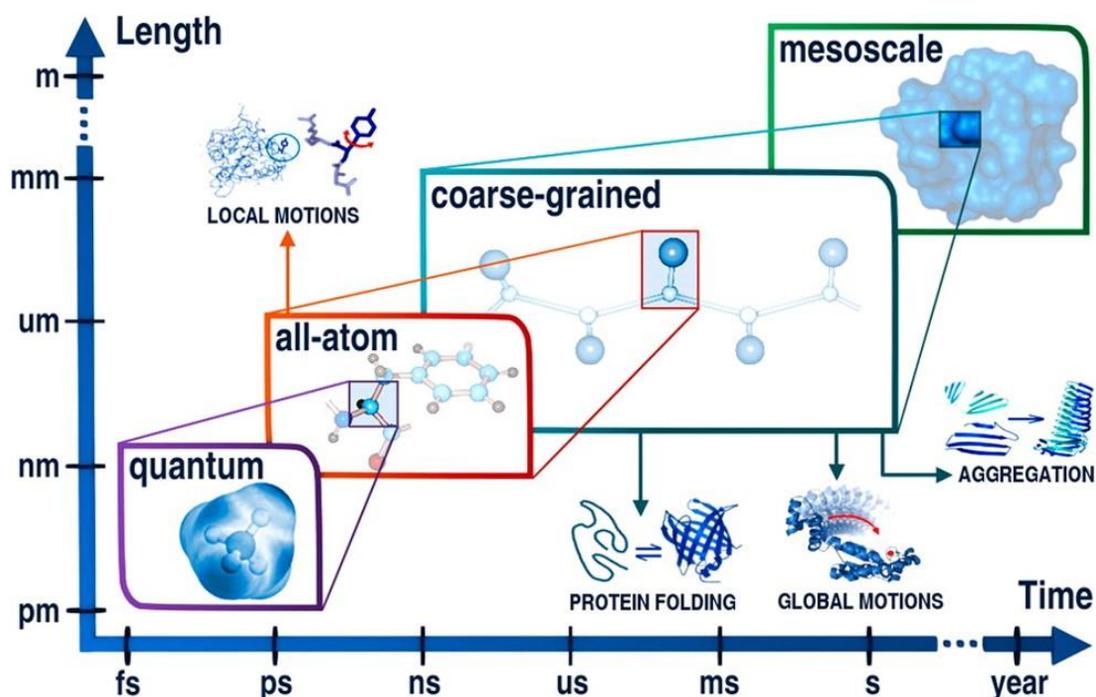


Figure 2.1: Application ranges for molecular modeling at different resolutions: quantum, all-atom, coarse-grained, and mesoscale. The plot shows approximate ranges of time scales and system sizes (lengths). The presented application ranges can be expanded by merging tools of different resolution into multiscale schemes. Reprinted under ACS open access policy http://pubs.acs.org/page/policy/authorchoice_termsfuse.html from (61).

Studying MPs is easier computationally than experimentally due to the fact that MPs adopt unique secondary and tertiary structures. The conformational search space for MPs is limited by the presence of membrane environment (62), and therefore, integral MPs form only two structural motifs in the bilayer: α -helices and β -sheets(63). However, MP 3D structure prediction is still challenging and requires accurate scoring functions and force-field parameters for dealing with the membrane. (63). There have been increasing efforts in developing various computational algorithms which predict from genome sequence the primary amino-acid sequence, and from the primary sequence the secondary structure, the TM spanning regions and the protein topology (α -helical bundles or β -barrels) as well as the different sequence motifs found in MPs. Computational methods such as homology modeling, fold recognition, and de novo predictions are used to predict tertiary structure (3D) of MPs.

In order to understand the effect of the lipid bilayer on an MP's function, various computational simulation approaches are employed, ranging from coarse-grained (CG) to all-atom (AA) molecular dynamics (MD) simulations. MD simulations help in exploring the structural dynamics of MPs in their native environment to bridge the gap between structure and function of these proteins. Advancements in the field of structural biology and experimental techniques in combination with advancements in the computer hardware, software and simulation algorithms have improved our understanding of biological structures, their mechanism and function. In this chapter, I will discuss the state-of-art computational modeling techniques used to predict the 3D structure of MPs and protein-membrane dynamics and interactions. Our focus will be mainly on multi-scale simulation studies, from atomistic to coarse-grained level resolution. I will briefly discuss the homology modeling technique for 3D structure prediction of MPs followed by discussion on computational modeling and simulation approaches.

2.2.1 Homology modeling

An important step in homology modeling is finding a template structure with high sequence similarity. A modeled structure with 70% sequence similarity to the template will have RMSD of 1-2 Å compared to 3-4 Å RMSD structures obtained from templates with 25% sequence similarity, which is twilight range which means the accuracy of model is low in this range. The quality of model also depends on how good the sequence alignment is. An incorrect alignment will result in inaccurate model generation. After target-template alignment, homology modeling tools are used to generate coordinates. The homology modeling tools most commonly used for MPs are: RosettaMP(64), Modeller(65) and Medeller(66), later tool is especially designed for MPs. There are various online servers used to predict MP 3D structures such as MEMOIR which uses Medeller together with other tools(67), HHPred(68) or SwissModel(69). The modeling tools generate coordinates using one of the following approaches: 1) assembly of rigid bodies (70) 2) segment matching or coordinate reconstruction (71), and 3) satisfactions of spatial constraints (65). The final step in homology modeling is

assessment of the model for its quality by using different scores. The above steps, beginning from sequence alignment, model generation and model assessment is performed, are iterated several times to obtain better quality of model (72).

2.2.2 Force-fields

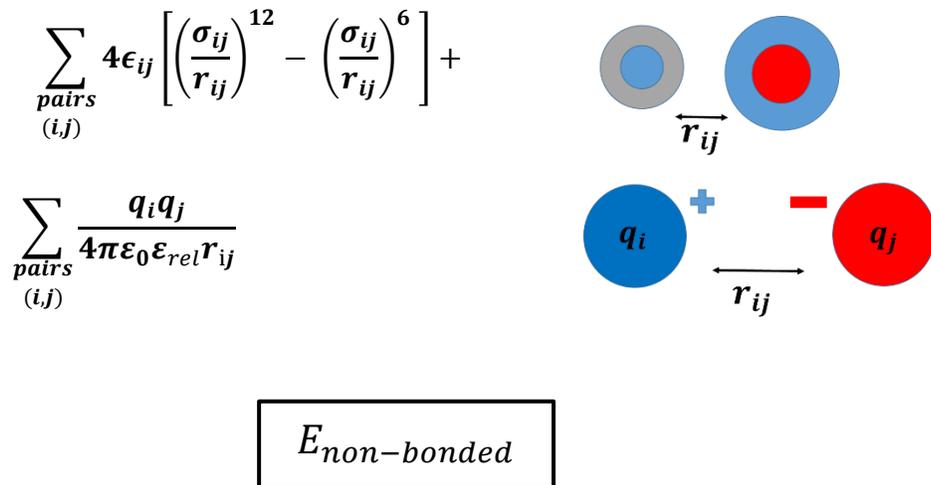
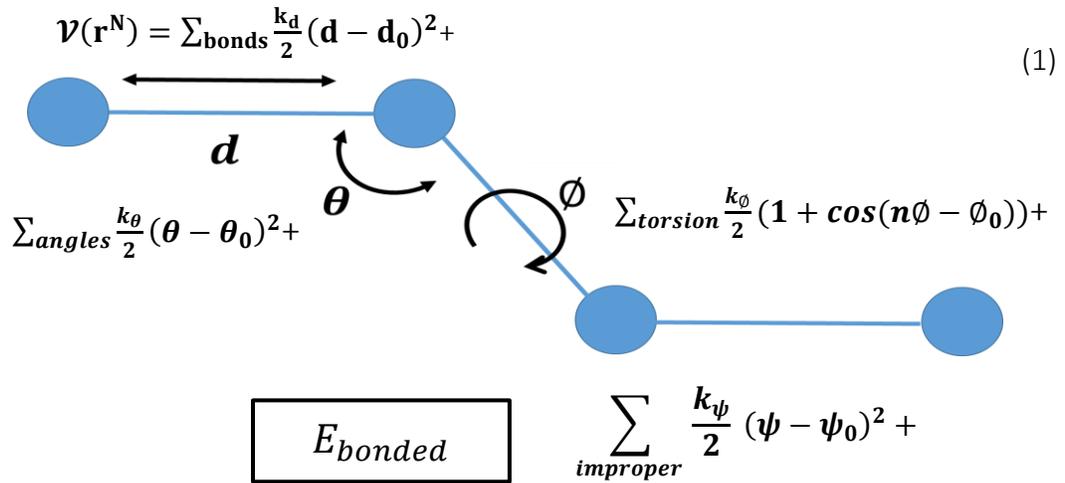
In classical MD, forces between particles and potential energy are calculated by empirically derived potential functions or molecular mechanics force-fields. Most molecular modeling force-fields represent the potential energy function for intramolecular (bonded) and intermolecular (non-bonded) interactions to model the proper geometry and structures of biomolecules. The potential energy change occurs when intramolecular interactions deviate from “equilibrium” or “reference” positions, such as by bond stretching, bond angle bending and rotation of bonds or changes in dihedral angle. The second contribution to the potential energy of the system is the forces on atoms due to non-bonded interactions such as van der Waals interactions and short and long range electrostatic interactions (which fall off slowly with distance).

2.2.2.1 Bonded interactions

In the equation (1), $\mathcal{V}(r^N)$ is the potential energy, r^N is the position of particles N . A harmonic potential with force constant k_d is used to calculate the energy change due to deviation from the reference equilibrium bond length d_0 . The second term in the force-field is a harmonic potential used to calculate the potential energy change due to bond angle bending where θ is the bond angle and θ_0 is the bond angle at equilibrium, and k_θ is the force constant. The third term is the torsional term, used to calculate the potential energy change resulting from rotation about bond. The torsional terms apply to atoms separated by three bonds. k_ϕ is the barrier height, n the number of minima, ϕ_0 is a phase factor which determines the position of the minima. The 4th term ψ is an improper dihedral, ψ_0 is the improper dihedral at equilibrium and k_ψ is the force constant for the improper dihedral angle harmonic potential. The improper term is used to retain planarity and chirality of certain groups. For example, peptide bonds in

proteins and aromatic rings in amino acid side chains. It applies a penalty to bending out of plane.

$$E_{total} = E_{bonded} + E_{non-bonded}$$



2.2.2.2 Non-bonded interactions

The forces arising due to non-bonded interactions are modelled by pair-potentials and are contributed by van der Waals and electrostatic forces. The van der Waals forces can be approximated by Lennard-Jones (LJ) 12-6 potentials which contain attractive (6) and

repulsive (12) terms. The attractive terms originate from dispersion forces caused by instantaneous dipole-dipole interactions, while repulsion results from Pauli's exclusion principle describing electron repulsion within a certain distance (~ 0.3 nm). In equation (1), r_{ij} is the distance between particles i and j . ϵ is the potential well depth, σ finite distance at which potential is zero. Electrostatic potential is sum of electrostatic forces arising from charges on atomic nuclei which is calculated by Coulomb's law. As shown in above equation r_{ij} is the distance between two nuclei i and j . q_i and q_j are the charges on particles i and j . ϵ_0 is the dielectric permittivity of vacuum, and relative dielectric screening ϵ_{rel} , which is 1 in case of all-atom water model and 15 when using MARTINI standard coarse-grained (CG) water model (discussed below).

Optimization of force-fields is done by fitting with quantum mechanical calculations and experimental spectroscopic and crystallographic data. In the early 80s, biomolecular force-fields and programs were developed such as AMBER, CHARMM and OPLS which are still widely in use. The software packages utilizing these force-fields include AMBER, NAMD and CHARMM, and GROMACS respectively (73). Although basic form of all force-fields is almost same, the accuracy may vary. The accuracy of force-field is major limiting factor in MD simulations, and due to increased computational resources, improved hardware and parallelization have enabled computationally expensive validation of force-fields against experimental data (51).

2.2.3 MARTINI coarse-grained force-field

One of the most widely used CG models is the MARTINI force field implemented in the GROMACS MD program (74). The MARTINI force field was initially developed for lipids in 2004, improved in version 2.0 in 2007,(75, 76) and extended to proteins and peptides in MARTINI version 2.1 released in 2008.(77) A re-parameterization of the MARTINI protein force field to version 2.2(78) and the introduction of a polarizable water (PW) model resulted in MARTINI version 2.2P(79). Currently, MARTINI offers parameters for a wide variety of lipids molecules, cholesterol, protein, DNA, carbohydrates, fullerene, collagen and dendrimers(80).

2.2.4 MARTINI Coarse-grained mapping

The CG approach used in MARTINI is based on a 4-to-1 mapping, with 4 heavy atoms plus associated hydrogen atoms represented by one CG bead. Four water molecules are represented by 1 CG bead in both the original standard non-polarizable water (NPW) model and the new PW model. CG ions are represented by one bead consisting of an ion and its hydration shell. A ring like molecule such as cholesterol, benzene and aromatic side chains of amino-acid residues are mapped using a 2 to 1 instead of 4 to 1 mappings.

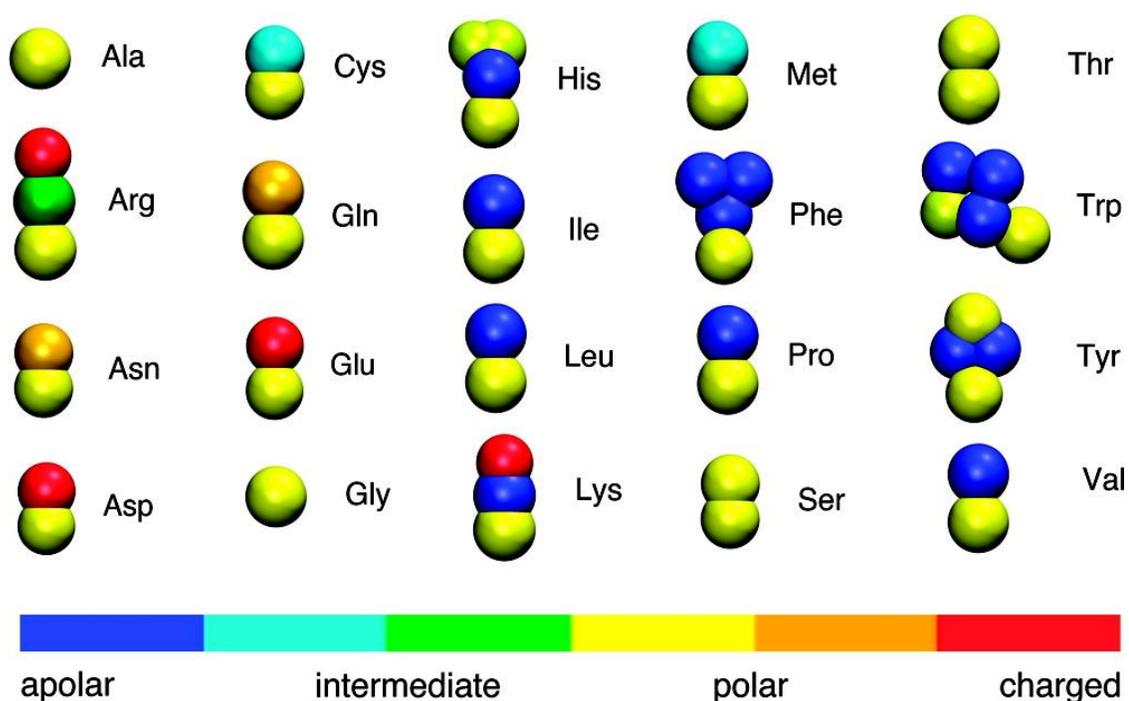


Figure 2.2: Coarse-grained representation of all amino acid residues. Different colors represent different particle types (81). Backbone is represented by yellow color.

In order to reproduce chemical properties, the MARTINI model consists of 4 main types of CG particles: polar (P), nonpolar (N), apolar (C) and charged (Q) and, within the 4 main types, there are subtypes based on hydrogen bonding ability (d=donor, a=acceptor, da=both, 0=none) or degree of polarity (1= low polarity, to 5=high polarity) (57) making a total of 18 bead types.

2.2.5 Bonded interactions in MARTINI

Bonded interactions in MARTINI CG models are described by the same potential energy functions as in classical force-fields. The potential energy functions contain a harmonic potential used for bonds, angles, dihedral and improper dihedral terms. Dihedral potentials are used for preserving secondary structure and to prevent out-of-plane distortion of planar groups (aromatic side chains in amino-acids). Bonded interactions are directly parametrized using structural data from atomic structure or by comparison with atomistic simulations(57). For proteins, the equilibrium values for bonded potentials (bond, angles and dihedrals) are optimized against 2000 atomistic reference geometries of proteins from Protein Data Bank (PDB)(82). The secondary structure information is provided additionally using the DSSP (“define secondary structure of proteins”) prediction algorithm(83). The secondary structure in MARTINI protein models remains restrained using additional restrains applied via elastic network to keep protein close to native fold.

2.2.6 Non-bonded interactions in MARTINI

The MARTINI force-field was developed by optimizing non-bonded interactions to reproduce experimental thermodynamic data such as free energies of hydration, free energies of vaporization, and partition free energies between water and a collection of organic phases, such as hexane, octanol and chloroform, for 18 bead types. Partition free energy is calculated from equilibrium densities ρ of CG particles in two phases ($\Delta G^{oil/water}$).

$$\Delta G^{oil/water} = k_B T \ln(\rho_{oil} / \rho_{water}) \quad (2)$$

The non-bonded interactions are described by Lennard-Jones (LJ) 12-6 potential (as shown in equation (1)) and interactions between charged particles, such as lipid head groups and charged amino-acid side chains, are calculated using Coulomb’s law. The strength of the interactions for each particle varies and is determined by well-depth ϵ_{ij} . ϵ_{ij} ranges from 5.6 KJ mol⁻¹ to 2.0 KJ mol⁻¹ for strongly polar groups to apolar groups

respectively (57). $\sigma=0.47$ nm describes the size of the particles except for ring like atoms where reduced $\sigma=0.43$ nm and ϵ_{ij} scaled down to 75% of its original value to model ring-ring interactions. The details of interaction matrix is given in the original article (84).

In CG simulations using standard water model or non-polarizable water (NPW), relative dielectric screening $\epsilon_{rel}=15$ is used, as shown in equation (1), which was decreased from its initially value of $\epsilon_{rel}=20$ (77).

In MARTINI force-field, non-bonded interactions are cutoff at $r_{cut}=1.2$ nm. The LJ interactions are truncated using a shifted potential from 0.9nm to 1.2nm while the electrostatic interaction potential is shifted from 0.0nm to 1.2nm. Recently, the MARTINI force-field has been tested using simple cut-off at 1.1 nm for LJ and Coulomb's interactions with a "potential modifier" implemented in a recent version of the Gromacs software package (5.0.x) (85, 86). It was observed that the use of a simple cut-off did not affect the system properties, such as partition free energies which remained within the range of the differences between the Martini force-field and experimental values. The use of a straight cut-off offered significant speedup in performance compared to shift method. Electrostatic interactions were treated by the RF method instead of the cut-off which gave slightly better results at negligible computational cost (86).

2.2.7 MARTINI protein and Elastic Network models

An Elastic Network (EN) model is used in combination with the MARTINI CG protein model to maintain the structure of biomolecules. In EN model, springs force constant K_{SPRING} are used to connect a network of point masses when the inter-particle distance is less than a predefined cutoff R_C . The MARTINI force-field together with the EN model, named EIneDyN, has been used to determine the structure and dynamics of the CG model with respect to atomistic reference simulations (87). In ELNEDIN model is applied on C α backbone bead instead of center of mass the, -N,C α ,C,O-, used in earlier version. EN parameters, K_{SPRING} and R_C were used ranging from 0.8 to 1.0nm and 500

to $1000 \text{ KJ mol}^{-1} \text{ nm}^{-2}$ respectively which provide agreement with atomistic simulations (87).

2.2.8 MARTINI coarse-grained water models

The MARTINI standard water (NPW) model does not contain charges and interacts by means of LJ interactions only. To prevent the CG water from freezing at 250 to 300 K, antifreeze particles are included in the NPW solvent (76). The use of the NPW model requires a relative dielectric constant $\epsilon_{\text{rel}} = 15$ offering uniform screening which is a reasonable approximation for bulk water. However, electrostatic screening varies at the water-lipid interface or around charged amino acids which is difficult to maintain using NPW model using $\epsilon_{\text{rel}} = 15$. Recently, a polarizable water (PW) model has been developed (79), which consists of three particles W, WP and WM. The two latter particles are positive and negative point charges, respectively, bonded to the central particle, W. The charged particles are able to rotate around W to model polarization. The central particle W is neutral, like NPW, and interacts with other particles by LJ interactions (79). The use of the PW model improves the modelling of electrostatic interactions with charged particles and the polarization effects of water. The ϵ (dielectric constant) of the PW model was computed between 75.6 to 77.1 at $\epsilon_{\text{rel}} = 2.5$, which is close to experimental value 78.4 at 298 (79). The reduced $\epsilon_{\text{rel}} = 2.5$ used with PW keeps dielectric constant more realistic in hydrophobic regions. The PW model reproduces the oil/water partition free energy, the freezing temperature and the free energy of pore formation in the membrane better than the NPW model. However, its effects on protein behavior were not explored when it was developed(79).

2.3 Molecular dynamics simulation

Molecular dynamics (MD) simulation is one of the most widely employed molecular mechanics techniques to study biochemical processes such as protein-folding, drug binding, membrane self-assembly, membrane transport, and conformational changes responsible for protein functions (51). Methods in molecular mechanics such as energy

minimization, normal mode analysis and Monte-Carlo simulations do not provide direct information about time dependence of dynamics of the system. Brownian dynamics (BD), dissipative particle dynamics (DPD) and molecular dynamics (MD) simulations methods are based on calculations of positions and motions of particles using equations of motions. MD simulations of large macromolecular systems, membrane protein complexes comprising more than tens of thousands of atoms, and smaller systems on longer time scales up to milliseconds have been performed to observe many critical biochemical processes.

MD simulation is based on integrating Newton's equation of motions ($\mathbf{F} = m \mathbf{a}$) numerically for a set of particles as a function of time. Given the initial position and velocity of each particle i , one can calculate the forces and therefore the acceleration \mathbf{a}_i at the time point t . The iteration of the equation gives the spatio-temporal evolution of the system (59).

2.3.1 Integration of equation of motion

In MD simulations, the forces acting on particles, positions and velocities are updated by integrating equations of the motion, which is computationally intensive and one of the limiting factor in the sampling the conformational changes (59). At each MD step energy, forces and velocity are calculated and updated for the next time step. A variety of integration algorithms is being used to integrate equation of motion at finite time steps, where new positions $\mathbf{r}_i(t + \Delta t)$ and acceleration at time $t + \Delta t$ are determined from previous step $\mathbf{r}_i(t)$ (88).The most common algorithms are: Verlet algorithm (89), its derivative velocity Verlet algorithm (90, 91) and the Leapfrog algorithm(92). These algorithms are designed to offer accuracy, efficiency, time-reversible and symplectic. These integrators are time-reversible, meaning that if the direction of velocity is reversed, the simulation will run in the reverse direction.

2.3.2 Treatment of non-bonded interactions

Biomolecular force-fields include long range electrostatic interactions where the summation of order of N^2 is required to account for all non-bonded interactions (93). In MD simulations, forces are computed at each time step and, long range interactions are the most computationally expensive in MD algorithm. In order to increase computational speed, a cut-off scheme is used to calculate non-bonded interactions between nearby atoms and which are truncated after cutoff distance (0.8-1.4nm) (94). However, abrupt disruption of forces at cutoff results in energetic discontinuity leading to unstable simulation (88). Smoothing functions such as Shift or Switch functions are used to decrease the energy gradually to zero at given cutoff value. The use of Shift or Switch functions for LJ interactions is suited well and does not cause problems as dispersion forces are short ranged and decay quickly.

The treatment of electrostatic interactions is a key issue in MD simulations. In particular, use of cutoff method in simulations of lipid-bilayer does not result in appropriate structure because of the high charge density at interface and a region of low dielectric constant which results in weakly shielded electrostatic interactions. The most common method to study long range interactions is particle mesh Ewald (PME)(95). PME is based on interpolation of the reciprocal-space Ewald sum. In PME, all the interactions are summed in a periodically replicated system including long range interactions. However, periodicity may cause artifacts due to artificial ordering leading to enhanced stability of system. The reaction field (RF) is another method to calculate long range electrostatic interactions. It was originally developed to simulate homogenous liquid systems. It adds a correction term to the cutoff results as solvent outside is considered as a continuum electrostatic. In the RF method, a charged particle beyond a certain distance (cut-off) does not see individual charges but rather the averaged homogenous medium of electrostatic field. It has been observed in simulations of a DPPC (Dipalmitoylphosphatidylcholine) lipid-bilayer that the reaction field gives similar results to PME (96). All the above methods have their own drawbacks when simulating

interfacial systems. The use of a cut-off creates artificial ordering, PME increases periodicity and RF neglects the heterogeneous nature of a membrane system. However, PME is generally considered to be the best choice with the fewest drawbacks (94). Force-fields are developed using one of these methods. Therefore, changing to a different method to calculate electrostatic interactions may affect the accuracy of the force-field.

2.3.3 Periodic boundary conditions

Molecular Dynamics simulations are often performed using periodic boundary conditions (PBC) to create bulk effect (an infinite system). In PBC, a virtual image of a box is replicated in all directions. The most common choices for PBC are the cubic and rectangular boxes or truncated octahedral box which is often used to decrease the solvent volume in the corners. In 3-dimensional space of a cubic box, PBC replicates into the 26 nearest neighbor boxes. If a particle exits the box from any side during simulations, it is replaced by the image particle entering from the opposite side. Thus, the total number of particles in the simulation box is kept constant. The use of PBC removes artifacts created due to surface effect (97).

2.3.4 Simulation conditions

Like in laboratories, *in silico* experiments can also be performed by selecting different thermodynamic conditions. MD simulation is the study of thermodynamic and kinetic properties of molecular systems. Thermodynamic state of the system is defined by macroscopic property of the system which depends on parameters such as the temperature T , pressure P and number of particles N . MD simulations are performed using different microscopic ensembles such as: canonical (NVT), isothermal-isobaric (NPT), microcanonical (NVE) and grand-canonical ensembles (μVT). The letters denote macroscopic observables which are kept constant such as number of atoms, N , volume, V , pressure, P , temperature, T and chemical potential, μ . Direct integration of Newton's equation of motion results in NVE ensemble while experiments are usually performed at constant temperature and volume and/or constant temperature and

pressure *NVT or NPT* (88). In the *NVT* ensemble, temperature is kept constant using an external thermal bath. There are different thermostat algorithms to keep temperature constant at the desired reference temperature. For example: Berendsen weak-coupling method (98), v-rescale (99), Nosé-Hoover (100), Langevin piston (101) etc. In the case of the *NPT* ensemble, an external barostat is used to control pressure while allowing the volume of the system to change by rescaling box-dimensions. The most commonly used barostat algorithms are Berendsen weak-coupling (98) and Parrinello-Rahman (102) methods.

2.3.5 Pressure control for membrane protein simulation

When simulating a membrane protein system, one of the most important parameters to deal with is pressure coupling. There are three methods of pressure coupling: 1) isotropic pressure coupling, where the x,y and z dimensions are coupled together; 2) semiisotropic pressure coupling, where x and y directions are coupled together while pressure in the z-direction is allowed to fluctuate independently; and 3) anisotropic pressure coupling which does not couple any directions of the pressure contribution (see Figure 2.3).

In the case of membrane simulations, semiisotropic pressure coupling is the recommended method which allows area fluctuations. The semiisotropic pressure coupling in x-y directions does not produce constant surface area, but the size adjustment in x-y dimensions allows area fluctuation in both dimensions equally. Therefore, the ratio of the box-size remains constant. On the other hand, anisotropic pressure coupling allows fluctuations in all directions independent of each other, which may result in large deformations of the whole simulation system. Anisotropic pressure coupling should be used carefully. While isotropic pressure coupling is not recommended, which leads to very small changes in box size. Isotropic pressure coupling is inappropriate as it does not allow fluctuation in surface area and does not specify surface tension (103).

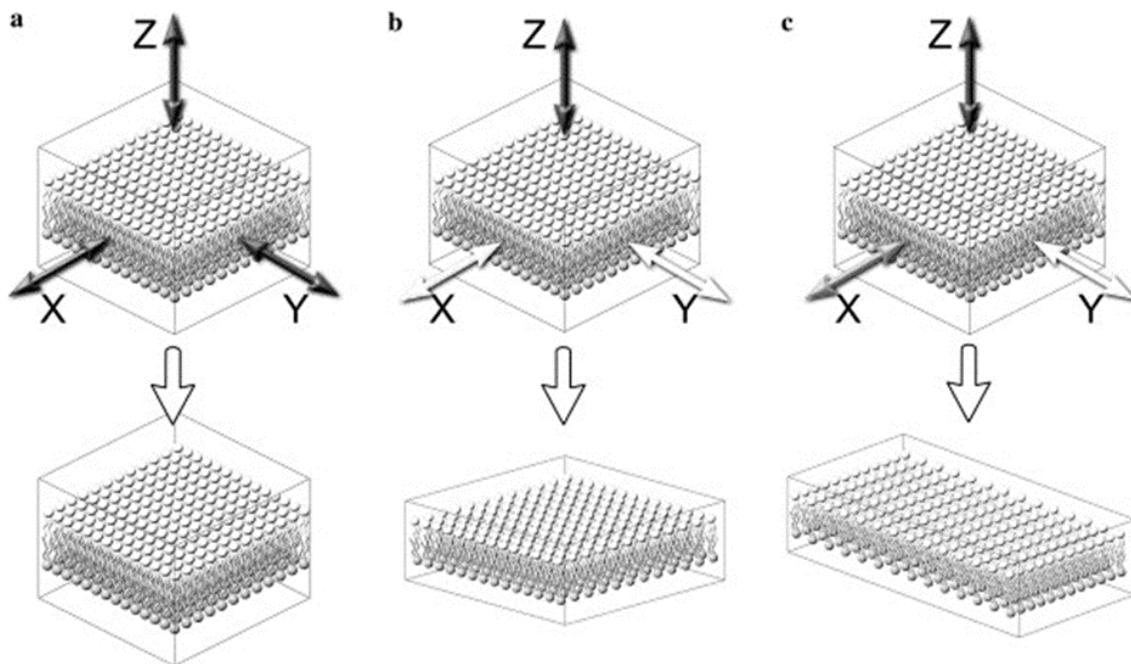


Figure 2.3: Different pressure coupling schemes and their effect on a membrane bilayer. a) Isotropic pressure coupling which does not allow large fluctuations. b) Semiisotropic or constant surface tension pressure coupling (recommended) for membrane system which allows area fluctuations as the pressure contribution in the x-y directions is coupled together while the z-direction is not. c) Anisotropic pressure coupling allows area fluctuations in all directions. Taken from the article (103).

2.4 Multiscale simulations of membrane-protein system

2.4.1 All-atomic molecular dynamics simulation

In classical MD, the all-atom model is coarse-grained in the sense of removing electronic degrees of freedom and treating electrons and nucleus as one sphere. For all-atom MD (AAMD) simulations, the forces contributed by all atoms in the molecule including non-polar hydrogens are computed. Due to fast vibrational frequency of H-bonds, time step to capture these motions is limited to 1-2 femtoseconds (59). The integration of equations of motions at such a small time-step currently limits AAMD simulations to study a system consisting of maximum 10^6 and timescale of one microsecond. AAMD simulations is limited to sample only local energy landscape. Therefore, it is still challenging to observe large conformational changes, biological events occurring at

longer timescale ($> \mu\text{s}$), and larger systems using equilibrium atomistic simulations. However, compared to quantum mechanics AAMD simulations offers longer computation time and simulation of larger system.

2.4.2 Coarse-grained simulation

Studying large conformational changes and larger system such as membrane protein complex requires zooming out the “computational microscope” to more coarse-grained (CG) resolution. CG simulations of biomolecules dates back to earlier modeling approach used to study protein folding (104). Recently, CG has gained much attention and is being used as part of a multiscale, coarse-grained combined with all-atomic (CG/AT) resolution, modeling approach. Michael Levitt, Ariel Warshel, and Martin Karplus were awarded Nobel prize in Chemistry in 2013 “For the development of multiscale models for complex chemical systems”, which included CG modeling of proteins (61). A widely-used level of coarse-graining for proteins and lipids use approximately 4 non-H atoms substituted by one particle. This simplified model allows a substantial increase in the simulation time. Due to decreased degree of freedom and smoother energy landscape, longer integration time steps (10-100fs) is used which result in the faster dynamics. The interactions in the CG models are rather short range where longer range electrostatic interactions are avoided which gives significant speedups of CG simulations. There is an increasing number of CG models developed for studying protein folding, protein-protein docking, protein structure predictions, peptide membrane interactions, protein-protein aggregation, lipid bilayer domain formations, bilayer self-assembly and formation of a lipid bilayer around MPs. These CG models are developed using different modeling strategies based on their applications as given in table 2.1. The major differences being in the degree of coarse-graining (number of beads representing backbone and side chains), explicit/implicit treatment of solvent and treatment of non-bonded interactions (105).

The CG modeling strategy is mainly divided in to “bottom up” force-matching and “top down” free-energy based approaches (106). The bottom up approach is based on

parametrization of models by comparing with atomic simulations or experiments to systematically derive the CG interactions. While top down approach is employed by MARTINI CG models in which chemical building blocks are extensively calibrated against thermodynamic data, such as partitioning free energies between polar and apolar phases (oil/water) (57). Thus, it offers a broader range of applications and transferability. The MARTINI CG models for lipids, used in this work, designed by Marrink group (75, 84) are also compatible with protein CG models developed by Schulten (107) and Sansom (108) groups.

Table 2.1: Different coarse-grained models used to study biomolecular systems. Modified table from chapter 5 of the book on “Advances in Protein Chemistry and Structural Biology”, Biomolecular Modeling and Simulations, Volume 96, (109).

Model	No of Particles		Membrane	Protein	Solvent	Speedup over AAMD
	*Back- Bone	*Side- Chains				
UNRES	1	1	-	X	Implicit ^a	~1000
MARTINI	1	1-5	X	X	CG ^b	75-100
OPEP	5	1	-	X	Implicit	30-40
SCORPION	1	1-2	-	X	CG	100-130
PaLac	3	1-2		X	CV ^c	~1000
PRIME	3	1	-	X	Implicit	~1000
PRIMO	3	1-5	X	X	GBMV ^d	10-15
Bereau and Deserno	3	1	-	X	Implicit	NA

* No of CG particles/beads

^a Implicit solvent included within non-bonded parameters

^b CG water model ^c Circular variance ^d Generalized Born/molecular volume

2.4.3 From coarse-grained to all-atom (CG2AA)

Recently, multiscale models have become popular, where coarse-grained in combination with fine-grained (all-atomic) resolutions, are used. CG simulations offers self-assembly and large conformational changes while AA simulations are used to study protein-membrane interactions at atomistic details. Thus, conformational orientation of MPs obtained by CG simulations provide starting structure for all-atomic simulations. Back-mapping procedures have been developed to obtain atomistic details from configurations achieved by CG simulations (110). The back-mapping strategy is useful to refine the lipid-binding sites on proteins (111, 112), membrane solvation of nanoparticles or transmembrane pores formed by antimicrobial peptides .

CHAPTER 3: CYTOCHROME P450 MEMBRANE ASSOCIATION: AN OPTIMIZATION OF THE MARTINI COARSE-GRAINED PROTOCOL

This chapter is mainly based on the published work in “The Journal of Chemical Physics (2016)” a Special Topic Issue on “Coarse Graining of Macromolecules, Biopolymers, and Membranes”; doi: <http://dx.doi.org/10.1063/1.4936909>, with contributions by authors G. Mustafa, P. P. Nandekar, X. Yu, and R. C. Wade (111).

Note: P. P. Nandekar performed CG self-assembly simulations of pure membrane system and compared the effect of varying parameters on computational efficiency, computed membrane properties of pure lipid bilayer and helped in analysis of simulation results.

3.1 Introduction

Human drug-metabolizing CYPs are anchored in the endoplasmic reticulum membrane by an N-terminal helix. The N-terminal helix of CYPs is truncated for crystal structure determination and in some cases, hydrophobic residues in the globular domain that interact with the membrane are mutated to increase the solubility (27, 113). The association of the CYP globular domain with the membrane has therefore been studied by other experimental techniques, such as site-specific antibody experiments,(114, 115), tryptophan fluorescence quenching and atomic force microscopy(30). The orientation of a few CYPs in the membrane has been experimentally determined by measuring the tilt angle of the active site heme by rotational diffusion measurements and found to vary from 38° to 78°(116). More recently, the heme-tilt angle of CYP3A4 bound in a POPC nanodisc membrane was determined by linear dichroism as 59.7±4.1°(117).

The orientations of different CYPs may vary depending on their primary sequence, 3D-fold or binding to other proteins or drugs. CYPs provide a particularly challenging case for prediction of membrane insertion because they possess a flexible linker of variable

length between the N-terminal transmembrane helix and the globular domain, and because the globular domain dips into the bilayer using a part of the protein (including the FG loop region) that is highly variable in sequence and structure. Thus, although CYP isoforms possess the same overall fold, they may show different interactions with membranes. In general, predicting the membrane interactions of monotopic membrane proteins or monotopic protein domains of transmembrane proteins, like the CYPs, is more difficult than for multitopic transmembrane proteins with more transmembrane helices, such as G-protein coupled receptors, which usually have a more obvious orientation in the bilayer enabling straightforward placement in the membrane for simulation studies (118–120).

Various computational methods have been used to predict the position, insertion and orientation of membrane proteins. For example, the Orientation of Protein in Membranes (OPM) database (<http://opm.phar.umich.edu/>)(121) and the Protein Data Bank of Transmembrane Proteins (<http://pdbtm.enzim.hu/>)(122) provide protein positions in a bilayer. However, these methods treat the membrane as a hydrophobic slab and do not account for charge or polar interactions between lipid head-groups and the protein or aqueous solvent. Alternatively, MD simulations have been used to study interactions at the atomic level and the conformational dynamics of transmembrane proteins in their physiological environment(103, 123). Many important biological processes require time scales of micro to milliseconds or longer and, therefore, the time scale and system size pose limitations on the use of all-atom molecular dynamics (AAMD) simulations. In comparison to AAMD simulations, Coarse-grained (CG) simulations allow greater exploration of conformational space by decreasing the number of degrees of freedom and allow the extension of simulations beyond the microsecond scale(124, 125). CG simulations are widely used to study large conformational changes such as bilayer self-assembly, the insertion of proteins in membranes and membrane protein folding(124–131). Sansom and colleagues have used CG bilayer self-assembly around proteins to insert about 2000 transmembrane

proteins into a bilayer and created a database of inserted membrane proteins MemProtMD (<http://sbc.bioch.ox.ac.uk/cgdb/>)(132, 133). Multiscale simulations, employing CG and AA simulations, have been used to study the orientations of several CYPs (CYP2C9, CYP3A4 and human CYP51) in a bilayer (31, 112, 134).

One of the most widely used CG models is the MARTINI force field implemented in the GROMACS MD program(74). The MARTINI force field was initially parameterized for use with the Shift function implemented in GROMACS versions up to 4.5.5 and below, where non-bonded interactions beyond the cutoff distance of 1.2 nm are neglected. In GROMACS versions 4.6.x and above, which have enhanced parallelization algorithms, the Shift/Switch algorithm for non-bonded interactions has been deprecated. The newer versions support a cutoff for short-range interactions and long-range electrostatic interactions are instead treated with Particle Mesh Ewald (PME) or a reaction field (RF)(85). Particle Mesh Ewald (PME) and the Shift function have been tested for lipid bilayer simulations with the polarizable water (PW) model, showing similar lipid properties such as area per lipid (APL)(79). Here, we have evaluated the effect of different MARTINI water models standard water model or non-polarizable water (NPW) and PW models, different treatments of long and short-range interactions (both van der Waals and electrostatic interactions) for CG simulation of protein-bilayer systems, and the effect of using different cutoff schemes implemented in different versions of GROMACS 4.5.5 vs 4.6.X i-e., a group vs Verlet cutoff schemes respectively.

In the present study, we use multiscale modeling for the insertion and orientation of CYP3A4 in a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) bilayer by MARTINI CG and AAMD simulations. The CG simulations are used to generate large conformational changes and find a converged orientation of CYP3A4 in the membrane, whereas AAMD simulations are used to refine the protein-membrane interactions at an atomically detailed level. The CYP3A4 orientation in the membrane has been predicted previously by using MD simulations as well as by experiments. Therefore, it provides an excellent test system for evaluating different protocols. We identify an optimum

procedure in terms of computational efficiency and the quality of the results for CYP3A4. We then show that the same protocol can be successfully applied to four other human drug metabolizing CYPs (1A1, 1A2, 2C9 and 2C19). We expect this approach to be applicable to other membrane proteins that have monotopic globular domains that dip into lipid bilayers.

3.2 Material and method

3.2.1 All-atom models of CYP3A4

The crystal structure of human CYP3A4 (PDB ID: 1TQN) was retrieved from the Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>). The crystal structure was resolved at 2.05 Å resolution by truncating the N-terminal residues from 1 to 27(113). The missing residues were modelled using Modeller9.10 (65) by applying constraints on residues 3-28 to form an alpha-helical conformation as predicted by the advanced protein structure prediction server (APSSP2: <http://www.imtech.res.in/raghava/apssp2/>). The complete structure of CYP3A4 consists of an N-terminal loop (residues 1-2), the N-terminal transmembrane (TM) region (3-28), a linker region (29-55) which includes the A' (31-36) and A'' (50-55) helices, and a globular domain (56-499). Five different orientations of the globular domain with respect to the TM region were generated by changing the dihedral angles of the residues in the linker region to obtain a diverse range of arrangements of the globular domain that would not clash with the membrane and that cover the available configurational space. A similar procedure was employed for the other CYPs simulated. The PDB coordinates used were: 2C9:1R9O, 2C19:4GQS, 1A1:4I8V and 1A2:2HI4.

3.2.2 Preparation of coarse-grained (CG) systems

All-atom models of the protein were converted to MARTINI CG models using the martinize.py script(110) available from <http://cgmartini.nl>. A pre-equilibrated CG TM-helix obtained from simulating CYP51(112) in a CG POPC membrane bilayer consisting of 594 POPC molecules (31) was used to superimpose the TM-helix of CYP3A4. In order

to maintain the secondary and tertiary structure of the protein during the simulation, harmonic restraints were applied to the backbone atoms of the protein with an elastic force constant of $500 \text{ kJ.mol}^{-1}.\text{nm}^{-2}$ and a distance cut-off of 5 to 9 Å (lower and upper limit) (87). Elastic restraints were removed from residues (28-36) to allow free movement in the linker region during simulations. The secondary structure information was provided in a DSSP file from the DSSP server (www.cmbi.ru.nl/dssp.html). Each protein-bilayer complex was solvated in a periodic box of water particles using the same x-y box dimensions as the POPC bilayer and the z dimension extending 13 nm above and 7 nm below the center of the bilayer (with the bilayer in the x-y plane and the z axis normal to the bilayer). The system was neutralized by adding counter ions.

3.2.3 Coarse-grained simulations using different parameters

In the present study, we have tested different simulation parameters for our optimization protocol for CG simulations. A comparison between Shift, PME and RF for calculations of Coulombic interactions and of different pressure coupling methods (Parrinello-Rahman vs Berendsen) for both PW and NPW models for the CYP-bilayer system was performed. For reference, we also simulated the CG self-assembly of a POPC bilayer without protein with the PW and NPW models to distinguish effects due to the protein.

All CG simulations were performed using the MARTINI CG force field for membrane, protein and solvent systems under periodic boundary conditions as implemented in the GROMACS software package (<http://www.gromacs.org>)(74, 85). The simulations were performed using two different water models, the standard non-polarizable water (NPW) and the polarizable water (PW) model with GROMACS versions 5.0.4 and 4.5.5 (Table 3.1). In GROMACS 5.0.4, the Verlet cutoff-scheme with automated buffered tolerance is used to generate the pair list for calculation of non-bonded forces. The group cutoff scheme is used in older GROMACS versions (4.5.5 and below) to generate the neighbor list of groups of particles (charge groups) with neighbor list distance

greater than the cutoff values (see the GROMACS Manual for details (www.gromacs.org/Documentation/Manual)).

The simulations started with a short steepest-descent energy minimization until the maximum force was less than 10 kJ/mol*nm (10,000 steps). The system was equilibrated in the NPT ensemble at a constant temperature of 310 K, where the temperature of the protein, POPC and the solvent was coupled separately by a velocity rescale (v-rescale) thermostat with a coupling constant of 1 ps. During the simulations, pressure was kept constant by a Parrinello-Rahman barostat with a coupling constant of 12 ps and a reference pressure of 1 bar. Semiisotropic pressure coupling was used with a compressibility of 3.0×10^{-5} . The time step was 20 fs.

The long-range non-bonded interactions were calculated by using RF or PME in GROMACS 5.0.4 and compared with the Shift function in GROMACS 4.5.5. In the CG PW simulations with PME, the Fourier grid spacing was set to 0.33 nm except in test simulations which were run with the Fourier spacing set to 0.12 nm. A relative dielectric constant of 15 was used with the NPW model and of 2.5 with the PW model. The default value for Verlet-buffer-tolerance (0.005 kJ/mol/ps) was used and the frequency of the neighbor list update was set to 10 steps. With the Verlet cutoff scheme, a potential shift was applied to the Coulomb and LJ interactions which shifts the potential by a constant such that potential is zero at the cutoff of 1.2 nm. With the group cutoff scheme, Coulomb interactions were shifted smoothly to zero between 0 and 1.2 nm and LJ interactions were smoothly shifted to 0 between 0.9 to 1.2 nm.

Table 3.1: CG Simulations performed for five CYPs.

CYP3A 4 (PDB)	Residues in regions			Water Model	NB [#] treatment	ϵ_{rel}^*	No of Simulations	GROMACS
	T M	Linke r	Globular					
1TQN	1-28	29-55	56-499	PW	Shift	15	5	4.5.5
1TQN	1-28	29-55	56-499	PW	RF	2.5	2	5.0.4
1TQN	1-28	29-55	56-499	PW	PME	2.5	3	5.0.4
1TQN	1-28	29-55	56-499	NPW	RF	15	5	5.0.4
1TQN	1-28	29-55	56-499	NPW	PME	15	1	5.0.4

#Non-bonded, *relative dielectric constant

3.2.4 Back conversion from coarse-grained to all-atom model

The CG simulations were performed to find the predominant arrangement of the protein with respect to the membrane. After convergence of CYP3A4 in the membrane obtained by CG simulations using NPW and RF, a representative frame was selected from each of the five CG simulations. The snapshot was required to have values for the geometric parameters monitored (see Analysis section below) within 1% of their mean value(112). The representative frame was then converted to an all-atom model. The procedure used for conversion of the POPC bilayer is described in Cojocaru *et al.*(31). For conversion of the globular domain, the TM helix, and flexible linker region, I used the back-mapping procedure using scripts backward.py and initram.sh, available at MARTINI website (<http://cgmartini.nl>) (110). The X-ray structure's globular domain residues 56-499 were superimposed on the back-mapped globular domain of CYP3A4. Then the TM-helix and flexible linker region from the back-mapped model were added to the X-ray globular domain and the resulting full length all-atom model was placed into the all atom model of the POPC bilayer. The heme cofactor, which was omitted in

the CG simulations, was thus added to the AA system by superposition from the crystal structure of the protein.

3.2.5 All-atom molecular dynamics simulation

AAMD simulation was performed for about 40 ns to relax the structure of the protein in the membrane. In the AAMD simulations, the ff12SB force field was used for protein residues(135), POPC lipids were simulated using the Lipid14 force field(136), and the heme parameters were provided by D. Harris with the partial atomic charges derived from DFT calculations(137). The ionic concentration was maintained at 150 mM using Na⁺ and Cl⁻ ions in a periodic box of TIP3P water molecules. Unlike for previous AMBER lipid force fields, the updated Lipid14 force field was used without applying surface tension in the simulations(136). The procedure for AAMD simulation was described in detail by Cojocaru *et al.*(31). Initially, the system was energy minimized with the Amber12 software with a decreasing force constant from 1000 to 0 kcal/mol·Å² on harmonic restraints on non-hydrogen atoms of the protein and lipid residues. The system was then equilibrated in NAMD2.9 at constant area and pressure at a temperature of 310 K (NPAT) for 1.5 ns with a time step of 1 fs by reducing restraints by scaling the force constant from 100 to 0 kcal/mol·Å² on non-hydrogen atoms of protein residues. The temperature was maintained at 310 K during simulations. Then the system was equilibrated for 5 ns without any constraints in the NPAT ensemble. Afterwards, the production simulation was run for 27.5 ns with a time step of 2 fs in an NPT ensemble with periodic boundary conditions. The electrostatic interactions were calculated using the PME method. All bonds to hydrogen atoms were constrained with the SHAKE algorithm(138). 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.

3.3 Data analysis

3.3.1 Angle and distance calculations

The orientation of the CYP globular domain above the membrane was quantified by calculating various angles and distances, similarly to our previously reported work on CYP2C9 and CYP51 (31, 112). Different vectors were defined: v_1 , from the center of mass (CoM) of the backbone particles or atoms of the first 4 residues to the CoM of the last 4 residues of the I-helix, v_2 from the CoM of the first 4 residues of the C-helix to the CoM of the last 4 residues of the F-helix, v_3 , the vector between the CoMs of the first and last four residues of the TM helix, and the z-axis perpendicular to the membrane (Figure 3.1).

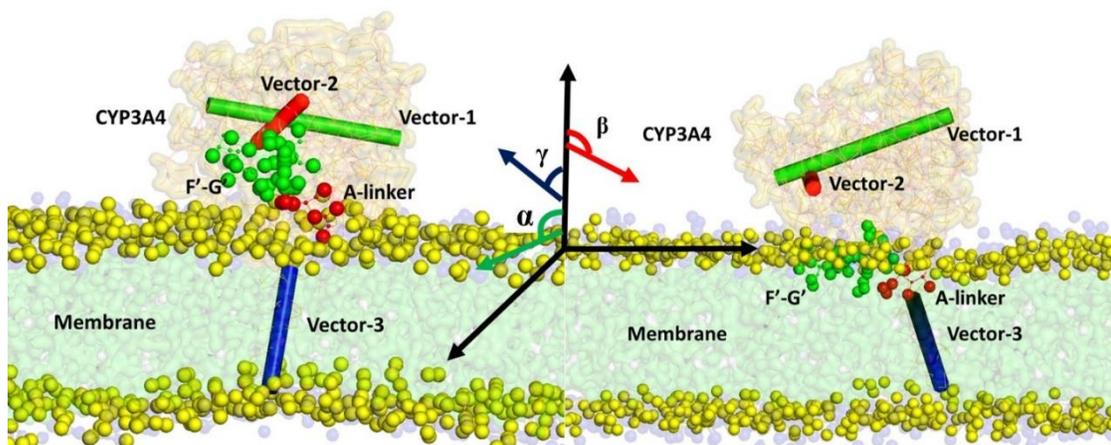


Figure 3.1: Representative initial and final configurations of CYP3A4 in a POPC bilayer. The transmembrane helix of the CYP3A4 protein is placed in the POPC membrane bilayer. A randomly generated position of the globular domain (brown surface representation) is shown in (a), where the A-linker (red spheres) and F'-G' helices (green spheres) are outside membrane. The α , β and TM tilt, γ , angles are determined by v_1 (vector-1 in green), v_2 (vector-2 in red) and v_3 (vector-3 blue), respectively (labeled as Vectors in figure). A converged orientation of the globular domain above the membrane is shown in (b) where the globular domain forms interactions with the lipid bilayer through the F'-G' helices and the A-linker regions, which are buried in the membrane.

The angle α was defined as the angle between v_1 and the z-axis and angle β was defined as the angle between v_2 and the z-axis. Similarly, the TM helix tilt angle or angle γ in the lipid membrane was defined as the angle between v_3 and the z-axis. The distances

of the CoM of the globular domain, linker region and F'-G' helices to the CoM of the lipid bilayer were monitored during the trajectories. The heme-tilt angle, defined as the angle between the heme plane (defined by the four nitrogen atoms coordinating the iron) and the z-axis, was monitored for the AAMD simulations.

3.3.2 Analysis of membrane properties

The different CG simulation methods and the presence of the protein may influence the physical properties of the membrane, such as the area per lipid (APL) and the membrane thickness. Therefore, the APL and the membrane thickness were calculated for CG and AAMD simulations. For the CG simulations, the time average APL was calculated simply by multiplying the x and y box dimensions and dividing by half the number of lipid molecules. The effect on APL and thickness due to the presence of the protein was checked after conversion of the CG model to the AA model. For the AA model, a Voronoi tessellation in combination with Monte Carlo (VTMC) integration method was used to compute the APL(139). The lipid molecules were separated into top and bottom layers and into those near the protein (boundary lipids) and those far from the protein (non-boundary lipids) (Figure 3.6). I also used the GridMAT-MD algorithm developed by the Bevan Lab for analysis of APL and thickness for protein-membrane systems (<http://www.bevanlab.biochem.vt.edu/GridMAT-MD/>) (140). MD frames were taken from the complete trajectory for 90 ns of the AAMD simulation at an interval of ~200 ps. The phosphate head group atom was used as a reference atom and 20 grid points were used. The APL and thickness of the top and bottom layer within and beyond 1.5 nm from the protein residues were calculated.

The lateral diffusion coefficient of the lipids in the CG simulations was calculated using the *g_msd* script available in the Gromacs 5.0.4 package. The jumps over the box boundaries and the overall center of mass motion were removed beforehand. A line was fitted to the linear regime of the mean-square-displacement curve to compute the lateral diffusion coefficient. This was then divided by 4 to account for the faster diffusion at the CG level than AA level due to the smoothed free energy landscape.

3.4 Results and discussion

3.4.1 Effect of different simulation parameters on performance

In the current study, two different CG water models, PW and NPW, are used for the POPC bilayer and CYP-bilayer simulations. A comparison is also made between two Gromacs versions 4.5.x vs 4.6.x which differ in the non-bonded cut-off schemes used. For example: a default cut-off scheme based on a classical buffered Verlet list is implemented in Gromacs 4.6.x and a group cut-off scheme, which is based on group of charged particles is used in Gromacs 4.5.x and below. When using the Verlet cut-off scheme, straight cut-off for VDW interactions is used together with PME and RF for Coulombic interactions. Whereas shifted potentials are used for both VDW and Coulombic interactions in Gromacs version 4.5.X and older. In another set of simulations, a comparison between Parrinello-Rahman and Berendsen pressure coupling is made. As shown in Table 3.2, the NPW model with the RF has the best computational performance, both with and without the protein, compared to the Shift method. Similarly, the PW model with the RF method showed better computational performance than with the PME method or the Shift method. It has been suggested by Gromacs developers that the Verlet scheme performs better and is more accurate than (<http://manual.gromacs.org/documentation/5.1/user-guide/cutoff-schemes.html>) group cutoff-scheme. The speed of the calculations with the PME method was sensitive to the Fourier spacing assigned. In simulations with PW and PME, changing the Fourier spacing from 0.12 to 0.33 nm increased simulation speed (for the protein system) more than 10-fold and reduced the time required for convergence of the protein orientations. However, increasing the grid spacing without increasing the cutoff affects the membrane properties as shown below (Section3.4.3). The NPW model with the Shift method (Gromacs version 4.5) was tested for other CYPs, namely CYP2C9 (31) and CYP51(112), and found to give reasonable protein-membrane interactions. Here, we observed that the computational performance for CYP3A4-membrane MD simulations with the NPW model and Shift method is approximately two times slower than with the

NPW model with the RF, as shown in Table 3.2. Overall, the best computational efficiency in terms of simulation time/CPU and the shortest simulated time for obtaining a converged protein orientation for the CYP-bilayer system (see Figure 3.2) was obtained with the new GROMACS version (5.0.4) and the NPW model with the RF treatment.

A recent study by MARTINI developers, described the testing of the MARTINI force field on different lipid systems by comparing the use of a shifted potential for both Coulomb and LJ interactions and new parameters using a straight cutoff for LJ and a RF for Coulomb interactions (86). Consistent with our study, the latter parameters were shown to give significantly improved computational performance without affecting computed system properties. However, simulation parameters were changed slightly by MARTINI developers. For example, time step was changed from 20 fs to 30fs, the neighbor list update was changed from 10 to 20 timestep, and the cutoff distance was decreased from 1.2 nm to 1.1 nm for both Coulomb and VDW interactions. These parameters were tested in a membrane only system. Therefore, we also tested sensitivity to the values assigned to these parameters for protein-membrane systems. We found that the simulations for protein-membrane system with the PW model and RF or PME crashed with the modified parameters. Therefore, we switched back to the set of parameters that we used in our previous studies (112), that is, time step 20 fs, nstlist (frequency to update the neighbor) after 10 steps and cutoff distance of 1.2 nm. From these tests, we found that calculations with the PW model and RF could also be run with the cutoff distance of 1.1 nm but a time step of 30 fs and/or a nstlist update after 20 steps resulted in simulation crashes. However, using NPW with RF and the modified parameters with a 20fs time-step, instead of 30fs, works fine with protein-membrane simulations.

3.4.2 Effect of different simulation parameters on convergence

The convergence of the CYP orientation in the membrane was determined by monitoring the angle α , the angle β , and the distances of the CoM of the globular

domain of the protein, the linker and the F'-G' region from the CoM of the lipid-bilayer in the trajectories. The convergence of the CYPs orientation in the membrane was also influenced by changing the water model and simulation parameters. For example, in the case of the PW model with the RF, insertion of CYPs in the bilayer and convergence to the final protein orientation took more than 6 μ sec in a few cases (for CYP2C9, CYP2C19 and CYP3A4). In CYP3A4 simulations using both NPW and PW water models with RF, most simulations converged quickly to a single orientation of CYP3A4 that remained stable throughout the rest of the simulations. In the first microsecond, the globular domain of CYP3A4 approached the lipid bilayer and made interactions with it. The protein-bilayer interactions were mainly formed by the linker region (residue 29 to 48) and the F'-G' helices (residue 218 to 236). The F'-G' helices and the linker residues penetrated below the lipid head groups and formed stable hydrophobic interactions with the interior lipid tail region resulting in a stable orientation of the globular domain in the bilayer (Figure 3.1). The convergence of the CYP3A4 position with respect to the membrane during PW with PME and NPW with RF simulations is illustrated in Figure 3.2a and Figure 3.2b. These plots show a transition of the globular domain towards the bilayer between about 400 and 600 ns for the PW model and in the first 200 ns for the NPW model, followed by a gradual reorientation that is completed within about 1 microsecond; thereafter the position and orientation of the protein is maintained to the end of the 4.5 microsecond simulations. The normalized distributions of the COM distance and angles α and β averaged over the converged part of the 5 PW with PME and 5 NPW with RF simulations show similar distributions (Figure 3.2c Figure 3.2d).

Table 3.2: The performance of CG simulations with two water models, three non-bonded interaction treatments and two different pressure coupling schemes is shown. The performance is compared for pure POPC and for CYP3A4-POPC bilayer systems. Two different pressure coupling schemes, Berendsen (Ber) and Parrinello-Rahman (PR), with τ_p values of 2 ps and 12 ps, respectively, are compared. ^(a) PME with Fourier spacing set to 0.12 nm whereas all other PME runs had a Fourier spacing of 0.33 nm.

Water model	NB Treatment	Pressure control	CG particles	No: of CPUs	Speed ns/day	GROMACS Version
POPC membrane only						
PW	PME	PR	3968	48	3268	5.0.4
PW	PME ^a	PR	3968	48	1614	5.0.4
PW	PME	Ber	3968	48	1448	5.0.4
PW	RF	PR	3968	48	7361	5.0.4
NPW	RF	PR	2432	48	24018	5.0.4
CYP-POPC membrane system						
PW	PME	PR	63817	192	1125	5.0.4
PW	PME ^a	PR	63817	192	76	5.0.4
PW	RF	PR	63817	192	2733	5.0.4
PW	Shift	PR	63817	192	2673	4.5.5
NPW	RF	PR	34995	192	4358	5.0.4
NPW	PME	PR	34995	192	556	5.0.4
NPW	Shift	Ber	34995	192	2012	4.5.5

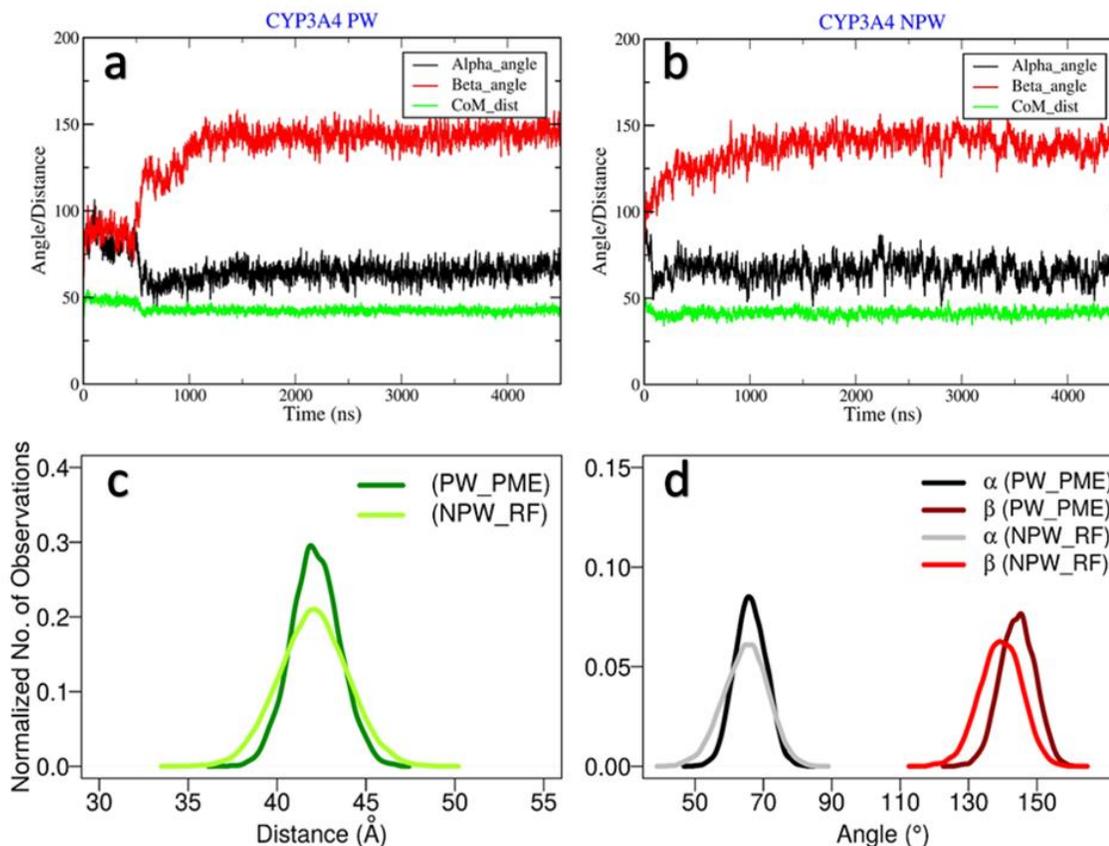


Figure 3.2: *a, b.* Convergence of the position of CYP3A4 with respect to the bilayer during a simulation with the PW model and PME (*a*) and with the NPW and RF (*b*). The distance from the center of mass (CoM) of the globular domain of the protein model, to the CoM of the lipid bilayer (green line). The angle α (black) between v_1 and the z -axis, and the angle β (red) between v_2 and the z -axis are shown. *c, d.* The normalized distribution of the distance between the CoM of the globular domain of the protein and the CoM of the lipid bilayer (*c*) and the angles α and β for snapshots at intervals of 1 ns from the converged part of the five simulations using NPW with RF and the five simulations with PW with PME (*d*).

The calculated means and standard deviations of the protein CoM to membrane CoM distance and the angles for the simulations with the two water models and different non-bonded interaction treatments are shown in Table 3.3. The converged positions of the globular domain did not differ significantly in the different CG simulations. CG simulations using PW and NPW models and different non-bonded interactions show similar converged angles and distances. The CYP3A4 globular domain formed an α angle of 64-66° and a β angle of 139-144°. The N-terminal TM-helix tilt angle γ ranged from

29-34°. The distance of the CoM of the globular domain of the protein from the CoM of the bilayer was 42-44 Å. Thus, the final orientation of CYP3A4 with respect to the bilayer and membrane APL as rather robust to the differences in the calculations although these significantly affected the computational efficiency Table 3.2

Table 3.3: The means and standard deviations for angles and distances defining the position of the CYP3A4 globular domain with respect to the bilayer were computed from averaging over snapshots obtained after convergence of the trajectories for CYP3A4.

Model	Water model	NB treatment	Epsr	Angles (°)			CoM distance (Å)
				Alpha (α)	Beta (β)	TM (γ)	
CG	PW	PME	2.5	66.2±4.6	144.3±5.0	34.4±8.3	42.1±1.4
CG	PW	RF	2.5	63.8±5.2	142.2±5.9	30.79±8.7	47.6±1.8
CG	PW	Shift	15	64.6±6.4	139.7±6.4	29.5±9.2	44.2±2.2
CG	NPW	RF	15	64.7±6.5	139.3±6.1	30.5±7.4	42.0±1.9
AA	TIP3P	PME	1	79.3±3.1	139.6±6.1	48.9±3.9	41.5±0.9

3.4.3 Effect of different simulation parameters on the bilayer properties

We tested the effect of water models and long-range interaction treatments on the POPC bilayer properties, in particular, the APL and the bilayer thickness. We evaluated CYP-membrane complexes and a pure POPC bilayer given in Table 3.4. For the pure POPC bilayer, the APL with PW and PME was a little higher at 0.63 ± 0.02 nm² than for PW with RF at 0.60 ± 0.01 nm². The experimentally determined APL for POPC is between 0.64 nm² and 0.68 nm² and the thickness is 3.7 nm (141, 142). The APL of POPC in the CYP3A4-membrane complex using the NPW model was the same with the RF and PME treatments: 0.63 ± 0.005 nm². The comparison between Shift and PME methods using the group cutoff scheme (GROMACS 4.0.5) with PW gave no significant difference in APL: 0.66 ± 0.005 nm² vs 0.68 ± 0.006 nm². For the PW model with the newer GROMACS version and the Verlet cutoff scheme, the APL decreased slightly with PME to 0.65 ± 0.008 nm² and to 0.66 ± 0.005 nm² with RF. In all CG simulations, the APL was found

in the range of the experimental APL of the POPC lipid membrane, and as reported earlier, there was no major change when treating long range interactions with RF or PME. However, when the Fourier grid spacing was increased from 0.12 to 0.33 nm in PW simulations with PME using the Verlet cutoff scheme, the performance speed increased more than 14-fold, the protein-membrane interaction and orientations were unaffected, but the APL increased from $0.65 \pm 0.008 \text{ nm}^2$ to $0.74 \pm 0.008 \text{ nm}^2$. A similarly large APL was obtained by using a dielectric constant of 15 with the PW model. These values of the APL are greater than those observed experimentally.

Table 3.4. The area per lipid (APL) computed for CG simulations of the pure POPC and CYP3A4-bilayer systems with different water models and simulation parameters. ^a Fourier grid spacing of 0.12 nm. ^b Fourier grid spacing of 0.33 nm. ^c Cutoff increased from 1.2 to 1.4 nm.

System	Water model	Coulombic interaction treatment	LJ* interaction treatment	Epsr	Cutoff scheme	Area per Lipid (APL)
POPC	PW	PME ^a	Cutoff	2.5	Verlet	0.63 ± 0.02
POPC	PW	RF	Cutoff	2.5	Verlet	0.60 ± 0.01
POPC	PW	PME ^b	Cutoff	2.5	Verlet	0.88 ± 0.04
POPC	PW	PME ^b	Cutoff ^c	2.5	Verlet	0.66 ± 0.02
POPC	NPW	RF	Cutoff	15	Verlet	0.67 ± 0.02
CYP-POPC	PW	PME ^a	Cutoff	2.5	Verlet	0.65 ± 0.008
CYP-POPC	PW	PME ^b	Cutoff	2.5	Verlet	0.75 ± 0.008
CYP-POPC	PW	RF	Cutoff	2.5	Verlet	0.66 ± 0.005
CYP-POPC	PW	Shift	Shift	2.5	Group	0.66 ± 0.006
CYP-POPC	PW	PME ^a	Shift	2.5	Group	0.68 ± 0.006
CYP-POPC	PW	Shift	Shift	15	Group	0.72 ± 0.007
CYP-POPC	NPW	RF	Cutoff	15	Verlet	0.63 ± 0.005
CYP-POPC	NPW	PME ^a	Cutoff	15	Verlet	0.63 ± 0.005

3.4.4 All-atom molecular dynamics simulation of CYP3A4

An AAMD simulation of about 90 ns was performed to study protein-bilayer interactions at the atomic detail. The heme-tilt angle, defined as the angle between the heme plane and the z-axis normal to the membrane, was previously determined to define the orientations of various CYPs with respect to the membrane and was used here to monitor the CYP orientation. The orientation of the globular domain of CYP3A4 in the membrane changed after 60 ns and remained stable up to 90 ns (Figure 3.3). Besides the TM helix (1-28), the most predominant hydrophobic interactions between CYP3A4 and the lipid tail were found for the linker region and the F'-G' and loop between G' and G helices (Figure 3.3). Compared to the CG simulations, the A'-helix (residues 32-36) reorients and buries into the lipid tail region whereas the A-linker region (42-48) remained buried in the hydrophobic tail region of the lipid bilayer which shows agreement with previous studies.(117, 134). Similar interactions of CYP3A4 in a DPPC membrane bilayer were observed by Navrátilová *et al.* (143). In AA simulations, the mean CoM distance initially increased from 42.0 ± 1.9 Å by about 2 Å to relieve atomic clashes after the back conversion to the AA representation and then decreased resulting in an average of 41.5 ± 0.9 in the last 30 ns of 90 ns simulations. The β angle remains stable during AA simulations and does not differ from CG simulations β angle while α angle increased after 60 ns from 64.7 ± 6.5 (CG) to 79.3 ± 3.1 (AA) (Table 3.3). The TM-helix remains highly flexible as seen from the B-factors and the TM tilt angle which increased from CG: $30.5 \pm 7.4^\circ$ to AA: 48.9 ± 3.9 . During AA simulation I-helix shows bending/kink in the middle which is similar to previously observed kink in TM helix (144). The heme-tilt angle initially decreased from starting value of 53.74° (after CG to AA conversion) to $52.7 \pm 3^\circ$, which increased after 60 ns to 61.0 ± 3.8 (computed for the last 30 ns simulations).

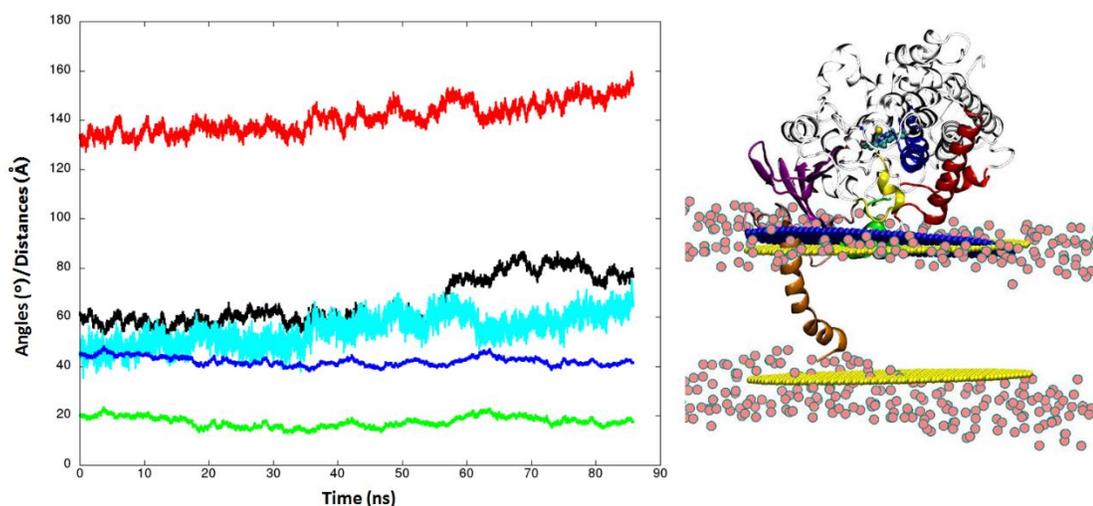


Figure 3.3: Angles and distances calculations. The angles alpha (black), beta (red) and heme-tilt (cyan) angles, calculated for complete trajectory of ~90 ns (left), are shown in thick lines. Distances to the center of mass (CoM) of membrane to the globular domain and F'-G' helices CoMs are shown in blue and green lines respectively. The orientation of CYP3A4 obtained from the last snapshot of the 90 ns AAMD simulation is shown in cartoon representation (right panel). The protein is shown in grey colored cartoon representation. TM helix and linker is shown by orange color, beta-strands in violet, BC-helices in yellow, F-G helices in red and F'-G' helices are shown in green. Central I-helix is colored blue, and heme is shown by cyan licorice representation. The lipid head groups, phosphate atoms (PO4), are show in red spheres. CYP3A4 obtained from OPM database for PDB 1TQN (blue) and PDB 5TE8 (yellow) are superimposed on last snapshot of the AAMD simulation. Sphere representation indicates the orientation CYP3A4 with respect to the hydrophobic slab.

The CYP3A4 interactions with the membrane were studied by defining residues within 5 Å of various groups in the membrane (choline, phosphate and lipid tail). The computed average B-factors (mean squared atomic positional fluctuations multiplied by $8\pi^3/3 \text{ \AA}^2$) in the 90 ns of this simulation were compared with the crystallographic B-factors and which showed lower B-factors for residues interacting with the lipid membrane (Figure 3.4).

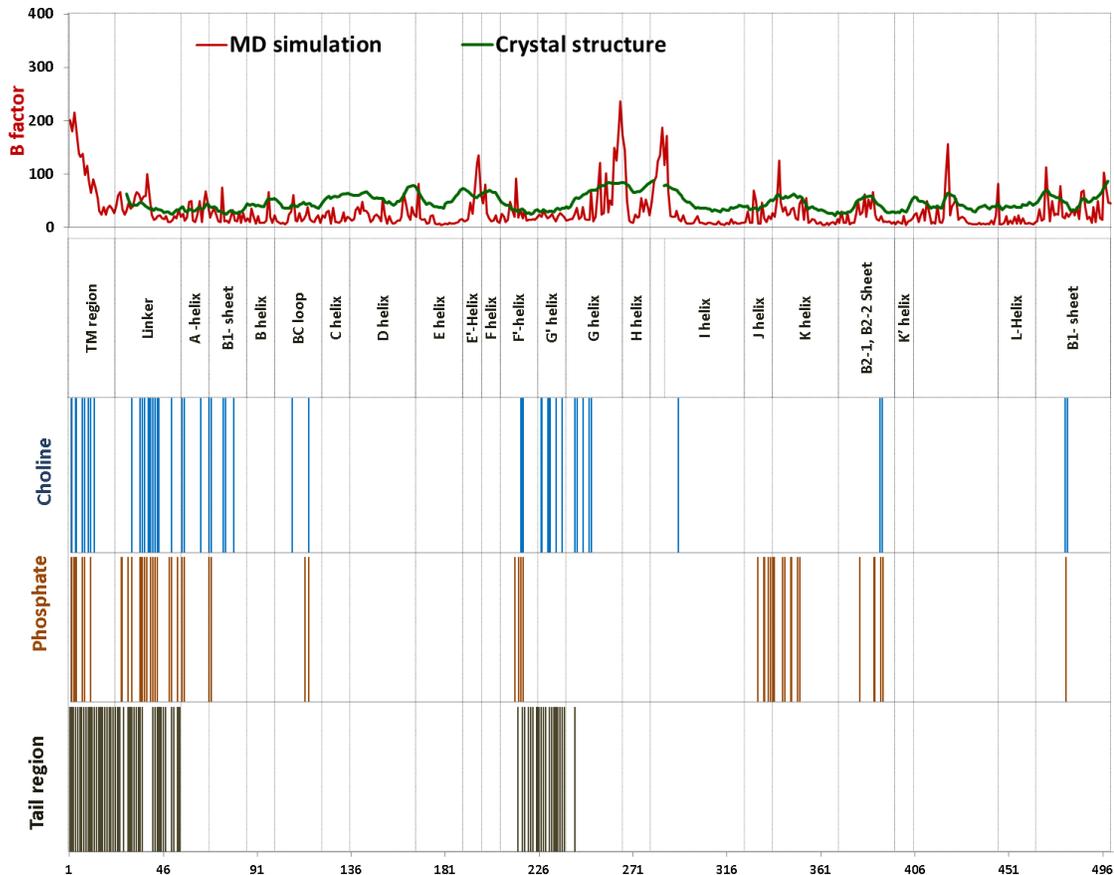


Figure 3.4: Comparison of representative B-factor values. X-ray crystal structure (1TQN) B-factor shown in thick green lines and B-factor values during simulations for Ca atoms are shown in red. The secondary structure of CYP3A4 is labelled below. The protein residues within 5 Å of the choline head, the phosphate group and the lipid tail regions are shown by lines.

The analysis of RMSD is shown in 2D-RMSD plot with respect to each frame after extracting frames at 200 ps intervals (Figure 3.5). As shown in the 2D-plot, initial RMSD increased between 1.5-2.5 Å for first ~30 ns simulations and after 30 ns RMSD is decreased between 0.5 to 2.0 Å suggesting simulations leading to equilibrated system (stable conformation) till the end of the simulation.

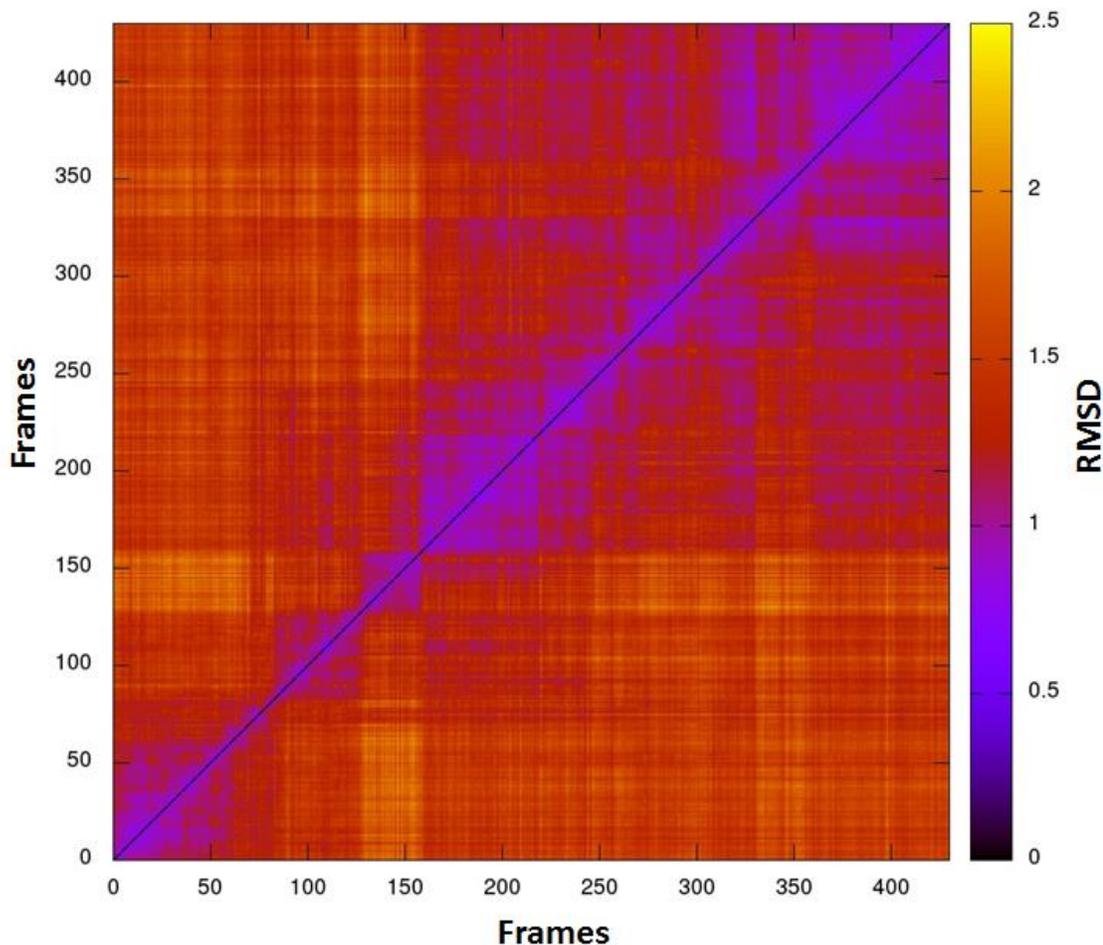


Figure 3.5: 2D-RMSD in simulations of CYP3A4. The root mean square deviation of every frame (X-axis) is compared to all other conformations (Y-axis) for backbone Ca-atom as a function of time during a 90 ns simulation. The color code is shown in the right color bar.

3.4.5 All-atom molecular dynamics simulation area per lipid

The APL values shown in Table 3.5 are average values for all lipid molecules in the top and bottom layers and do not account for the presence of the N-terminal TM helix and the globular domain of the protein. Therefore, APL and membrane thickness were checked after conversion of the representative frame from the CG NPW and RF simulation to an AA model and after running AAMD simulations. The APL was computed with the VTMC algorithm which calculates APL separately for the upper and the lower layer and for the boundary lipids and the non-boundary lipids (Figure 3.6). The APL was

also calculated for a AAMD simulations trajectory of ~90 ns by extracting coordinates frames after every 200ps.

The comparison is shown between initial and final frames in Table 3.5. The APL calculated for initial frame taken after 200 ps of AAMD trajectory for the boundary lipids was 0.47 nm² for the upper layer and 0.53 nm² for the lower layer, indicating distortion of the lipid immediately surrounding the protein. The non-boundary lipids had APL values in the upper layer of 0.64 nm² and the lower layer of 0.66 nm² consistent with experimental values for lipid alone. After running ~90 ns AAMD simulations the APL changed slightly from 0.64 nm² to 0.65 nm² for non-boundary lipids in the upper layer and from 0.66 nm² to 0.68 nm² in the lower layer. The average APL for all lipids increased from 0.636 nm² to 0.646 nm² which shows close agreement with experimental values e.g. experimental APL is between 0.64 nm² to 0.68 nm².

Table 3.5: APL computed and shown for first and last frames from AAMD trajectory analysis. The green columns show the APL for all-lipids, boundary and non-boundary lipids in the upper layer, the yellow columns show APL for all lipids, boundary and non-boundary in the bottom layer and blue columns show averages of APL for all lipids, all boundary lipid all non-boundary lipids.

CYP3A4	Time (ns)	Box Dimension (Å)		Average APL (Å)				
		X-Size	Y-Size	Boundary	Non-Boundary	Boundary	Non-Boundary	All
LIPID14	Initial	142.	138.	58	536	48.3	65.3	63.6
	90	134.	149.	66	528	47.8	66.8	64.7

*(142, 145),(136)

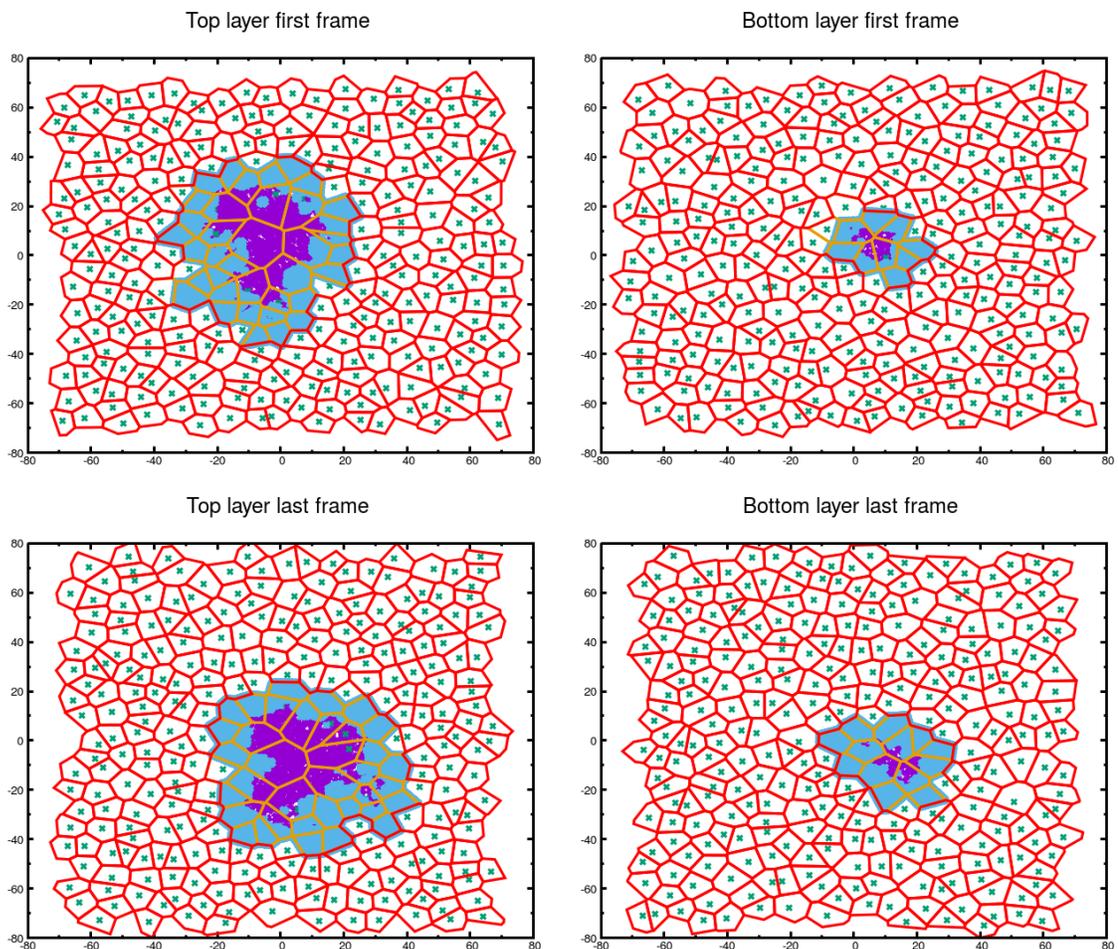


Figure 3.6: Area per lipid calculations. The assignment of lipid molecules as boundary (cyan) and non-boundary (white) lipids are represented by polygons. The violet regions show protein atoms. The + marks shows center of mass of each lipid molecule. The area per lipid is computed by Voronoi tessellation for non-boundary and Monte Carlo integration for boundary lipids.

3.4.6 Comparison of results with experimental and simulation data

We compared the results of our CG and AA simulations for the CYP3A4-bilayer orientation and interactions with experimental data and previous simulation results. The orientations of various CYPs with respect to the membrane have been experimentally determined by measurement of the heme-tilt angle and we used this angle for the comparison (Table 3.6). In all cases, the same CYP3A4 apo structure (PDB 1TQN), which is a closed structure, has been selected. All measurements and simulations were performed for a POPC bilayer except those of Navrátilová *et al* (144).

The heme-tilt angles observed at the end of our CG trajectory and in the subsequent AAMD simulation agree reasonably with the experimental data. However, there are three different X-ray structures used in the OPM database with PDB ids 1TQN, 3NXU and 5TE8. The later X-ray structure is published recently complexed with

Table 3.6. Comparison of the heme-tilt angle from experiments and simulation results for CYP3A4.

<i>Source</i>	<i>Reference</i>	<i>Heme tilt angle (°)</i>
<i>CYP3A4 in nanodiscs</i>	Baylon <i>et al</i> (117)	59.7±4
<i>MD simulation</i>	Baylon <i>et al</i> (117)	72.6±5
<i>MD simulation</i>	Navrátilová <i>et al.</i> (143)	52 (pure DPPC) 68 (DPPC with 50 wt% chol)
<i>MD simulation</i>	Lonsdale <i>et al.</i> (134)	72±14
<i>OPM (1TQN)</i>	(121)	67±11
<i>OPM (3NXU)</i>	(121)	52±12
<i>OPM (5TE8)</i>	(121)	62 ± 5*
<i>CGMD simulation</i>	this work	53.7
<i>AAMD simulation</i>	this work	61.0±3.8

* New crystal structure of CYP3A4 bound with midazolam released on 7th Dec 2016 and paper published on Jan 2017.

Recently, Sligar and colleagues measured the heme-tilt angle for CYP3A4 in a POPC nanodisc as 59.7±4°, which is in close agreement with the heme-tilt angle of 53.7° and 61.0±3.8° for the representative structure from our CG simulations and the last 30 ns of the AAMD simulation, respectively. Navrátilová *et al.*(143) found that the orientations and insertion depth in the membrane of CYP3A4 were influenced by the addition of cholesterol to a DPPC bilayer with the heme-tilt angle changing from 52° for pure DPPC to 68° with 50 wt% cholesterol. Two different orientations of CYP3A4 are reported in the OPM database, where CYP3A4 (PDB id: 3NXU) has different orientation in the hydrophobic slab membrane with heme-tilt angle 52±12° than the structure with PDB id 1TQN which has higher heme-tilt angle (67±11°). Former is structure of CYP3A4

is complexed with ritonavir inhibitor, and may influence on the dynamics of protein(146). Recently, new structure of 3A4 bound with midazolam (PDB id 5TE8)(147) has been resolved and orientation suggested by OPM database for 3A4 PDB id 5TE8 ($62\pm 5^\circ$) and 1TQN structure ($67\pm 11^\circ$) is similar to our simulations results (61.0 ± 3.8). Superimposition of our structure with coordinates obtained from the OPM database also showed similar orientation and position with respect to the membrane to the 1TQN and 5TE8 structures (Figure 3.3), which is in agreement with the heme-tilt angles obtained after AAMD simulations.

3.5 Conclusion

Our results show that a combination of simulations with the MARTINI CG model followed by AAMD simulations can be successfully used to immerse CYP proteins in a POPC bilayer. We expect this approach to be applicable to other membrane proteins with monotopic domains and to bilayers with different phospholipid content. The CG simulations are used to efficiently sample the configurational space of the system, reproduction of the dynamics of the system is not required in these CG simulations and was not evaluated here. Our tests of different water models, long-range interaction treatments and implementations for the CG simulations, showed that the computation times varied widely for different model and parameter choices. On the other hand, the final position of the immersed CYP protein with respect to the membrane and the membrane structural properties were similar for most model and parameter choices and these were consistent with available experimental data. We conclude that for a computationally efficient multiscale procedure to immerse a protein in a membrane, the CG simulations should be carried out with the MARTINI force field version 2.2 with the standard NPW water model and a RF treatment of long-range Coulombic interactions.

CHAPTER 4: UNDERSTANDING THE MEMBRANE PROTEIN

INTERACTIONS OF TWO ISOFORMS OF THE CYP2C SUBFAMILY: CYP2C9 AND CYP2C19

4.1 Introduction

Among different CYP isozymes, human subfamily CYP2C contributes significantly in the hepatic clearance of several drugs. Although members of the subfamily exhibit >82% sequence identity, they have unique substrate specificity. The subfamily of CYP2C consists of 4 isoforms of CYPs i.e., CYP2C9, CYP2C19, CYP2C8, and CYP2C18. CYP2C9 is the most expressed CYP after CYP3A4, and is responsible for the metabolism of >12.8% of drugs that are mainly weak acids like non-steroidal drugs (NSAID). CYP2C19's contribution to drug metabolism is 6.8% with a 10-fold lower expression level than CYP2C9 (16). The expression of CYP2C19 is increased in response to the treatment of *Helicobacter pylori* (*H. pylori*) with proton-pump inhibitors such as omeprazole, lansoprazole, and pantoprazole, thereby increasing drug clearance. The polymorphism of CYP2C19 dramatically influences the treatment of *H.pylori* (148).

Although CYP2C19 and CYP2C9 have high sequence conservation with 94% sequence identity (see sequence alignment (Figure 4.11)), they have distinct substrate specificities. The crystal structure of CYP2C19 resolved at 2.8 Å (PDB 4GQS)(23) differs in two residues (V208 and I362) in the binding pockets from CYP2C9 structure (L208 and L362), complexed with flurbiprofen (PDB 1R9O)(149). Other differences are seen on the outer surface of the protein (23). The CYP2C19 crystal structure has a 3D fold closer to the CYP2C8 structure (PDB 2NNI) than the structure of CYP2C9 and CYP2C9m7 (PDBs 1R9O, 10G2). The latter structure was resolved after making 7 substitutions (K206E, I215V, C216Y, S220P, P221A, I222L and I223L) in the F-G loop region. The CYP2C19 structure shows more than 3.0 Å *Ca* deviation from CYP2C9 (1R9O) and CYP2C9m7 (10G2) on the outer surface of the protein responsible for substrate access and selectivity. The main differences are observed in helices F, F'-G', G and their turns, the turn in the beta-

strand1 and B-C loop regions. CYP2C9m7 (1OG2) structure differs from CYP2C9 (1R9O) structure in BC loop and F'-G' helices conformations and position of Arg108 which is pointing out of the binding cavity in the 1OG2 (19) and inside in the 1R9O. 1R9O has a highly flexible BC loop conformation and missing F'-G' loop/helices information. It has been suggested that the sequence difference and conformational changes outside the binding cavity are responsible for the differential selection of drugs to enter the binding pocket (23). These differences are mainly found in the substrate recognition regions identified by Zawaira et al (29). A description of SRS regions is given in chapter 1.

Since most residues that differ between CYP2C9 and CYP2C19 are found in these SRS regions, I proposed that the sequence differences in the SRS regions and, thereby, the conformational differences observed in the two CYPs can contribute to the different membrane-protein interactions and orientations which lead to differences in the substrate access tunnels to the binding cavity and the product release tunnels. To investigate the effect of sequence differences between CYP2C9 and CYP2C19 on the membrane-protein interactions and orientations, I used our optimized multiscale modeling and simulation protocol (chapter 3) and compared our results with previously reported experimental and computational data.

4.2 Material and method

4.2.1 Preparation of full length CYP2C9 and CYP2C19

The crystal structure of CYP2C9 (PDB 1R9O)(149), resolved at 2.0 Å resolution in complex with flurbiprofen (FLP), was taken from the Protein Data Bank <http://www.rcsb.org>. The structure was solved after removal of the N-terminal residues 1-25. The 1R9O structure has missing residues in the linker region (residues 38-42) and in the F'-G' region (residues 214-220). Recently, a crystal structure of CYP2C19 (PDB id 4GQS) was resolved at 2.87 Å, after truncating residues 1-28 from the N-terminus, in complex with the inhibitor (2-methyl-1-benzofuran-3-yl)-(4-hydroxy-3,5-dimethylphenyl)methanone (Protein Data Bank chemical component 0XV)(23). The

sequence alignment of CYP2C9 with CYP2C19 shows 92% sequence identity (Figure 4.11). The missing residues in the globular domain of CYP2C9 are identical to those in CYP2C19. Therefore, CYP2C19 was used as a template for modeling the missing linker residues and F'-G' residues in CYP2C9. The TM-helix (residues 3-21) and missing linker (residues 22-25) were modeled similarly as in Cojocar *et al* (31). The TM-helix for CYP2C19 were predicted from residues 4-20 and 3-22 by online server Psipred using membrane protein structure and topology (MEMSAT3) software and transmembrane protein topology prediction using support vector machines (SVM-MEMSAT) software (150). PredictProtein server (<https://www.predictprotein.org/>) was also used which suggested N-terminal alpha-helical conformation starting from residues 2-23. I modeled two different TM helix lengths of CYP2C19: one with residues 3-21 (same as CYP2C9 TM helix length) and second with residues 3-24. Final models consisted of X-ray structure of globular domain and modeled missing regions (mentioned as systems (S)). Various starting orientations of globular domains above the membrane were generated by changing the dihedral angles in the linker regions before converting to CG systems.

4.2.2 Preparation of additional models of CYP2C9:

Additional CG systems of CYP2C9 were prepared using modelled structures. Since the x-ray structure is missing the F'-G' helices (loop) in CYP2C9 (PDB 1R9O), different modeling approaches were used to prepare a full length CYP2C9 structure. The CG systems prepared from fully modeled structures of CYP2C9 are named by letter M for models, which differ from the CG systems (S), for which X-ray structures were used after modelling missing regions (Table 4.1). Different modeling procedures for CYP2C9 were used, brief descriptions of the resultant models are given below:

CYP2C9 model 1: CYP2C9 model was generated using CYP2C9 X-ray structure (PDB 1R9O) as template. For modeling missing residues in the linker region (residues 38-42) and in the F'-G' region (residues 214-220), CYP2C19 was used as template only for these missing regions. The CG system consisting of CYP2C9 model 1 is referred as M1.

CYP2C9 model 2: Another model of CYP2C9 was generated using complete 1R9O and 4GQS X-ray structures as templates, which generated intermediate side chain conformations between CYP2C9 and CYP2C19 structures. The CG system using model2 is referred as M2.

CYP2C9 model 3: Modeling of CYP2C9 was also performed by using only single template of CYP2C19 X-ray structure (PDB 4GQS). This model of CYP2C9 resembles CYP2C19 in side chain conformations and CG system is called M3.

CYP2C9 model 4: Modelled structure of CYP2C9 (residues 26-490) with F'-G' helices was taken from previous studies by Cojocaru *et al* (31), detail of modeling procedure of CG system is given in the paper. The missing TM helix was modeled as discussed earlier. This modeled is referred as M4.

4.2.3 Mutants of CYP2C9 and CYP2C19

The residues different at the membrane-protein interface were substituted to create chimera structures of CYP2C9 and CYP2C19. The substituted residues between CYP2C9 and 2C19 were in the linker (G46D), beta-sheet1 (K72E and P73R), B-C loop (I99H) and F'-G' helices (S220P and P221T). The five different orientations of wild-type all-atom models (S1) were selected to make substitution mutations, thereby keeping the initial orientations of the globular domain of the mutant and wild-type structures the same.

4.2.4 Preparation of coarse-grained systems

A similar procedure was used to generate coarse-grained protein-membrane systems for CYP2C9 and CYP2C19 as given in chapter 3. The effect of different linker flexibility on the final orientations of CYP2C9 and CYP2C19 was checked by defining two different linker flexible regions, residues 22-36 (CG system 1 (S1)) and residues 26-38 (CG system 2 (S2)). The linker was kept flexible by removing the restraints from specified residues, applied by elastic network (Chapter 3). CG system consisting of globular domain only (S3), residues 47-490, was prepared for CYP2C9 and CYP2C19 to allow an unbiased

conformational search of the protein orientation and to evaluate convergence of the orientations in the membrane. CG systems (S1-S3) were solvated using the MARTINI standard water model (NPW). However, the MARTINI polarizable water (PW) model was also tested (S4-S5) (Table 4.1).

After preparation of CG models, different simulation parameters were tested as given in the following section:

4.2.5 Coarse-grained simulations with different parameters:

In the current study, the MARTINI CG force-field was used to study the orientation of two closely related isoforms, CYP2C9 and CYP2C19. The same CG simulation protocol was used as mentioned in chapter 3. Here, I extended our optimized protocol to CYP2C9 and CYP2C19 to further validate it and tested different parameters to observe the effect of these parameters on the final orientations of the two isoforms of CYPs in the membrane. I used both NPW and PW water models and different treatments of non-bonded interactions; use of Shift, PME or RF (reaction field) for the calculation of Coulombic interactions and Shift or cut-off methods for calculation of VDW interactions. With NPW model, long range electrostatic interactions were calculated using RF and VDW interactions by cut-off method (S1-S3). In the case of the PW model, electrostatic and VDW interactions were calculated by Shift method (Gromacs 4.5.5) (S4), PME and cut-off (S5) or RF and cut-off (S6) methods using Gromacs 5.0.4. All CG simulations for CYP2C9 modelled system (M1-M4) and mutant CYP2C9 and CYP2C19 were performed using NPW water model and RF and cut-off for calculating electrostatic and VDW interactions, respectively. The simulation procedure was kept the same as mentioned in chapter 3.

Table 4.1: Coarse-grained simulation protocols used for CYP2C9 and CYP2C19.

CYP System	Residues in regions					No. of Simulations	Water Model	Non-bonded treatment	GROMACS version
	PDB ID	TM Helix	Linker	Flexible linker*	Globular domain				
CYP2C9 globular domain from X-ray structure									
S1	1R9O	3-21	22-49	22-36	50-490	10	NPW	RF	5.0.4
S2	1R9O	3-21	22-49	26-38	50-490	5	NPW	RF	5.0.4
S3	1R9O	Globular domain only				1	NPW	RF	5.0.4
S4	1R9O	3-21	22-49	22-36	50-490	10	PW	Shift	4.5.5
S5	1R9O	3-21	22-49	22-36	50-490	5	PW	PME	5.0.4
S6	1R9O	3-21	22-49	22-36	50-490	5	PW	RF	5.0.4
CYP2C9 modeled globular domain									
M1	Model1	3-21	22-49	22-36	50-490	6	NPW	RF	5.0.4
M2	Model2	3-21	22-49	22-36	50-490	6	NPW	RF	5.0.4
M3	Model3	3-21	22-49	22-36	50-490	5	NPW	RF	5.0.4
M4	Model4	3-21	22-49	22-36	50-490	10	NPW	RF	5.0.4
mtCYP2C9	Mutant	3-21	22-49	22-36	50-490	5	NPW	RF	5.0.4

CYP2C19									
S1	4GQS	3-24	25-49	26-38	50-490	10	NPW	RF	5.0.4
S2	4GQS	3-21	22-49	22-36	50-490	5	NPW	RF	5.0.4
S3	4GQS	Globular domain only				1	NPW	RF	5.0.4
S4	4GQS	3-24	25-49	26-38	50-490	5	PW	Shift	4.5.5
S5	4GQS	3-24	25-49	26-38	50-490	5	PW	RF	5.0.4
CYP2C19 modeled structure									
mtCYP2C19	Mutant	3-21	22-49	22-36	50-490	6	NPW	RF	5.0.4

*Linker region which was set flexible by removing elastic constraints from residues mentioned. #Mutant structures of CYP2C9 and CYP2C19 prepared after substitution of linker, beta-strand1, BC and F'-G' loop's residues.

4.2.6 Convergence of coarse-grained simulation and back conversion

The simulations were considered converged when no changes were observed in the orientations of CYPs above the membrane. Experimentally, the orientation of the CYPs above the membrane is determined by measuring the heme-tilt angle. Due to absence of heme in CG models of CYPs, the orientations were quantified by calculating the angles and distances mentioned in our previous work (31, 111, 112). Angles alpha and beta define the orientation of the globular domain above the lipid-membrane. The interactions of TM-helix in the membrane were quantified by measuring the TM-tilt angle. Various distance criteria were set as indicators of convergence of the positions of the CYPs in the membrane and insertion depth, for example, the distance of the center of mass (CoM) of the globular domain (residue 50-490), the distance of the linker (residues 22-49), the distance of the F'-G' helices (residues 210-220) from the CoM of the POPC membrane.

Based on the convergence criterion, the representative frames were selected for back-conversion to an all-atom model. The representative frame was chosen to have the angle and distance values within 1% of their mean value.(112). The back-conversion of POPC bilayer was performed as described in Cojocaru *et al.*(31), whereas the protein back-conversion was done using scripts backward.py and initram.sh, available at the MARTINI website (<http://cgmartini.nl>) (110). In the absence of the heme cofactor in the CG model, conformational changes in the side chains of heme-binding pocket residues were observed. Therefore, globular domain (residues 50-490) from crystal structure was superimposed on the back-mapped structure and used in subsequent AA simulations. The AA model of the globular domain contained the heme-cofactor. If there was a co-crystallized ligand in the crystal structure, it was also incorporated in the model. The TM-helix and flexible linker region obtained from the back-conversion procedure were then connected to the globular domain, resulting in a full length all-atom model. Finally, the all-atom model of CYP was placed into the all atom model of the POPC bilayer to obtain a complete CYP-membrane complex.

4.2.7 All-atom molecular dynamics simulations of CYP2C9 AND CYP2C19

AAMD simulations were performed with two different starting orientations of CYPs in the membrane for each CYP. Different orientations were obtained from two different CG simulation systems S1 and S2. AA force-fields ff12SB and LIPID14 were used for protein residues (135) and for POPC lipids, respectively,(136). Additionally, the GAFF lipid force-field and ff99SB for protein were also used for test purposes. The LIPID14 simulation results were compared with our previous simulations of CYP2C9 in a membrane (31), and the current GAFF lipid simulations. The GAFF lipid force-field requires surface tension to maintain the structural properties of the membrane bilayer, while LIPID14 parameters are optimized for use without application of surface tension. The heme parameters were provided by D. Harris with the partial atomic charges derived from DFT calculations.(137) The ionic concentration was maintained at 150 mM using Na⁺ and Cl⁻ ions in a periodic box of TIP3P water molecules. Same procedure for AAMD simulation was used as described in detail by Cojocaru *et al.*(31). The detailed procedures of minimization, equilibration and production simulations are given in Chapter 3. However, with LIPID14 force-field, I have tested different pressure coupling mechanisms: Anisotropic (cell fluctuates independently in x, y and z cell dimensions), Constant Ratio (constant shape in x and y dimension) and Constant Area (keeping x and y dimension constant) <http://www.ks.uiuc.edu/Research/namd/2.9/ug/node37.html>. The suggested pressure coupling method by LIPID14 force-field developers is anisotropic method, which has been tested with pure membrane system without protein. Use of anisotropic pressure coupling in protein-membrane system on longer simulations (> 100 ns) resulted in unusual x and y box dimensions. Therefore, above mentioned pressure coupling options given in NAMD software package were tested. With the GAFF lipid, surface tension coupling was used. AAMD simulations performed for CYP2C9 and 2C19 are given in following table 4.2:

Table 4.2: AAMD Simulations performed for CYP2C9 and CYP2C19. Different starting structures, parameters and force-fields were used.

Simulation	AA System	CG System	Force-field	Ensemble	Pressure coupling	Time (ns)
CYP2C9						
SIM1	Apo	S1	LIPID14	NPT	Anisotropic	193
SIM2	Apo	S1	LIPID14	NPT	Constant Ratio	97
SIM3	Apo	S1	LIPID14	NPAT	Constant Area	120
SIM4	Ligand*	S1	LIPID14	NPT	Anisotropic	156
SIM5	Apo	S1	GAFF	NPγT	Surface-tension	80
SIM6	Apo	S2	LIPID14	NPT	Anisotropic	50
SIM7	Apo	S2	LIPID14	NPT	Constant Area	170
SIM8	Apo	S2	LIPID14	NPT	Constant Ratio	105
CYP2C19						
SIM1	Apo	S1	LIPID14	NPT	Anisotropic	39
SIM2	Apo	S1	LIPID14	NPAT	ConstantArea	120
SIM3	Ligand*	S1	LIPID14	NPT	Ansiotropic	46
SIM4	Apo	S1	GAFF	NPgT	Surface Tension	30
SIM5	Apo	S2	LIPID14	NPT	Anisotropic	95
SIM6	Apo	S2	LIPID14	NPT	Constant Ratio	56

*Ligand: Flurbiprofen bound in X-ray structure of in CYP2C9 and protein data bank chemical component OXV inhibitor in CYP2C19

4.3 Coarse-grained simulation results

4.3.1 Distinct orientations adopted by CYP2C9 and CYP2C19:

The CG simulation trajectories were analyzed for two isoforms (CYP2C9 and CYP2C19) to assess convergence of their orientation and interactions with the membrane. Different starting structures, flexible linker regions (residues 22-36 (S1) and residues 26-38 (S2)), and simulation methods were tested to obtain reliable CYP-membrane interactions and CYP orientations.

In this study, both NPW and PW solvent models and different treatments of non-bonded interactions were used. The final orientations of all sets of simulations (CG systems) are given in Table 4.3. The two isoforms consistently show different orientations from each other in the membrane regardless of simulation parameters used. For the full protein, the CYP2C9 angles range from 87 to 91° and 109 to 118° for alpha and beta, respectively. The CYP2C19 alpha and beta angles range from 89 to 100° and 123 to 144°, respectively. The distance of the center of mass (CoM) of the protein (globular domain residues 50-490) from the membrane CoM is 43 Å in CYP2C9 and 45-50 Å in CYP2C19. In CG simulations of CYP2C9 with PW (S5 and S6) using PME or RF, none of the orientations converged to a stable orientation. The globular domain did not approach the membrane surface even after running 6 μs long simulations. PW with PME resulted in too slow simulations (see Chapter 3 Table 3.2), giving only a few hundred ns per day, and did not show convergence. Therefore, PME was not used for CYP2C19 simulations. On the other hand, 3 out of 5 CG simulations of CYP2C19 with PW and RF (S5) converged to a final orientation. The angles formed by S5 simulations are the same as observed in S1-S2 simulations. However, the mean CoM distance in S5 simulation is higher (50±2 Å) than the S1-S2 simulations (48±2 Å). The different orientations of the two isoforms were classified into three different classes, A, A/B and B. The CYP with beta angle below 125° were categorized in class A, from 125° to 130° into class A/B, and above 130° into class B. All CYP2C9 CG systems (S1-S4) converged to class A and the

CYP2C19 CG systems S1-S3 and S5 converged to class B. CYP2C19 S4 simulations converged into class A/B (Table 4.3).

The angles and distance values are plotted in Figure 4.1, where left panel shows comparison between CYP2C9 and CYP2C19 orientations observed in CG systems (S1-S5). The plot in the right panel (Figure 4.1) shows CG system S1 which comprises of 10 CG simulations with different initial orientations but the same simulations protocol i.e., NPW and RF. In CYP2C9 S1 simulations three out of 10 simulations gave beta angle values between 115-118° (class A), in 8 simulations beta values are below 115° (class A). Similarly, in CYP2C19 (S1), beta values in 2 simulations out of 10 lie in the average 120° (class A) and 130° (class A/B), the others are above 140° (B).

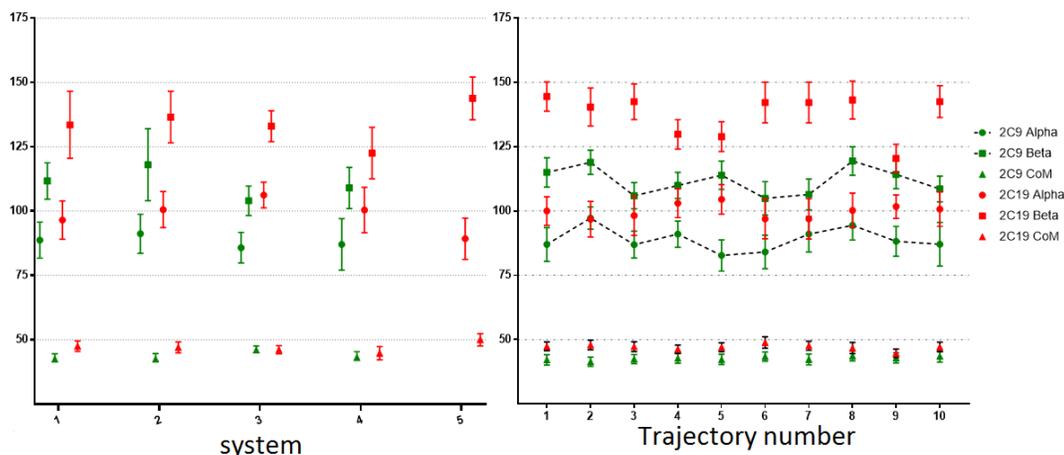


Figure 4.1: CG simulation results. Comparison between CYP2C9 (green) and CYP2C19 (red). CG simulation results of different systems are shown in the plot on left panel. Results for 10 trajectories from CG simulation systems 1 (S1) of CYP2C9 (green) and CYP2C19 (red) are shown in the right panel. Angles (°) (alpha and beta) and CoM distance (Å) are given by squares, circles and triangles, respectively.

Table 4.3: CG simulation results for CYP2C9 and CYP2C19. Different simulation parameters are shown. Mean and standard deviation are shown for angles (alpha and beta) and distance of the center of mass of proteins globular domain (CoM) from the lipid center.

CG Systems	CG Simulation Parameters					Angles (°)		Distance (Å)	Class
	TM-Helix	Flexible Linker	Water Model	NB [#] Interactions	No of Simulations	Alpha	Beta	Protein CoM	
CYP2C9									
S1	3-21	22-36	NPW	RF	10	89±7	111.7±7	43±2	A
S2	3-21	26-38	NPW	RF	5	91±8	118±14	43±2	A
S3	Glob-domain ^a		NPW	RF	1	86±6	104±6	46±2	A
S4	3-21	22-36	PW	Shift ^c	10	87±10	109±8	43±2	A
S5	3-21	22-36	PW	PME	5	---	---	---	--
S6	3-21	22-36	PW	RF	5	---	---	---	--
mtCYP2C9 ^b	3-21	22-36	NPW	RF	5	91±10	121±11	43±2	A
CYP2C19									
S1	3-24	26-38	NPW	RF	10	100±7	137±10	47±2	B
S2	3-21	22-36	NPW	RF	9	97±8	134±13	48±2	B
S3	Glob-domain ^a		NPW	RF	1	106±5	133±6	46±2	B
S4	3-24	26-38	PW	Shift ^c	5	100±9	123±10	45±3	A/B
S5	3-21	26-38	PW	RF	3*	89±8	144±8	50±3	B
mtCYP2C19 ^b	3-24	26-38	NPW	RF	5	95±8	128±13	47±2	B

^a globular domain only system (CYP2C9 residues 47-490 and CYP2C19 residues 47-490), ^bMutant CYPs, ^cthe value of epsilon_r=15 used in PW model with Shift method, [#]Non-bonded interactions, *values for 3 out of 5 simulations showed convergence

4.3.2 Full length vs globular domain simulation results

In order to ensure initial orientations do not bias the results, two separate simulations of CYP2C9 and CYP2C19 were performed using X-ray crystal structures which consist of the only globular domain (S3). The globular-domain-only-CYPs could explore various configurations before reaching a stable orientation. In both isoforms, the final orientation of the CYPs were stabilized by insertion of the F'-G' helices in the bilayer (Figure 4.2). Snapshots from CG simulations (S3) are shown at different time intervals. The F'-G' helices region is one of the hydrophobic regions in CYPs which keep globular domain anchored in the membrane even after truncating the TM-helix. In CYP2C9 globular domain simulations, the orientation converged in 3.5 microseconds with sharp decrease in the CoM distance of F'-G' helices, shown by the arrow in Figure 4.3 B. CYP2C19 (S3) converged quickly in 200 ns and remained stable throughout simulations (Figure 4.3 D). After the F'-G' helices in both CYPs developed contacts with the membrane, there was no further change observed in the orientation of the CYPs. The angles formed by the full length CYP2C9 CG simulations (S1) are same as CYP2C9 globular domain CG system (S3) (Table 4.3). The full length CYP2C19 (S1 and S2) and CYP2C19 globular domain (S3) systems also converged to similar orientations. The difference in the orientations of two CYPs in membrane is maintained in both full length (S1) and globular domain simulations (S3). However, in the globular domain simulations (S3), the CoM distance does not vary (46 ± 2 Å). Overall, the CYP2C9 CG simulations (Table 4.3, Figure 4.1 and Figure 4.2) show difference in the final orientations from CYP2C19 despite using the same simulations parameters, water models and protein components (Globular or full length).

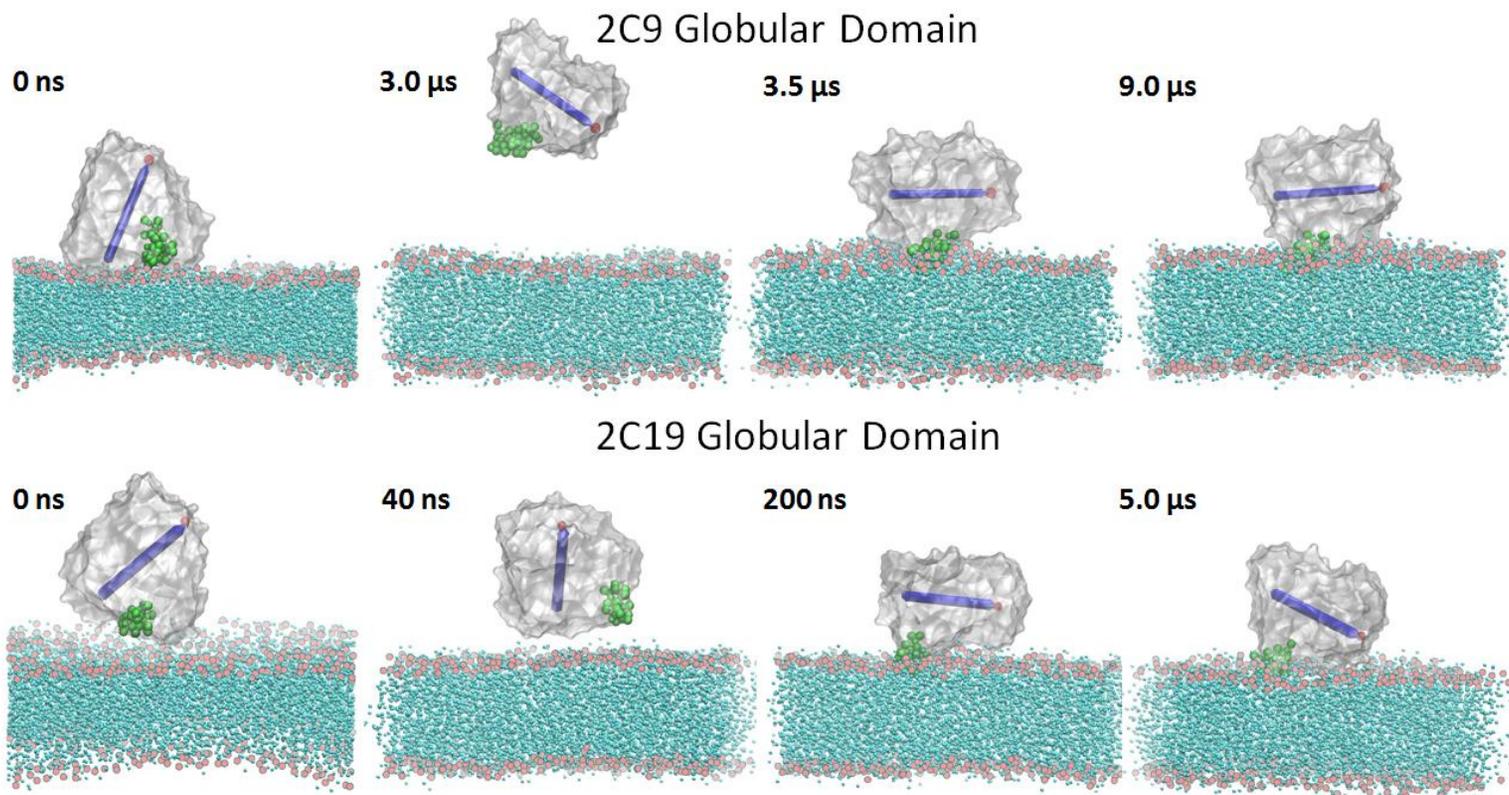


Figure 4.2: Snapshots from CG simulations of globular domains (S3) of CYP2C9 and CYP2C19. The globular domain is shown as silver surface representation, where F'-G' region is shown in VDW spheres colored in green, I-Helix is shown in blue cylinder with arrow at the end (residue 286-316 in full protein). Different orientations of globular domain are shown at different time steps. The POPC membrane is shown in cyan color and phosphate atoms (membrane head group) by red spheres.

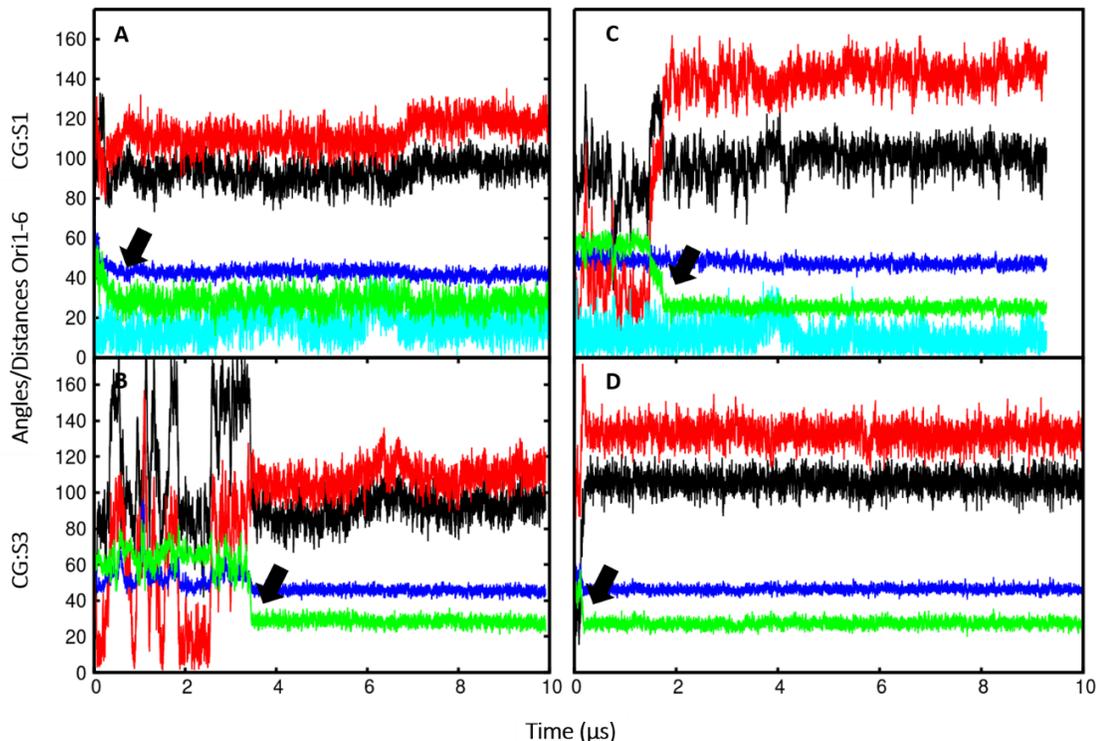


Figure 4.3: Angles and distance graphs of CYP2C9 and CYP2C19. The angles ($^{\circ}$) and distance (\AA) values vs time (μs) are shown for selected trajectories from CG systems (S1 and S3). A) Full length CYP2C9 CG system (S1) B) globular domain of CYP2C9 (S3), C) full length CYP2C19 CG system (S1), D) globular domain of CYP2C19 (S3). The angles are colored as: alpha (black), beta (red), and TM (cyan). The distances; CoM of the globular domain (blue) and the CoM of the F'-G' helices shown in green color. The thick black arrows point to the decreased CoM distance of F'-G' helices leading to stable orientations.

4.3.3 Orientations of CYP2C9 modeled structures:

Furthermore, I tested different models of CYP2C9 (M1-M4) simulated using NPW and RF method. The full-length models of CYP2C9 differ slightly in the side chain conformations in the globular domain, due to different modeling approach (templates) used (see method section). CYP2C9 is more extensively studied here due to the missing F'-G' helices in its crystal structure, and therefore, when different templates are used (see method section) different side chain conformations are expected in different models. Important differences are observed in the beta-strand1, BC and F'-G' helices conformations, which are critical for developing interactions between the CYPs and the

membrane, which determine the final orientations of the CYPs in the membrane. In Table 4.4, I compare the results of the CG simulations using different models of CYP2C9.

Table 4.4: CG simulation results of CYP2C9 models to show dependence of final orientation in the membrane on the initial conformation of protein. Angles, alpha and beta, and distances (mean and standard deviation) values are compared between different models.

CYP2C9 Models	Angles (°)		Distance (Å)			No of simul ations	Class
	Alpha	Beta	Linker	F'-G'	Globular		
M1	94±6	119±8	20±2	25±2	46±2	6	A
M2	92±8	120±12	19±2	25±2	46±2	6	A
M3	95±6	138±6	22±2	24±2	48±2	6	B
M4	85±9	106±9	25±2	26±3	47±2	5	A
mtCYP2C9	98±7	130±10	27±4	29±4	44±2	5	B
mtCYP2C19	95±8	127±13	20±2	25±3	46±2	5	A/B

In all CG simulations (M1-M4), CoM distance of the globular domain to the CoM distance of the lipid bilayer was increased from 43±2 to 47±2 Å. The new CoM distance value for CYP2C9 is the same as observed for CYP2C19. The angles (alpha and beta) increased in CG simulations using modeled structures compared to the x-ray structure of CYP2C9. In M3 CG simulations, using the 2C19 crystal structure as the template, 50% of the simulations (3 out of 6 trajectories) show higher angles (class B), resembling the orientation of CYP2C19 in the membrane. The increased angle and distance values can be attributed to the initial conformations of the globular domain, based on the selection of the template. From CG simulations of CYP2C9 models (M1-M4), it was found that not only primary sequence but also conformational differences in the linker, BC loop and F'-G' helices regions contribute to the final orientation of CYPs in the membrane.

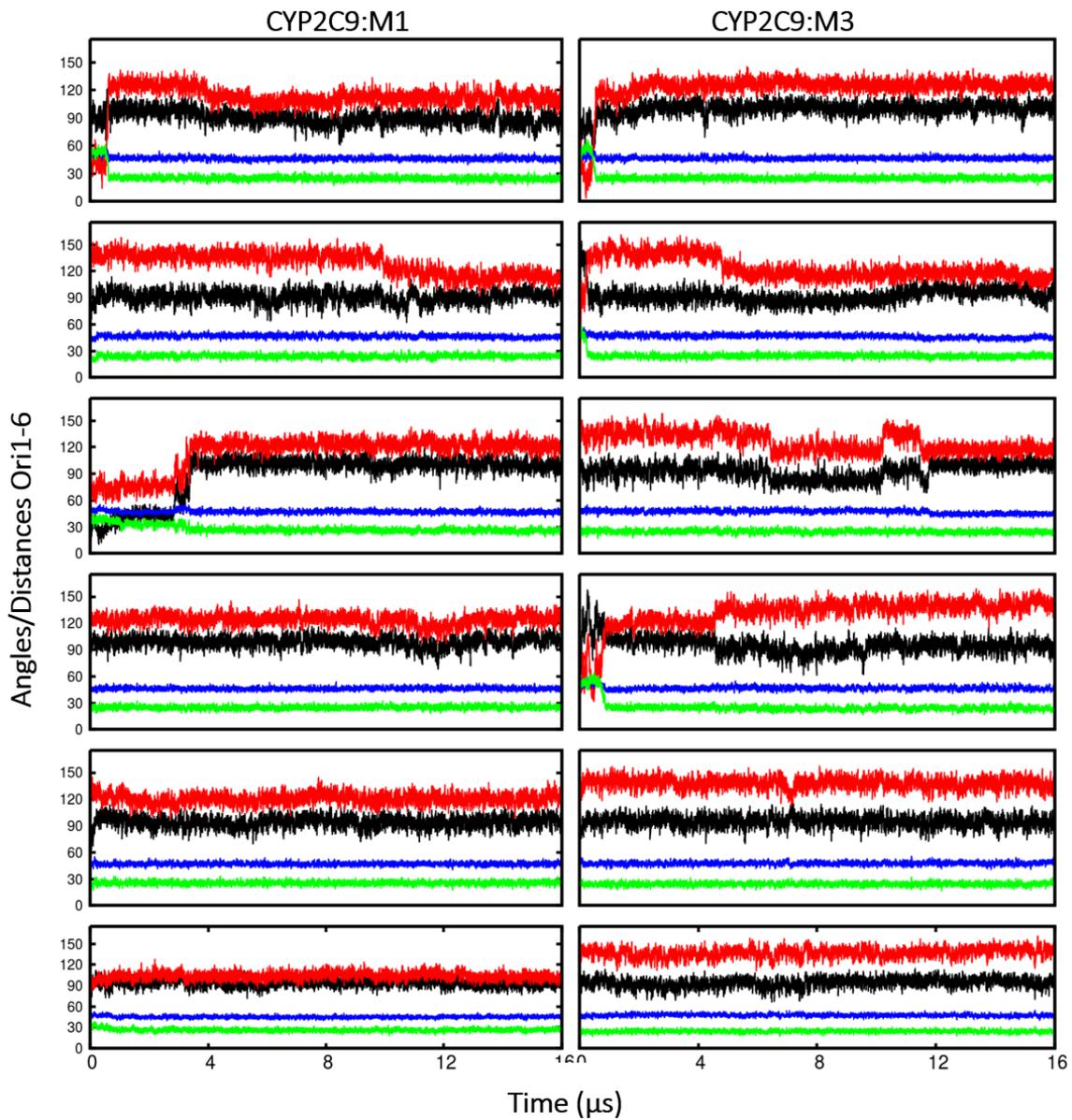


Figure 4.4: Angles ($^{\circ}$) and distance (\AA) values vs time (μs) for CG simulations of CYP2C9 M1 (left) and M3 (right). Color scheme same as in Figure 4.3.

Table 4.5: Angles and distance values obtained from CYP2C9 M3 CG simulations. The results are shown for six orientations after excluding the first 5 microseconds of simulation. * Mean and standard deviation for M1 and ori3 was calculated after 12 microseconds. For all other calculations first 5 microseconds were considered as equilibration time and was skipped.

CYP2C9 Orientations	Angles (°)		Distances (Å)			Class
	Alpha (α)	Beta (β)	Linker	F'-G'	Globular	
CYP2C9:M1						
Ori1	88±6	110±6	20±2	25±2	46±2	A
Ori2*	90±7	102±12	19±2	24±2	46±2	A
Ori3	101±6	123±6	21±2	26±2	47±2	A
Ori4	99±7	123±7	20±2	25±2	46±2	A
Ori5	94±6	120±6	21±2	25±2	47±2	A
Ori6	94±6	104±6	19±2	26±2	45±2	A
CYP2C9:M3						
Ori1	101±6	126±5	20±2	25±2	47±2	A/B
Ori2	92±7	116±6	18±2	24±2	46±2	A
Ori3	94±9	120±8	19±2	25±2	46±2	A
Ori4	92±7	139±7	22±2	24±2	47±2	B
Ori5	95±6	139±7	21±2	24±2	47±2	B
Ori6	95±6	138±6	22±2	24±2	48±2	B

4.3.4 Orientations of mutant models of CY2C9 and CYP2C19

Mutant structures of CYP2C9 (mtCYP2C9) and CYP2C19 (mtCYP2C19) were prepared by substituting residues different at the membrane interface, mainly in the linker, beta-strand1, BC loop and F'-G' helices (method section). For mtCYP2C9, 2 out of 5 orientations converged to CYP2C19-like orientations (class B) and one intermediate orientation (A/B) while 2 simulations converged to same orientation of CYP2C9 (class A). Same trend was observed in mtCYP2C19 simulations which converged to CYP2C9 orientations, 2 out of 5 converged to class A, one to the intermediate orientation (A/B)

and 2 simulations retained CYP2C19 wildtype orientations (class B), as shown in Table 4.6. Moreover, mutant simulations resulted in differences in the insertion depth of linker and F'-G' helices CoM distances, the regions where substitution mutations were introduced.

Table 4.6: The angle and distance values of mtCY2C9 and mtCYP2C19 calculated for the 5 orientations. Mean and standard deviations are calculated after 8.5 microseconds for mtCYP2C9 and after 1 microsecond for mtCYP2C19, to ensure correctness (see Figure 4.5)

mtCYPs Orientation*	Angles (°)		Distances (Å)			Class
	Alpha (α)	Beta (β)	Linker	F'-G'	Globular	
mtCYP2C9						
Ori1	102±5	125±5	27±4	30±4	44±2	A/B
Ori2	99±6	139±6	28±4	29±4	45±2	B
Ori3	98±6	123±6	27±4	29±4	44±2	A
Ori4	93±7	133±10	27±4	29±4	44±2	B
Ori5	98±6	124±6	26±4	29±4	44±2	A
Ori1-5	98±7	130±10	27±4	29±4	44±2	B
mtCYP2C19						
Ori1	93±6	114±6	20±2	25±2	47±2	A
Ori2	94±8	139±8	18±2	24±2	47±2	B
Ori3	102±6	128±5	19±2	25±2	46±2	A/B
Ori4	94±8	116±12	18±2	24±4	45±2	A
Ori5	94±7	138±6	21±2	24±2	47±2	B
Ori1-5	95±8	127±13	20±2	25±3	46±2	A/B

*initial orientations for all mutant simulations were the same as their wildtype CG system (S1)

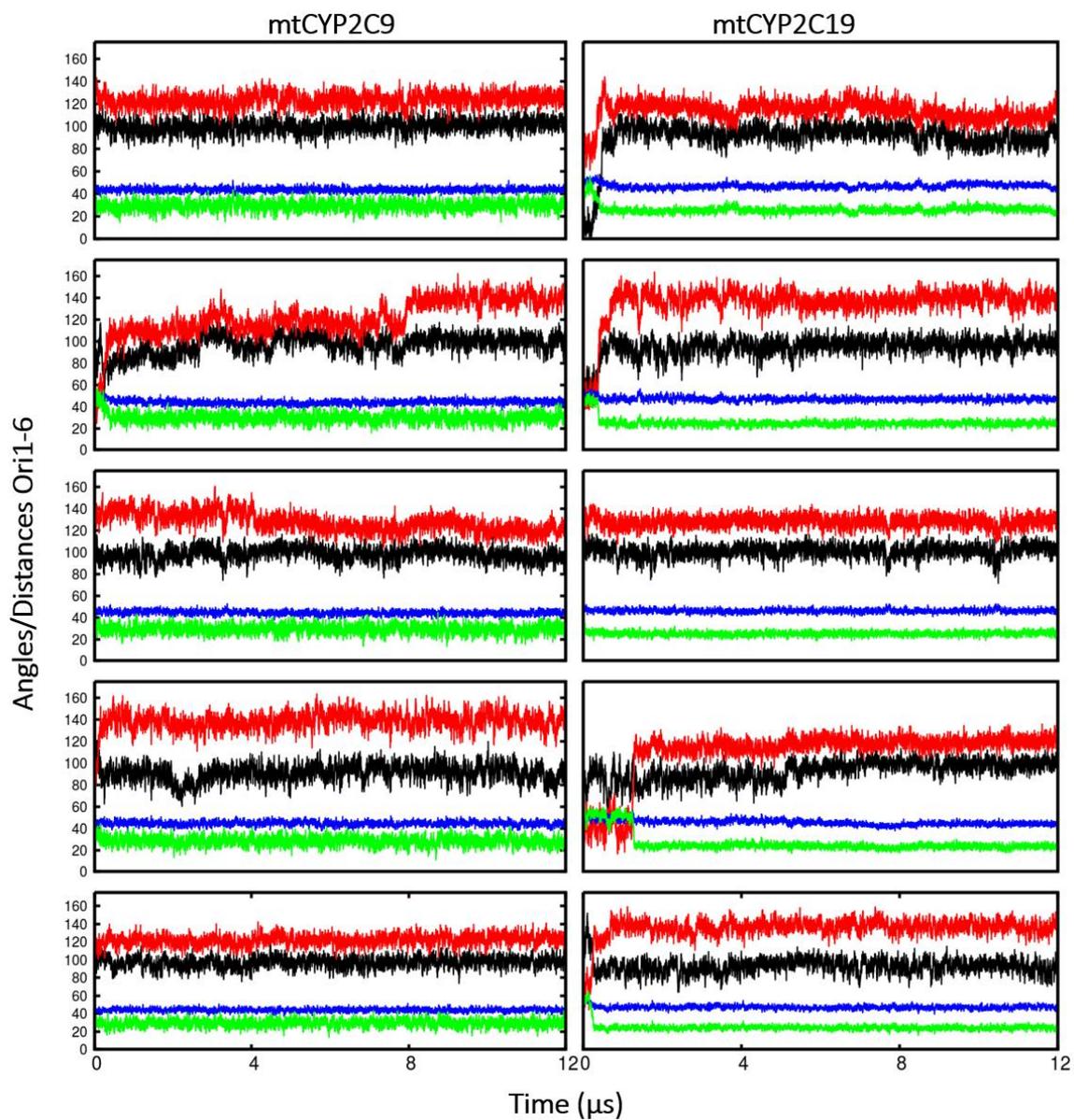


Figure 4.5: Angles ($^{\circ}$) and distance (\AA) values vs time (μs) for CG simulations of mtCYP2C9 (left) and mtCYP2C19 (right). Color scheme same as in Figure 4.4

4.4 All-atom simulation results

All atom simulations were performed to obtain refined atomic level interactions between the membrane and the proteins. Two different initial configurations obtained from CG simulations (S1 and S2) were used for AA simulations. To observe the effect of force-fields on protein membrane interactions and orientations, I also used two different combination of force-fields: LIPID14 and ff12SB, and GAFF lipid and ff99SB for the lipid and the protein respectively. The results were compared with previously reported work on CYP2C9. As mentioned in the Methods section, different pressure control methods were applied with the LIPID14 force-field. When using NPT MD ensemble for LIPID14, it is recommended to use anisotropic pressure coupling, which allows all three cell dimensions to fluctuate independently. However, in longer simulations (> 100 ns), the box dimensions were changed resulting in unequal x and y dimensions. Therefore, I also used semiisotropic pressure coupling using two different options given in NAMD software package: constant area and constant ratio. With the GAFF lipid, a surface tension of 60 dyn/cm was applied in the x-y plane. The effect of the two force-fields and different simulation protocols on the orientations of the proteins in the membrane, protein structure and dynamics and structural properties of membrane are compared in the following sections.

4.4.1 CYP2C9 orientations in the membrane

All atom simulations of CYP2C9 (SIM1-SIM8) with different starting structures, different initial velocities and different force-fields were analyzed for the orientations of the globular domain above membrane by calculating angles and distances as in CG simulations. In AA-simulations, heme-tilt angle is calculated in addition to angles calculated in CG simulations. All CYP2C9 simulations using LIPID14 force-field showed changes in the orientations from the starting conformations (Figure 4.6 to Figure 4.8). The angle and distance values increased to $89\pm 7^\circ$ to $101\pm 5^\circ$ and to $117\pm 5^\circ$ to $124\pm 6^\circ$ for alpha and beta angles, respectively. The heme-tilt angle increased from 30.2° and ranged from $35\pm 5^\circ$ to $44\pm 7^\circ$. In simulations using anisotropic pressure control (SIM1),

the heme-tilt angle decreased to $25\pm 5^\circ$ during first equilibration step and remained on lower side. In simulations with GAFF force-field (SIM5), no difference in the heme-tilt angle is observed from CG orientation. However, CoM distance of protein is decreased from initial position 45 to 39 Å. In previous studies with GAFF lipid, CoM distance was reported to decrease from 39.5 ± 2.5 Å (CG) to 34.1 ± 1.0 Å (31). In contrast, in simulation with LIPID14 force-field, CoM distance increased from starting distance 44 Å (CG_S1) to AA distance 48 ± 2 Å (upper limit), and from 42 Å (CG_S2) to AA 44 ± 2 Å (upper limit). Final orientation of CYPs in the membrane and normalized angle and distance distribution plots are shown in Figure 4.6 to Figure 4.8.

Table 4.7: Mean and standard deviations of angles and distances calculated for all-atom simulations of CYP2C9 using different starting configuration and different simulation methods

CYP2C9	Simulations methods		Angles (°)			Distance (Å)		Class
	Force-field	Pressure Control	Alpha (α)	Beta (β)	TM (γ)	Heme Tilt	Protein CoM	
CG_S1	NPW	Semiisotropic	91.9	111.9	17.6	30.2	45.0	A
SIM1	LIPID14	ConstantRatio	96.9±5.4	117.3±5.2	23.2±4.7	34.5±5.0	44.6±2.8	A
SIM2	LIPID14	ConstantArea	100.8±7.0	121.3±7.0	24.7±6.9	44.3±6.8	45.5±1.8	A
SIM3	LIPID14	Anisotropic	91.4±5.6	110.5±7.9	25.2±4.9	25.4±4.9	48.0±1.8	A
SIM4*	LIPID14	Anisotropic	96.7±4.0	124.3±5.6	16.2±4.9	38.9±6.2	47.1±2.0	A
SIM5	GAFF	Surface Tension	83.1±5.9	115±5.6	15.5±3.2	30.5±5.8	38.7±0.9	A
CG_S2	NPW	Semiisotropic	90.5	111.8	13.9	33.8	42.3	A
SIM6	LIPID14	Anisotropic	92.9±5.8	122.7±4.5	15.3±5.7	41.6±5.6	44.4±1.6	A
SIM7	LIPID14	ConstantRatio	94.8±4.9	122.1±4.0	17.0±4.0	40.8±5.6	44.2±1.5	A
SIM8	LIPID14	ConstantArea	88.8±6.8	122.6±3.9	11.3±6.4	41.3±5.4	43.9±1.9	A

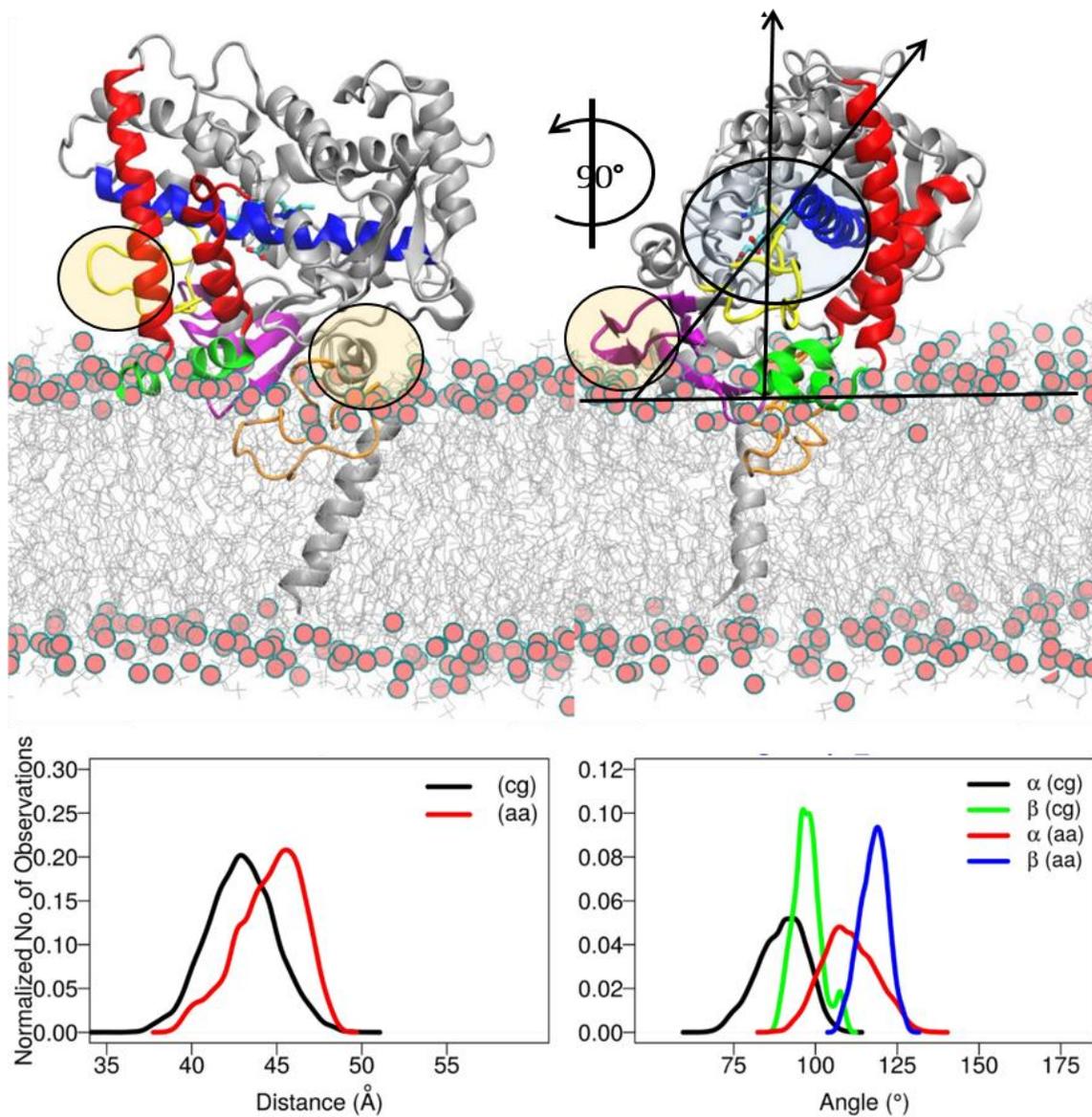


Figure 4.6: Last frame from CYP2C9:SIM1 all-atom simulation. Orientation of CYP in the membrane is shown in the top panels with different views of the same frame rotated by 90° around the z-axis. The protein is shown in grey colored cartoon representation. Linker is shown by orange, beta-strands in violet, BC-loop/helices in yellow, F-G helices in red and F'-G' helices in green. Central I-helix is colored blue, and heme is shown in cyan licorice representation. In the top left panel, the orange circle show positions of BC loop and A-helix above the membrane. Transparent circles in the right panel indicate heme position, and thick lines show angle formed by heme plane with membrane normal. Position of beta-strand2 (residues 370-380) in violet above the membrane is highlighted by transparent circle. POPC membrane is shown in grey lines with phosphate atoms in red spheres. In the bottom panel, the distance and angle distribution plots are compared between CG and AA simulations.

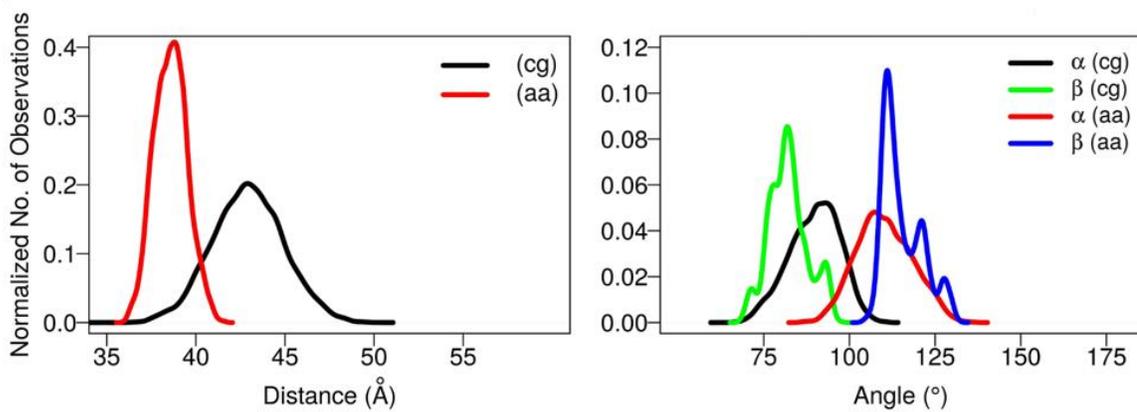
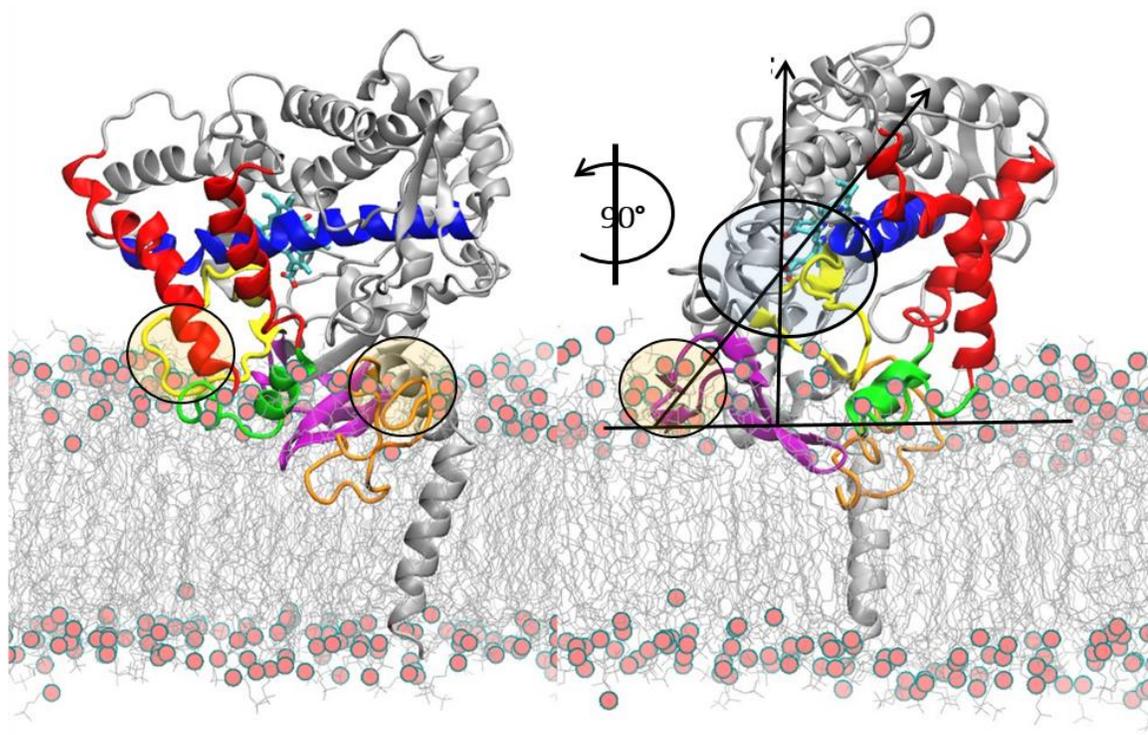


Figure 4.7: Last frame from CYP2C9:SIM5 all-atom simulation. Simulations were done using GAFF lipid. Top panel shows different view of the same frame rotated by 90° about z-axis. The representation is the same as in figure 4.5.

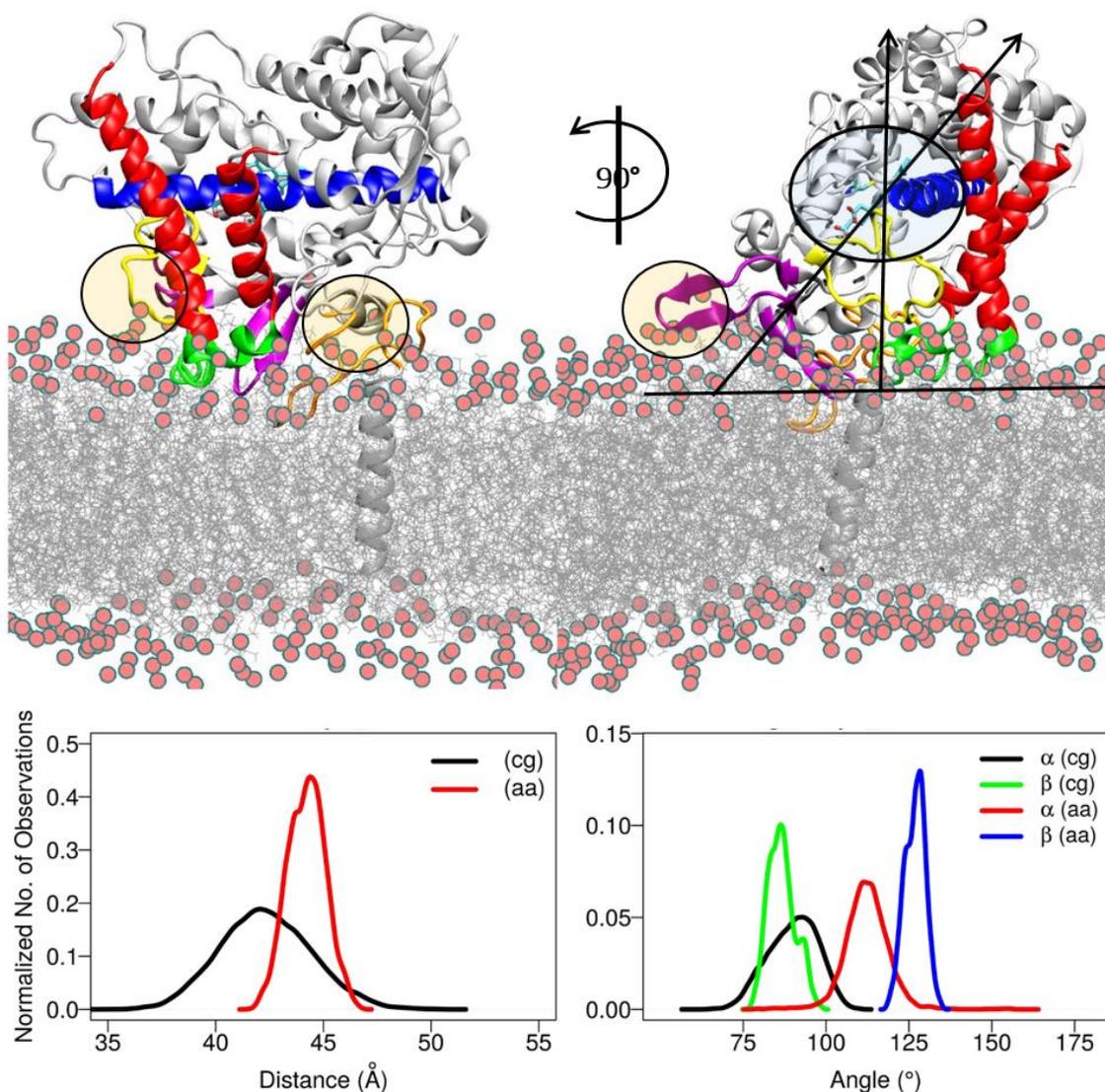


Figure 4.8: Last frame from CYP2C9:SIM7 all-atom simulation. CYP2C9:SIM7 showing different view of the same frame rotated by 90° about z-axis. Same representation as in figure 4.5.

4.4.2 CYP2C19 orientation in the membrane

CYP2C19 AAMD simulations were performed using both LIPID14 and GAFF lipid and with two different starting structures obtained from CG simulations (S1 and S2) as for CYP2C9. The starting structures of the two AA models in the membrane vary slightly in the heme-tilt angles, between 52° and 46° in CG_S1 and CG_S2, respectively. Distance CoM of globular domain is also different in two starting structures: 47 Å in CG_S1 and 50 in CG_S2 Å. After AA simulations using LIPID14 force-fields (SIM1-SIM3 and SIM5),

the orientations of CYP2C19 above the membranes fluctuate around same starting position as measured by alpha, beta and heme-tilt angles. The angles formed by CYP2C19 range between $93\pm 8^\circ$ to $103\pm 6^\circ$ for alpha, $132\pm 5^\circ$ to $147\pm 6^\circ$ for beta (class B), and $50\pm 6^\circ$ to $60\pm 6^\circ$ for heme-tilt angle. Distance CoM range from 46 to 50 Å in AA simulation as in CG simulations (CG_S1 and CG_S2). In simulations using anisotropic pressure (SIM1 and SIM5), an increase in heme-tilt angle is observed from initial value 52° (CG_S1) to 60° and from 46° (CG_S2) to 58° respectively, while distance CoM (46 Å) is consistent in two simulations (Table 4.8). A test simulation with GAFF lipid force-field (SIM4) was also performed for CYP2C19. As seen in CYP2C9 simulations with GAFF lipid force-field, CoM of protein decreased from 47 (CG_S1) to 44.6 Å. Orientation of CYP2C19 converged to same position as seen in CYP2C9 LIPID14 simulation (class A/B). Different orientations obtained in two force-fields and normalized distribution graphs for angles and CoM distance are shown in Figure 4.9 and Figure 4.10.

Table 4.8: Mean and standard deviations for angles and distances for all atom simulations of CYP2C19. Different starting configurations and different simulation methods were used.

CYP2C19	Simulations methods		Angles (°)				Distance	Class
			Alpha (α)	Beta (β)	TM (γ)	Heme Tilt	(\AA) Protein CoM	
AAMD	Force-field	Pressure Control						
CG_S1	NPW	Semiisotropic	99.6	135.3	13.0	52.4	46.7	B
SIM1	LIPID14	Anisotropic	102.8 \pm 5.5	146.9 \pm 5.5	19.7 \pm 9.0	60.2 \pm 5.5	46.6 \pm 2.6	B
SIM2	LIPID14	ConstantArea	93.8 \pm 7.9	132.0 \pm 4.9	15.5 \pm 4.7	50.3 \pm 5.6	50.3 \pm 2.3	B
SIM3*	LIPID14	Anisotropic	97.2 \pm 6.9	138.9 \pm 7.9	16.9 \pm 9.3	52.1 \pm 7.5	47.7 \pm 2.1	B
SIM4	GAFF	Surface Tension	82.6 \pm 4.7	128.8 \pm 3.6	14.2 \pm 3.3	46.3 \pm 1.1	44.6 \pm 1.1	A/B
CG_S2	NPW	Semiisotropic	99.5	133.3	10.2	45.9	50.3	B
SIM5	LIPID14	Anisotropic	96.4 \pm 4.6	138.5 \pm 4.2	16.6 \pm 4.1	57.9 \pm 4.5	46.3 \pm 1.4	B
SIM6	LIPID14	Semiisotropic	93.4 \pm 6.5	135.3 \pm 5.4	15.8 \pm 5.1	57.0 \pm 4.6	49.7 \pm 2.9	B

*CYP2C19 complexed with (2-methyl-1-benzofuran-3-yl)-(4-hydroxy-3,5-dimethylphenyl)methanone (Protein Data Bank chemical component 0XV)(23)

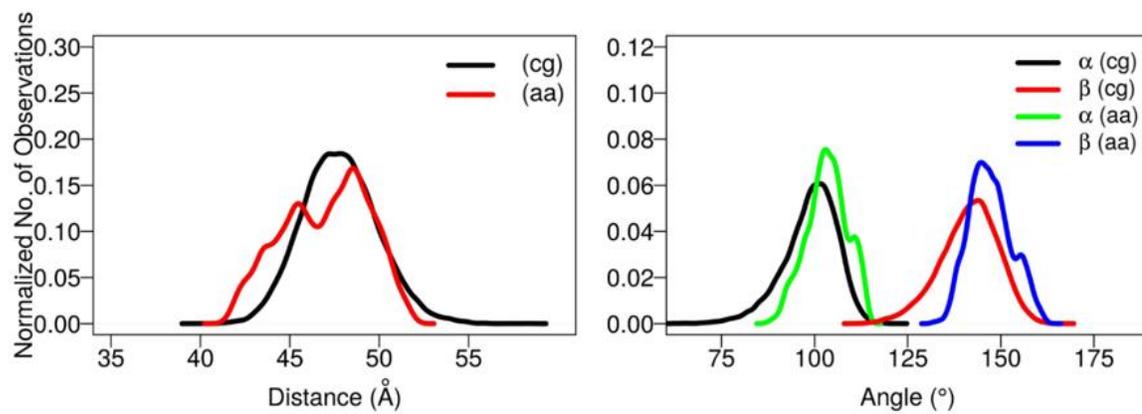
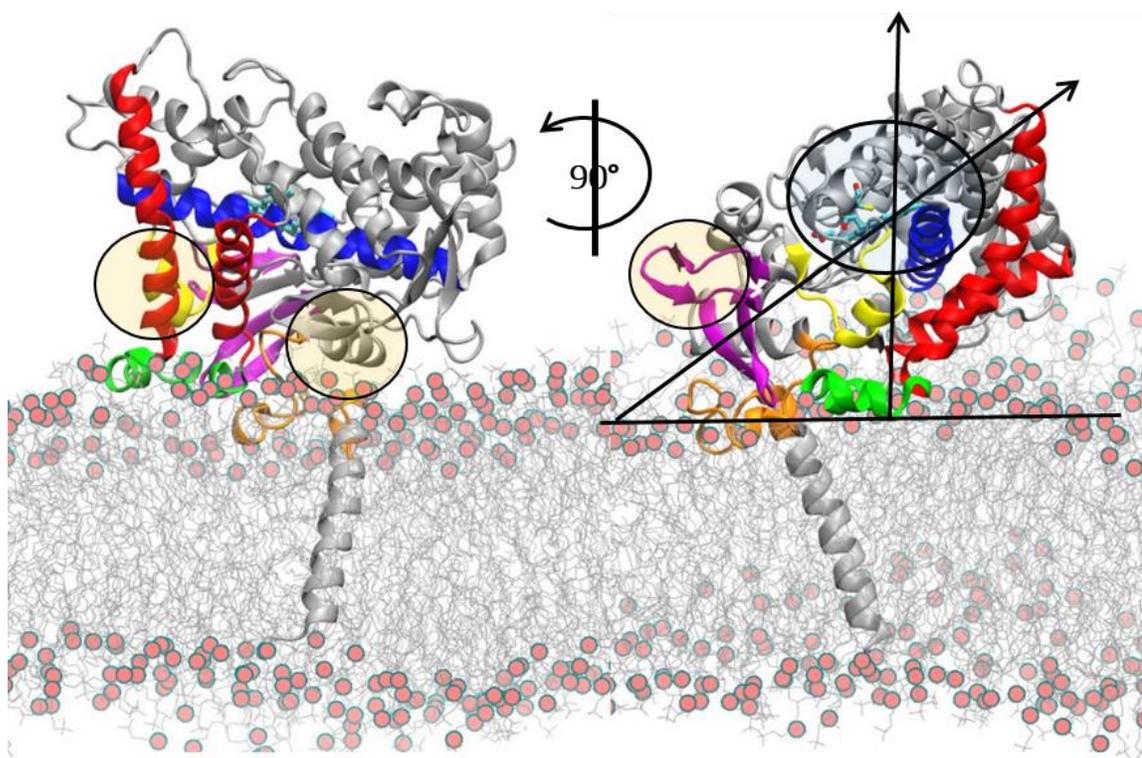


Figure 4.9: Last frame from CYP2C19:SIM1 all-atom simulation. Top panel showing different view of the same frame rotated by 90° about z-axis. The representation is the same as in Figure 4.6.

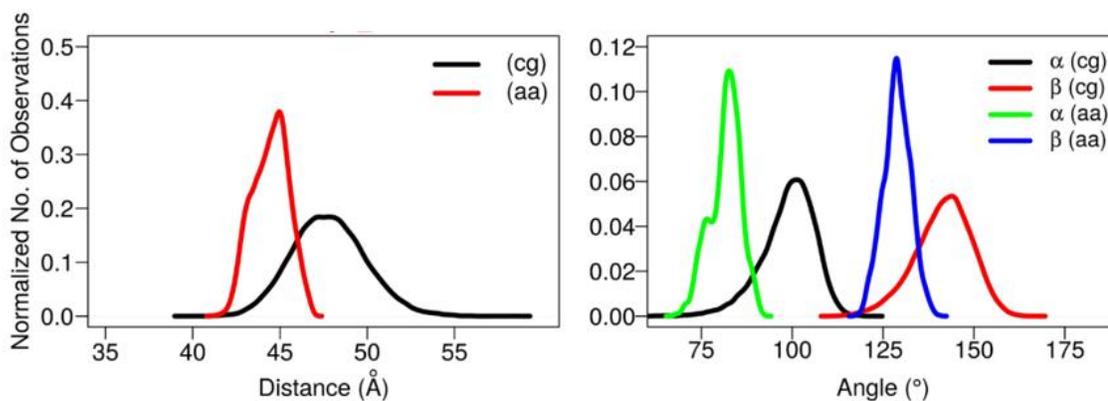
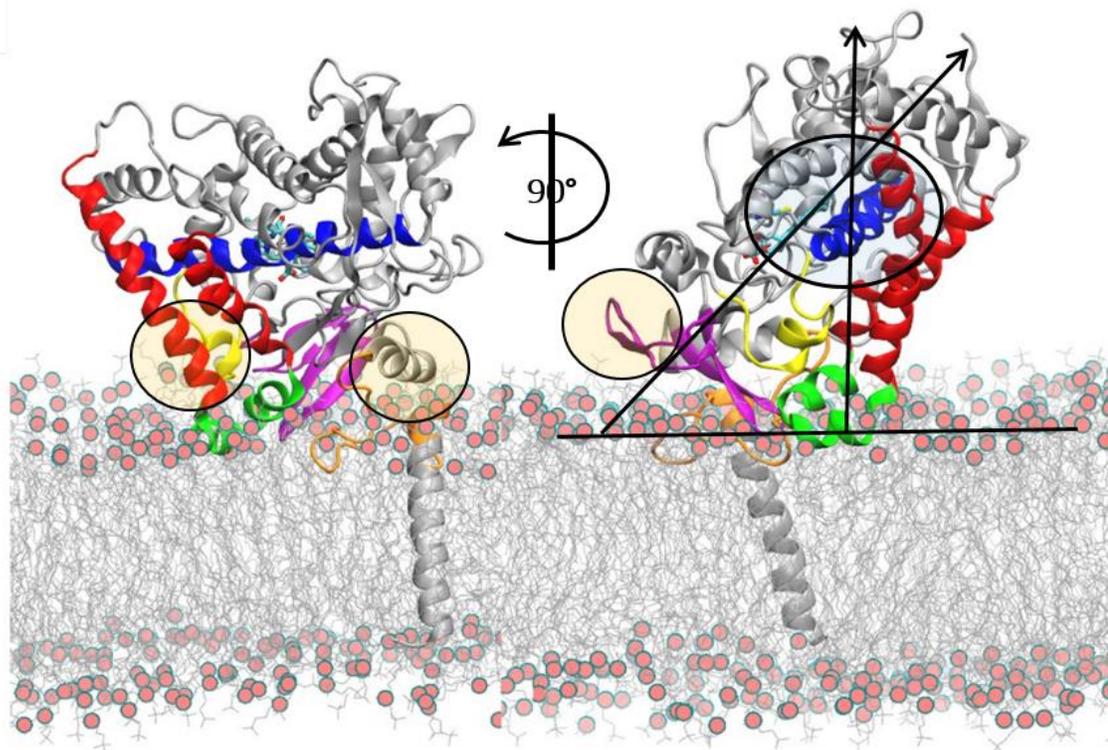


Figure 4.10: Last frame from CYP2C19:SIM4 all-atom simulation. Top panel showing different view of the same frame rotated by 90° around z-axis. The representation is the same as in Figure 4.6.

The differences in the orientations and interactions of two proteins (CYP2C9 and CYP2C19) were maintained in both CG and AA simulations using LIPID14 force-field. However, AA simulations with two different force-fields show different angle and distance values. Simulations with LIPID14 results in increased angles and distances compared to initial values. On the other hand, in simulations using GAFF lipid force-

field, CoM distance decreased, corresponding to further insertion of protein in the membrane. In the following sections, I analyzed differences in protein-membrane interactions in relation to primary sequence and 3D conformations and highlight the regions that show direct interactions with the membrane. I also compared the effect of the two force-fields on the membrane-protein interactions and orientations.

4.4.3 Effect of sequence on the membrane-protein interactions:

From above AA and CG simulation results, it has been established that two isoforms of CYP2C family, despite high sequence conservation (92% sequence identity), maintain differences in the interactions, orientations and degree of insertion in the membrane. The most important residue differences are found in the substrate recognition sites (SRSs) as seen in the primary sequence analysis. For example, SRS1' a, SRS1' b, SRS1 and SRS2,3 regions defined by Zawaira *et al* (29) cover residues in the linker region (G46D), beta-strand (72-73 KP- ER), BC loop (I99H) and F'-G' helices (220-221 SP-PT), respectively (Figure 4.11). I have observed through simulations that secondary interactions are mainly developed through beta-sheet1 (residues 64-74), and F'-G' helices (residues 210-226). However, CYP2C9 shows further interaction through A-helix (residues 50-60), B-C loop (residues 95-110) and C-terminal beta-sheet2 (residues 370-385). The peripheral interactions developed by CYP2C9 are similar to the hydrophobic surface identified in CYP2C5 (residues 30-45, 60-69 after A-helix, 376-379 beta-strand 2 and the F'-G' helices) (27). These secondary interactions, with either lipid tail or head regions, are established by the SRS regions, that show primary sequence differences in two CYPs. Therefore, differences in SRS regions can be crucial for CYP-membrane interactions and orientation, and lead to distinct substrate specificity.

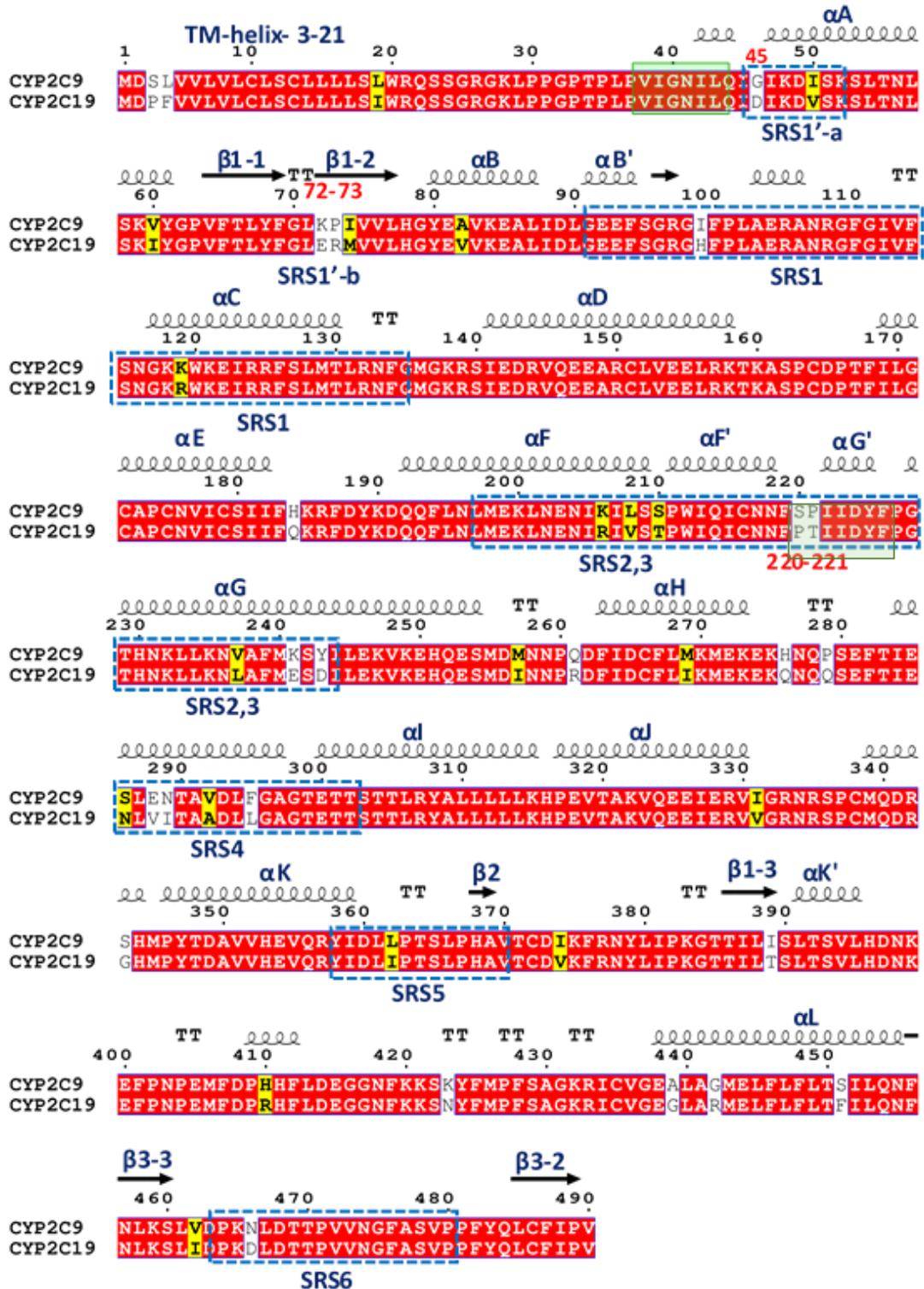


Figure 4.11: Sequence alignment between CYP2C9 and CYP2C19. The residues that are identical are shown in red and similar residues are shown with a yellow background. The residues different in two sequences have white background. The secondary structure information obtained from CYP2C19 crystal structure is given by arrows for beta-strands

of different residues in SRS regions in distinct orientations of two CYPs, leading to different substrate access and selectivity.

Furthermore, trajectory analysis was performed to differentiate residues interacting with membrane head or tail regions, calculated by defining 5 Å distance cutoff between protein and lipid head group (phosphate atoms) and hydrophobic tail separately. The % contact time or occupancy of these residues was calculated (Figure 4.13) in trajectory. From the occupancy graph, the different interacting regions and therefore different orientations adopted by two proteins in the membrane can be explained. As shown in the graph, most interacting residues are found on the proximal side (BC-loop side) of the protein and in substrate recognition regions SRS1'-a, SRS1'-b and SRS5 while few or no contacts are seen in these regions by CYP2C19. The residues in A-helix and beta-sheet1 are more in contact with lipid head and tail regions in CYP2C9 as compared to CYP2C19 which showed no or fewer contacts with membrane. In CYP2C19, secondary interactions are mainly formed by a patch of F-G, F'-G' helices (residue 207-231) which resulted in the tilting of CYP2C19 from the distal side. The tilted F-G helices resulted in the increased beta-angles (discussed in earlier section) and higher heme-tilt angle in all CYP2C19 simulations.

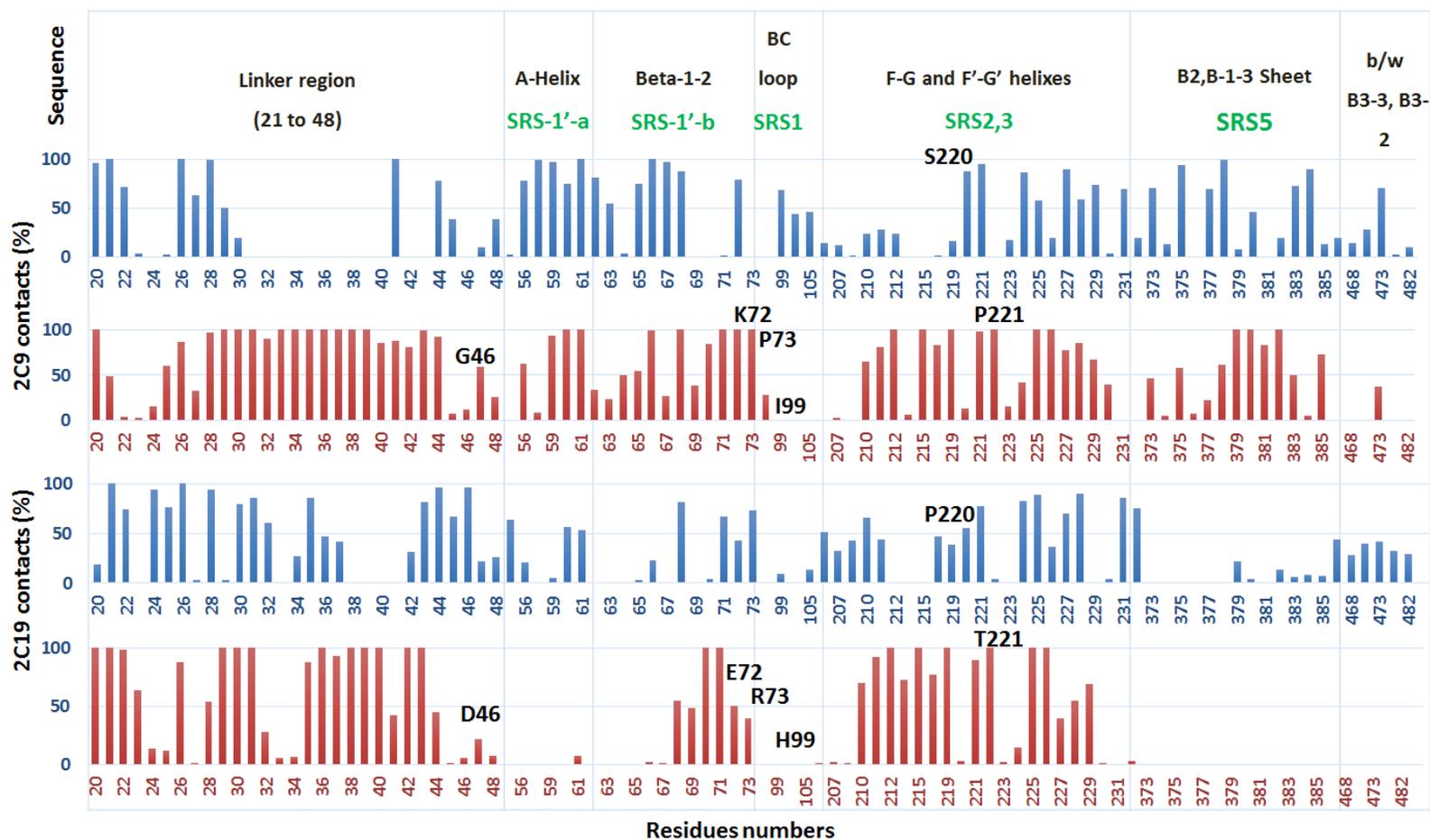


Figure 4.13: Residues in contact with lipid headgroup (blue) and tail region (red). The percentage is shown on the y-axis and residues (numbers) interacting with membrane are given on x-axis. The secondary structures and substrate recognition regions are shown on the top. The residues differing in the interacting regions between CYP2C9 and CYP2C19 are labeled.

4.4.4 Structural differences result in different membrane-protein interactions

Apart from sequence differences, the structural analysis of CYP2C9 and CYP2C19 (PDB 1R90 and 4GQS, respectively) revealed different conformations of beta-sheet1 and the B-C loop. The beta-sheet1 in CYP2C9 differs in sequence, (residues K72-P73), from CYP2C19, (residues E72-R73), as well as conformation. The turn in beta-sheet1 in CYP2C9 points away from globular domain towards membrane surface. During simulations, beta-sheet1 remained buried in the membrane or interacted with lipid head groups in CYP2C9 (SIM1, SIM5 and SIM7 shown in Figure 4.6-4.8. K72 in CYP2C9 is pointed towards binding pocket and formed hydrogen bond with S220 in the F'-G' helices as well as phosphate head groups of lipid molecules. During simulations, positively charged ϵ -amino group (cation) of K72 made hydrogen bonds with the phosphate headgroups of lipids, which resulted in further insertion of beta-strand residues in the membrane (Figure 4.14 A-B). K72 has been suggested to play important role in the selection of anionic substrates in CYP2C9 and is found in pathway 2b (descriptions of pathways is given in Chapter 1) for ligand entrance into the binding pocket from the membrane (153); it is replaced by E72 in 2C19. Besides K72, presence of P73 in CYP2C9 favors interactions with hydrophobic interior of the membrane (also seen in the % occupancy graph). Thus, K72 and P73 position in beta-strand could be responsible for the orientation difference in CYP2C9. The conformation and orientation of beta-sheet1 also favors interactions of beta-sheet2 (residues 370-385) with membrane head groups in CYP2C9. In CYP2C19 simulations, beta-sheet1 remained out and showed fewer interactions with the membrane surface (Figure 4.9-4.10) which could be attributed partly to its charged residues E72 and R73 and the conformational difference observed in crystal structure. R73 in CY2C19 restricts insertion of beta-strand in the membrane (Figure 4.14 C-D). In various studies on membrane-protein interactions, it is found that arginine has propensity to stay in the lipid headgroup region (154–156, 156, 157).

An important difference between CYP2C9 and CYP2C19 is seen in the BC loop, which differs in only one residue I99H (H99 in CYP2C19) but has loop conformation in CYP2C9 that is highly mobile compared to the B'-C' helical conformation in CYP2C19. The BC loop also differs in the side chain conformations of R105 and R108 in the two CYPs. R105 in CYP2C19 points downward and shows electrostatic interactions with D224 in G' helix. In CYP2C9 X-ray structure (PDB 1R90), R105 has a different conformation and no interactions are reported with D224 due to the missing G'-helix. However, after modeling of F'-G' helices and simulations, similar R105-D224 interactions were observed in most simulations. Another important difference between the two CYPs is in F'-G' helices (S220P and P221T). P221 in CYP2C9 is located on the outer surface of the G'-helix which is in direct contact with membrane and would favor insertion of P221 in the lipid tail, while T221 at same position in 2C19 has more propensity to stay in the lipid head group region.

In CG simulations section, I have discussed the role of F'-G' helices in stabilizing the interactions and orientations of CYPs in the membrane. By introducing mutations or changing the initial conformations of models in SRS regions of CYP2C9, the CG simulations revealed convergence to different orientations. This implies that not only sequence differences but conformational changes in the regions involved in membrane-protein interactions contribute to the differences in the orientations adopted by two isoforms in the membrane. While, in AA simulations with different starting structures and velocities, CYP2C9 converged to the orientations which agree with previously reported orientations formed by CYP2C9 (31).

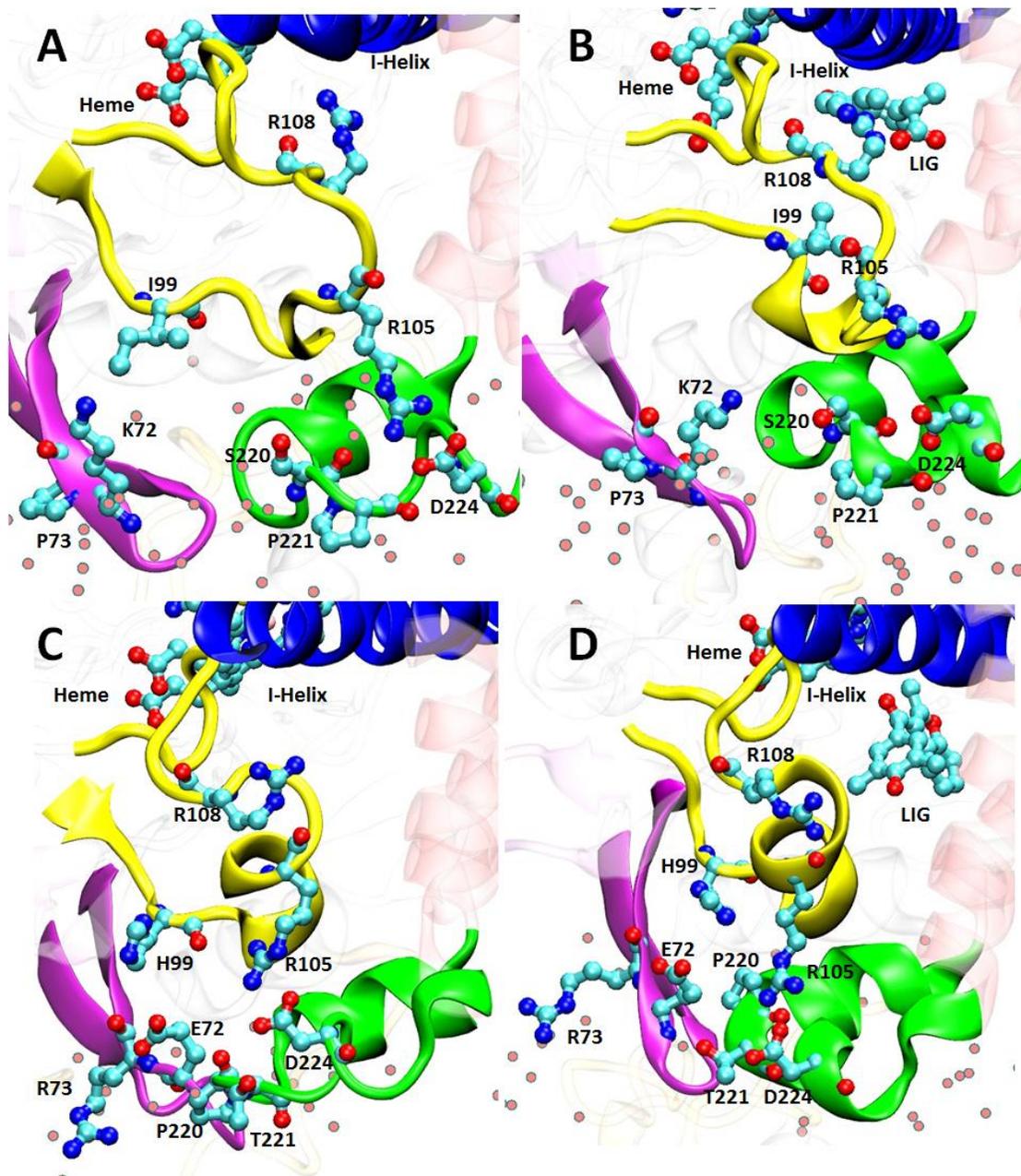


Figure 4.14: Differences in the CYP2C9 and CYP2C19 residues at the membrane interface. The residues in CYP2C9 apo form (A) and ligand (flurbiprofen) bound from (B) are shown in cyan colored stick representation. The protein is shown in cartoon representation with different colors assigned to secondary structures. Beta-sheet1 is shown in violet, BC loop in yellow, F'-G' helices in green and I-Helix on the top is colored blue. The rest of protein is shown in transparent grey and red color. Same representation is followed for CYP2C19 apo (C) and ligand (OXV) bound form (D).

4.4.5 Different all-atom force-fields and simulations parameters affect dynamics of proteins

The root mean square deviation (RMSD) was calculated for CYP2C9 and CYP2C19 by superimposition of backbone C α atoms of globular domain (50-490) on the reference minimized structure. RMSDs were compared for all simulations using different force-fields, pressure control and ligand free and ligand bound conformations (Figure 4.15). The RMSD graph suggests that the structure deviated during initial equilibration step by 2.5 Å and then remained stable during production simulations. MD simulations using ff12SB and LIPID14 with different pressure coupling, show stable RMSD in CYP2C9. Fluctuations in RMSD graph for ff99SB and GAFF lipid force-field are observed during simulations. In case of CYP2C19, difference in RMSD in different simulations is visible. A large difference is seen between apo and ligand bound simulations with LIPID14 force-fields and anisotropic pressure coupling. Higher RMSD is also seen in GAFF lipid force-field.

Furthermore, B-factor values were computed and compared with crystal structure B-factors (Figure 4.16). Large fluctuations are observed in the regions which also show high crystallographic B-factor values. CYP2C9 and CYP2C19 crystal structures show different flexible regions, as seen by crystallographic B-factor values. Comparing the CYP2C9 and CYP2C19 crystal structures, higher B-factor values are seen in CYP2C19 in the HI loop, the meander region (consisting β -bulge region and the K' helix region), and beta-strand 3 (residues 460-475) (Figure 4.17). In CYP2C9 crystal structure and in simulations with LIPID14, the membrane interacting regions such as linker, beta strand2, BC loop and F-G helices and beta-strand 2 remain flexible and the presence of membrane does not restrict the flexibility in these regions (Figure 4.16). There is increased flexibility in the GH and HI loop when using LIPID14 and anisotropic pressure control or constant area pressure control methods. The same effect is observed with GAFF lipid and surface tension simulations. LIPID14 simulations with constant ratio show less flexibility except in linker region, where reorientation of linker residues in the

membrane occurs. For ff99SB and GAFF lipid, there is increased flexibility in the loop regions outside the membrane, especially BC loop, EF loop, GH, HI loops and F and H helices (Figure 4.16). The increased flexibility in the BC loop results in a wide tunnel (tunnel 2) leading to the binding pocket. With GAFF lipid and ff99SB force-field, the BC loop shows increased interactions with the membrane headgroups. In CYP2C9 simulations with GAFF lipid force-field, secondary structure distortion or bending of helices is observed in the C-terminal end of the F-helix (residues 203-208), C-terminal end of G-helix (residues 242-250) and unwinding of central I-helix (residues 293-295) in CYP2C9 simulations (Figure 4.18). Secondary structure calculation is performed on last 4 ns trajectory time (2000 frames) for each simulation using LIPID14 and ff12SB and compared with GAFF lipid force-field.

In apo CYP2C19 simulations using LIPID14 and ff12SB force-fields, similar fluctuations are observed corresponding to the X-ray B-factor values (Figure 4.16). However, changing the pressure coupling to anisotropic, higher fluctuations are observed in the linker residues preceding A-helix, F'-G' helices and Beta-strands 3 and 4. Presence of membrane does not decrease the flexibility of these regions, rather, increased B-factor is seen in linker and F'-G' helices. The linker flexibility is consistent with the changing transmembrane helix angle. Conformational changes in the linker (patch of polar residues) result in the re-orientation of charged residues in the linker to interact with polar headgroup of the membrane. High fluctuations in F'-G' helix could be partly due to unwinding of the G'-helix (

Figure 4.19). In simulations using GAFF lipid and ff99SB force-fields, high fluctuations are seen in the meander region, beta-strand 3 and 4 (residues 460-475). Besides, distortion in the F-helix (residues 205-208) prior to the F'-helix, and unwinding in the I-helix is also observed in CYP2C19 with GAFF lipid and ff99SB (

Figure 4.19).

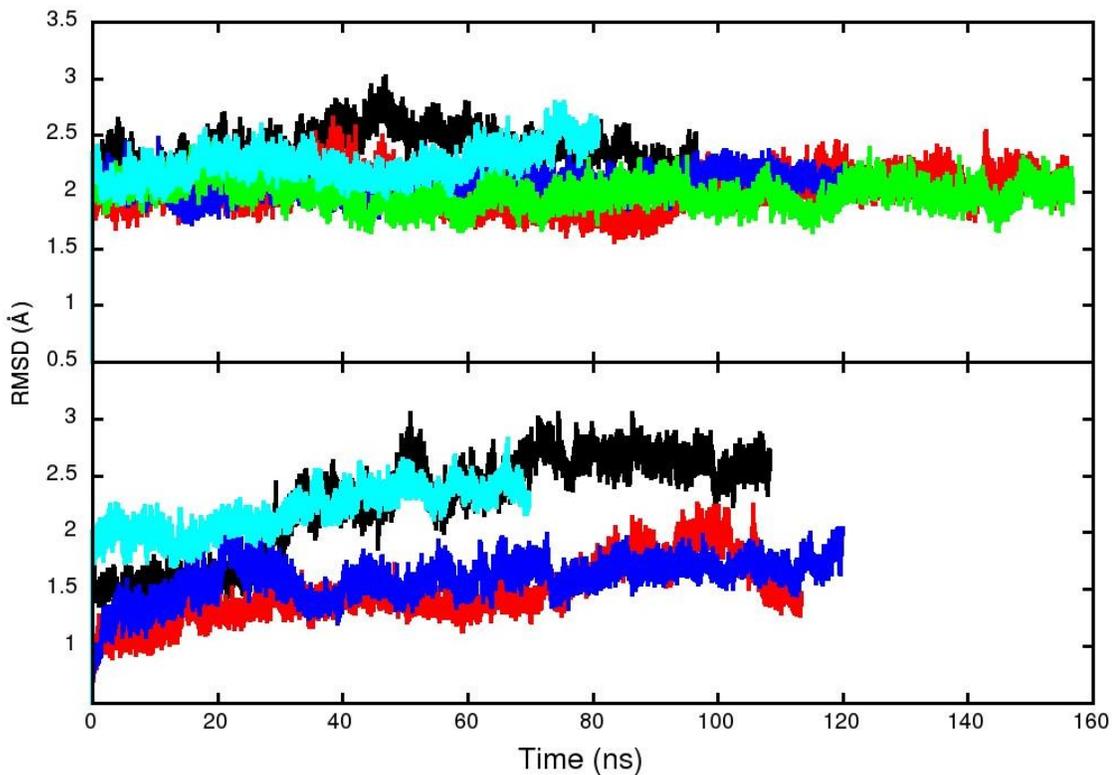


Figure 4.15: Root mean square deviation (RMSD) plots of CYP2C9 (top) and CYP2C19 (lower) backbone C α atoms with respect to minimized structure are shown. RMSD results from AAMD simulation starting with structure from CG_S1 are shown for CYP2C9 in top panel and for CYP2C19 in lower panel. Apo simulations using LIPID14 and anisotropic pressure control (thick black lines) and ligand bound (red). Apo simulation with constant area in blue and with constant ratio shown in green. Apo simulations using GAFF lipid are shown in cyan.

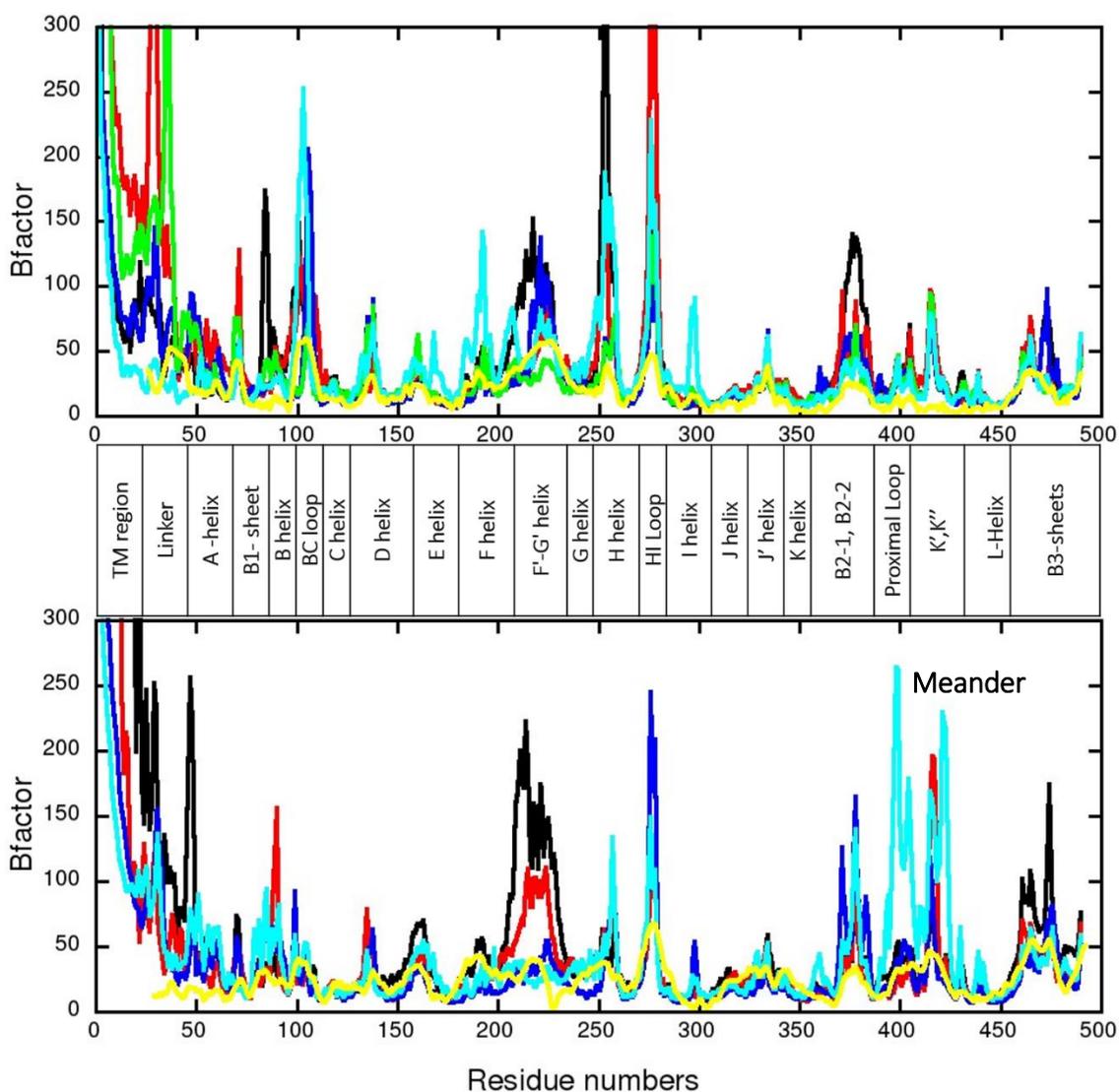


Figure 4.16: Comparison of the average B-factor values ($8\pi^2 rmsf^2/3$) of CYP2C9 (top) and CYP2C19 (bottom) in AAMD simulations are compared with X-ray Structure B-factor (yellow). Simulation with LIPID14 and anisotropic pressure control shown in thick black lines (apo) and in red (ligand bound), with constant area pressure control (blue), constant ratio (green) and GAFF lipid with surface tension (cyan).

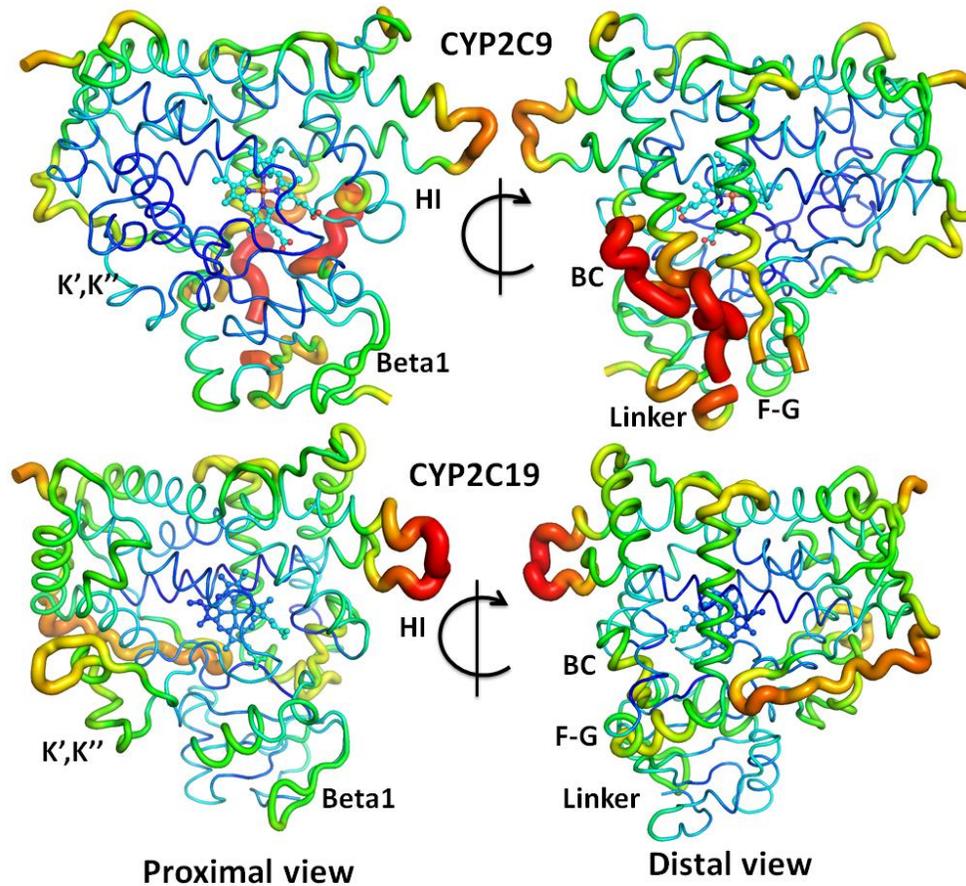


Figure 4.17: X-ray crystal structure B-factor values for the two CYPs (CYP2C9 PDB 1R90 and CYP2C19 PDB 4GQS). The proximal view is from where heme-cofactor is interacting with CYP reductase enzyme. The distal view is where solvent tunnel opens. Increasing B-factor is represented by color scale and thickness of ribbon from blue to red and thin to thick respectively. Heme is shown in the center in stick representation.

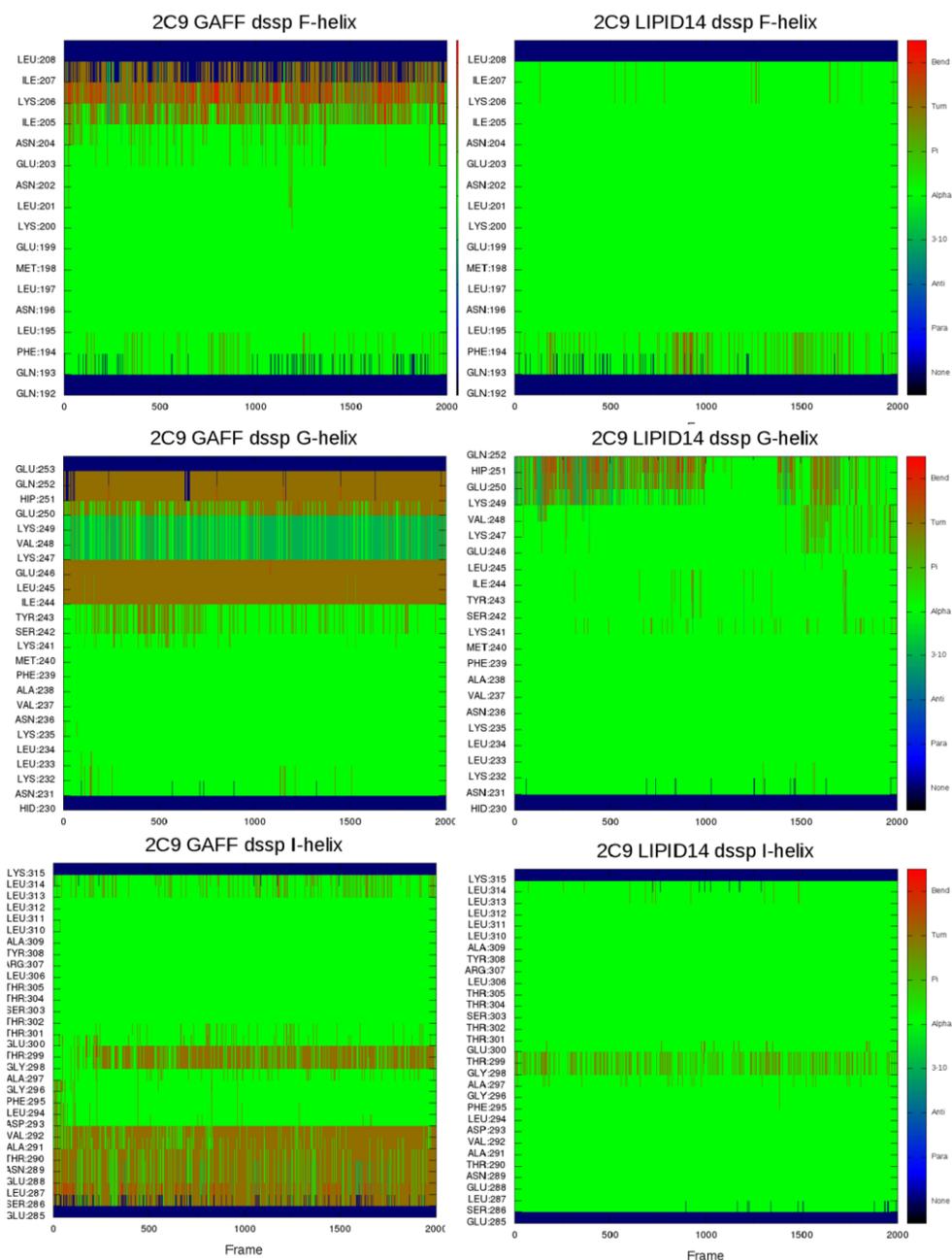


Figure 4.18: Secondary structure analysis of CYP2C9. Comparison between two force-fields, GAFF with ff99SB (left) vs LIPID14 with ff12SB (right). The color bar on the right shows different secondary structures. Secondary structures of F, G and I helix regions are compared.

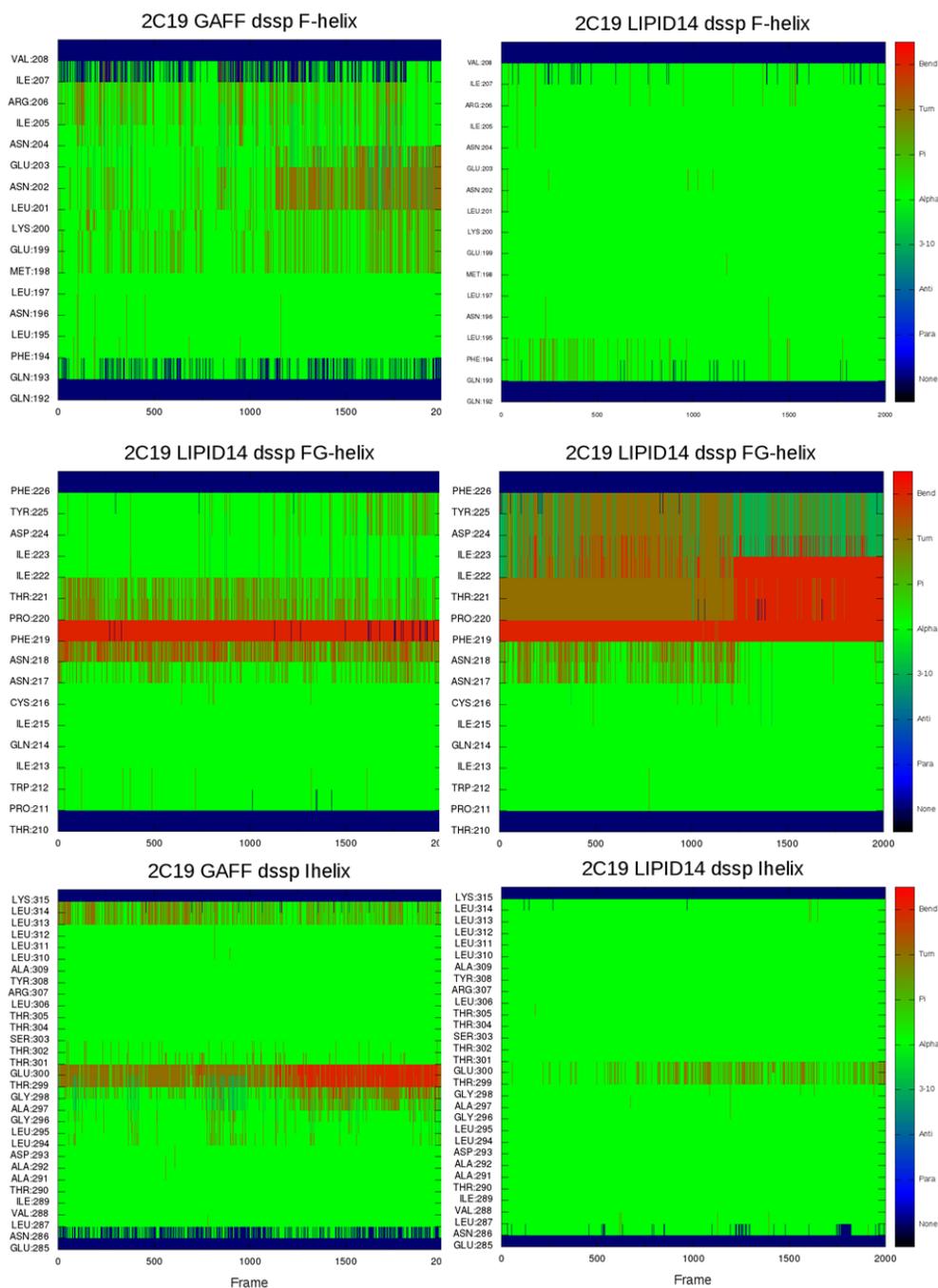


Figure 4.19: Secondary structure analysis of 2C19. Comparison between two force-fields, GAFF lipid and ff99SB (left) vs LIPID14 and ff12SB (right) force-fields. The color bar on the right shows different secondary structures. F, F'-G' and I-helices in CYP2C19 are compared and shown in descending order.

4.4.6 Structure and dynamics of the membrane is affected by choice of force-field

While studying membrane protein interactions, it is important to study the bilayer properties. Proteins can influence membrane properties as much as membranes shape the structure and function of proteins. Therefore, I examined the basic structural characteristics and dynamics of the membrane. The most common structural properties considered are membrane thickness, surface area occupied by each lipid, commonly referred as area per lipid (APL), and electron density profile. The lipid molecules in the vicinity of the protein are called boundary lipids and those having no contact with protein are named non-boundary lipids. The non-boundary lipids or bulk lipids are expected to have characteristics of pure lipid. Compared to previous AMBER force-fields for lipids, the recently optimized LIPID14 force-field has shown improved structural properties without applying surface tension in MD simulations. However, these MD simulations were performed on pure phospholipids and to date no data is available on the compatibility of protein and lipid force-fields together. Since the Amber parametrization strategy has been followed, it is assumed that the new LIPID14 force-field will be compatible with Amber protein and small molecules force-fields.

4.4.7 Area per lipid

The APL for pure lipid is calculated simply by multiplying the dimensions of the simulation box and dividing by the number of lipids in one layer. To calculate the APL of the lipids interacting with protein molecules, I have used a tool developed by Mori et al. 2012 (139) which combines both Voronoi tessellation and Monte Carlo methods (VTMC). Each lipid is assigned to a Voronoi polygon. Where the Voronoi polygon contains protein atoms, the lipids are termed boundary lipids. For calculation of APL for boundary lipids, the Monte Carlo method is used to calculate explicitly the area occupied by one lipid molecule in that polygon. APL for non-boundary lipids is computed simply from the area of Voronoi polygon. Changes in the box-dimensions, number of boundary and non-boundary lipids and average APL for boundary, non-boundary lipids

and total lipids are shown in Table 4.9 for first and last frames of the MD simulations and Figure 4.20.

In all simulations with LIPID14, average APL of bulk lipids and all lipids is consistent with experiments. However, APL of boundary lipids is decreased due to interactions with protein. In an anisotropic system, despite changes in box dimensions, there is no effect on the average APL. In simulations using constant area pressure control, the box dimensions remain constant which prevents further relaxation of the system. Thus, increased number of boundary lipids and decreased average APL of the boundary lipids is seen. Simultaneously, APL of the bulk lipids and all lipids in constant area simulations was within the experimental range. The opposite effect is observed in the CYP2C9 simulations with the GAFF lipid force-field, where box dimensions and number of boundary lipids increased along with the APL of non-boundary lipids. Visual inspection of GAFF lipid simulations and order parameters discussed below suggests highly ordered lipids. These might be expected to have decreased APL but voids between lipid molecules may contribute to the increased average APL. The increase in APL can be seen in the increased size of Voronoi polygons, Figure 4.20 (right panel). The increase in the number of boundary lipids corresponds to increased protein-membrane contacts and decreased CoM distance of protein to the CoM of the membrane with deeper insertion of the protein in the membrane.

Table 4.9: Area per lipid and box dimensions are shown for the initial and final frames obtained from MD trajectories of CYP2C9 under different conditions.

CYP2C9 AAMD	Time (ns)	Cell Dimensions (Å)		Number of lipids		Average APL (Å) (Exp: 64.3,68.3 (141, 142), Sim 65.6 (136))		
		X-Y length		Boundary	Non-Boundary	Boundary	Non-Boundary	All
LIPID14	Initial	142.4	138.1	49	545	51.7	64.9	63.8
Anisotropic	193	131.3	150.8	48	546	50.2	65.4	64.1
Constant Area	120	142.4	138.1	54	540	47.5	64.9	63.4
Constant Ratio	97	143.6	139.3	49	545	49.5	66.2	64.9
GAFF Lipid	80	150.6	150.1	78	516	57.9	74.2	72.1

LIPID14

GAFF Lipid

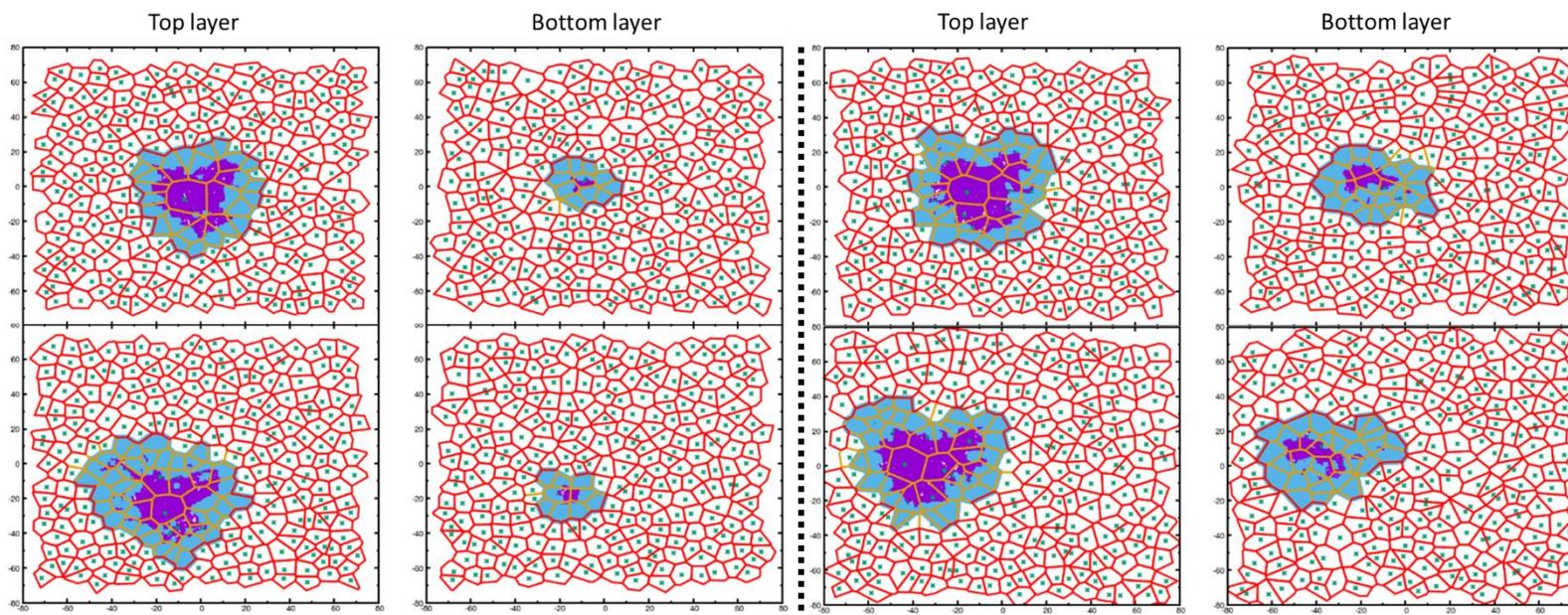


Figure 4.20: The area per lipid calculations. CYP2C9 simulations with LIPID14 (SIM1) (left panel) are compared with GAFF lipid simulation (SIM5) (right panel), separated by dotted line. APL is, computed for first frame and final frame by Voronoi tessellation for non-boundary and Monte Carlo integration for boundary lipids. The APL is shown for top and bottom layers separately. The assignment of lipid molecules as boundary (cyan) and non-boundary (white) lipids are represented by polygons. The violet regions show protein atoms. The + marks show the center of mass of each lipid molecule.

4.4.8 Electron density and membrane thickness

Bilayer thickness was calculated using the peak-to-peak distance of phosphate head groups in the electron density map. The membrane thickness has a direct relation with APL as increased APL results in decreased thickness. The experimental thickness for fully hydrated POPC is 37Å (142) and the reported thickness for pure lipid bilayer simulations using LIPID14 force-field is 36.9±0.6 Å (136).

Experiments such as X-ray scattering, electron, or neutron diffraction are used to obtain structural information on lipid bilayers (141, 142, 158–160). The structural information on lipid bilayers is important for understanding protein-membrane interactions. The electron density profile is computed from atomic distribution in membrane normal. The membrane in Z-dimension is divided into slabs of 0.1 Å thickness. The calculated electron density in the slab is then divided by the average cross-sectional area to obtain electron density in $e/\text{Å}^3$. The electron density is computed and time averaged for the last approximately 66 ns, 58 ns, 58 ns and 76 ns for anisotropic, constant area, constant ratio and simulations of CYP2C9 with GAFF lipid force-fields, respectively. The most important information the electron density profile can provide is the thickness of bilayer from the distance between the two highest peaks. The density profile can also give information about the hydration shell and insertion of water molecules in the lipid membrane (161). In Figure 4.21, electron densities are shown for head groups (max 0.43 $e/\text{Å}^3$) and the bilayer tail regions and indicated by ρ_H and ρ_C , respectively. The electron density below 0.3 $e/\text{Å}^3$ is contributed by methylene groups ρ_{CH_2} (145). Bilayer thickness is defined as the distance between two headgroup peaks (d_{HH}). Some water molecules interacting with the lipids are considered as part of the bilayer and contribute to the bilayer thickness (d_{HH}). Therefore, the headgroup size (d_H) is estimated to be the full width at half maximum (FWHM) of the Gaussian curve for headgroup (d_H). According to a recent definition of bilayer thickness, total bilayer thickness is calculated as ($d_B = d_H + d_{HH}$) which also includes the first hydration shell of water. From total

bilayer thickness, the thickness of the hydrocarbon core d_C is derived as $(d_C = d_H - d_{HH})$ (161, 162).

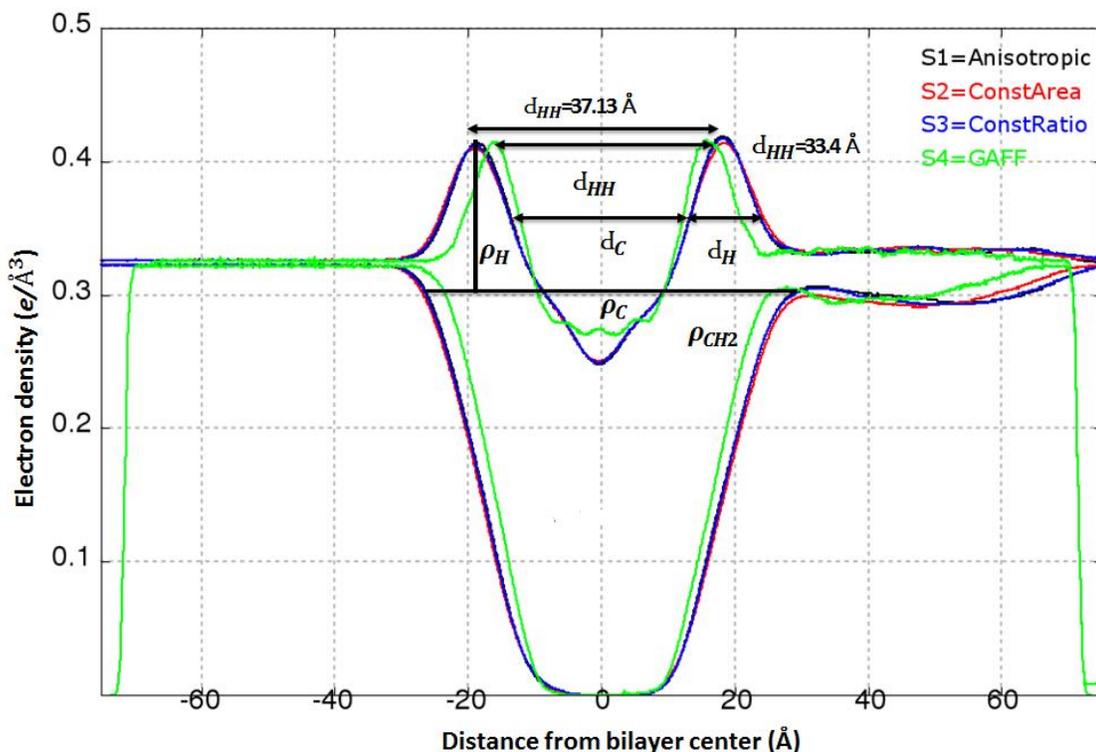


Figure 4.21: Electron density profile comparison between LIPID14 and GAFF lipid force-field calculated for total system of CYP2C9, which include protein-membrane and water system. All three pressure control methods gave same electron density profile. Electron density of GAFF lipid (green) shows decreased thickness (d_{HH}) and distance in the z -dimension. Asymmetry in the z -axis electron distribution is due to protein globular domain interaction on the +ve z -axis of bilayer.

The electron density profile from LIPID14 and GAFF lipid simulations differ in shape. All simulations with LIPID14 show similar bilayer shape and the bilayer thickness calculated from d_{HH} is 37.1 Å which matches with previous simulations (36.9 ± 0.6 Å) and experiments (37 Å). The thickness of POPC in GAFF lipid simulations decreased to $d_{HH} = 33.4$ Å and showed a broad trough in the tail region, indicating that an interdigitated bilayer is formed. Similar effect is observed by Small- and wide-angle X-ray scattering (SWAXS) experiments of multifunctional human peptide LL-37 which causes membrane disruption (163) by various mechanisms, depending on the membrane composition. The electron density profile in presence of LL-37 peptide

showed decreased membrane thickness, quasi-interdigitated phase interdigitation in the lipid bilayer and formation of ill-defined trough in the center of bilayer.

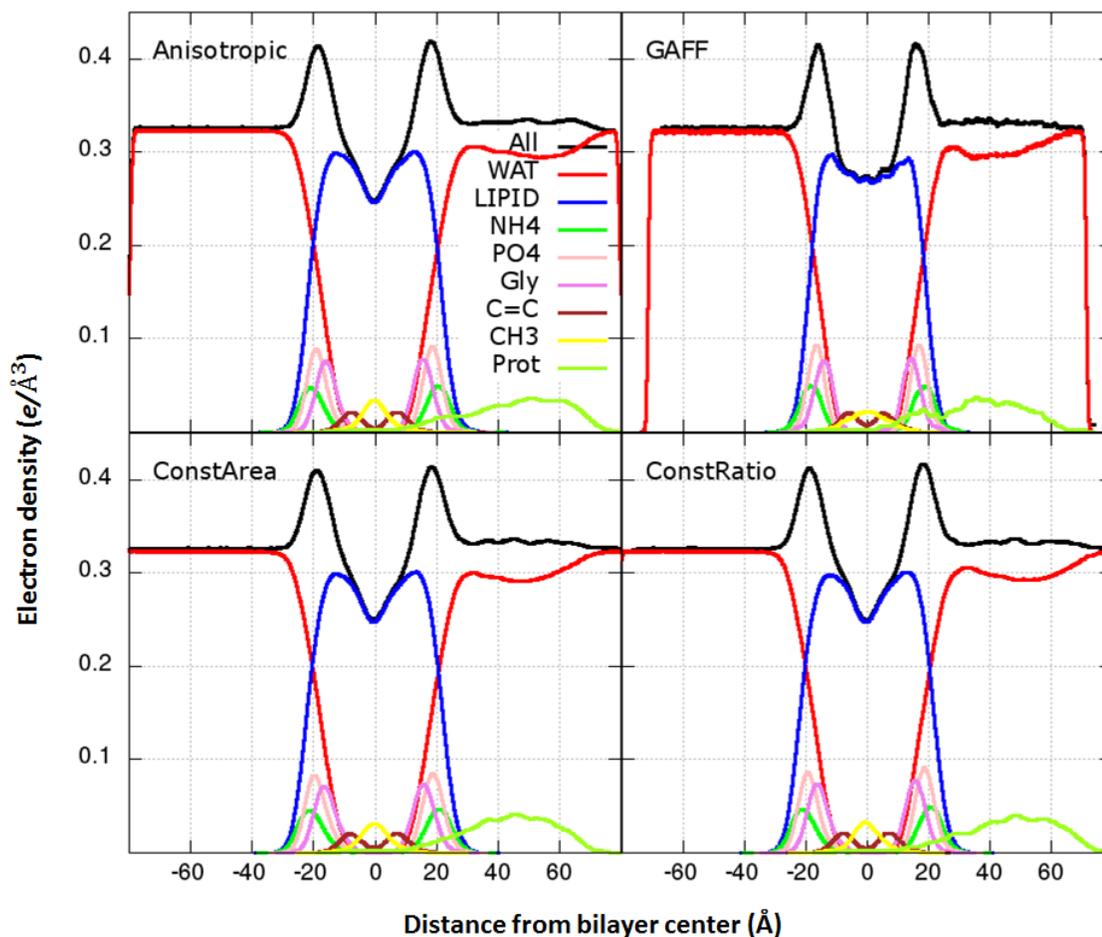


Figure 4.22: Electron density profile. Comparison between LIPID14 and GAFF lipid force-field for CYP2C9-POPC membrane system. All three pressure control methods for LIPID14 are compared with GAFF lipid surface tension. The total electron density is shown in black which is further decomposed in to different segments of membrane bilayer, water and protein. A similar shape of trough is seen in all LIPID14 simulations, while for GAFF lipid, a broader trough is visible.

The electron density profile for total system and decomposed electron density into water, lipid, phosphate, choline, glycerol, methylene and terminal methyl groups is shown together with density of protein (Figure 4.22). From protein density graph, the degree of insertion of the protein in the membrane can be seen. As seen also from decreased CoM distance of protein globular domain to membrane CoM, protein density

can be seen slightly higher in the membrane region in GAFF lipid simulations (Figure 4.22).

4.4.9 Order parameters

The 2-oleoyl-1-palmitoyl-sn-glycerol-3-phosphocholine (POPC) membrane contains two fatty acyl chains. Palmitoyl is a 16-carbon chain which is ester linked to glycerol C1 and named as Sn1 chain. The Sn2 chain is an unsaturated oleoyl (cis-9) ester linked to C2 of glycerol. The unsaturated Sn2 develops a kink and therefore results in less tightly packed hydrocarbon chains, which decreases melting temperature for transition from gel (solid ordered phase) to liquid crystalline. The phase transition to liquid crystalline results in liquid ordered phase (L_a) in the presence of cholesterol or liquid disordered (L_d) phase(6). The phase transition is important for studying membrane-protein interactions and biological processes that take place in the membrane. In the gel phase, all hydrocarbons occur in trans conformations and are highly ordered. The membrane in the liquid crystalline phase is highly dynamic and has a higher degree of freedom of motion resulting in the rotation around chemical bonds, trans/gauche isomerization, rotation around the axis and lateral diffusion(164).

Experimental techniques such nuclear magnetic resonance (NMR) and electron spin resonance have been widely used to study the dynamics of membranes. Lipid order parameters such as deuterium order parameters are determined by NMR (S_{CD}) and can be compared with MD simulation order parameters. Order parameters are used to measure spatial motion (orientational flexibility) around C-H (C-D) bonds or the C-C vector and indicates the degree of disorder in the hydrophobic tail. Higher order parameter values are observed in the first 6 carbons in the acyl chain starting from 0.2 in Sn1 and near 0.1 in Sn2, which gradually decrease to a minimum at the terminus. A decreased order parameter indicates high mobility (flexibility) in the hydrocarbon chains which increases along the lipid tail and reaches a maximum at the termini. Order parameters calculated for LIPID14 simulations match well with experimental order parameters [166], [167] and simulation results for both Sn1 and Sn2 chains (136) and

do not differ with different pressure control methods (Figure 4.23). GAFF lipid order parameters are higher in Sn1 saturated palmitoyl chain as well the Sn2 unsaturated oleoyl chain, suggesting a rigid gel like configuration of GAFF lipids. Higher order parameters in GAFF lipids were also observed in an earlier simulation Jójárt et al (167).

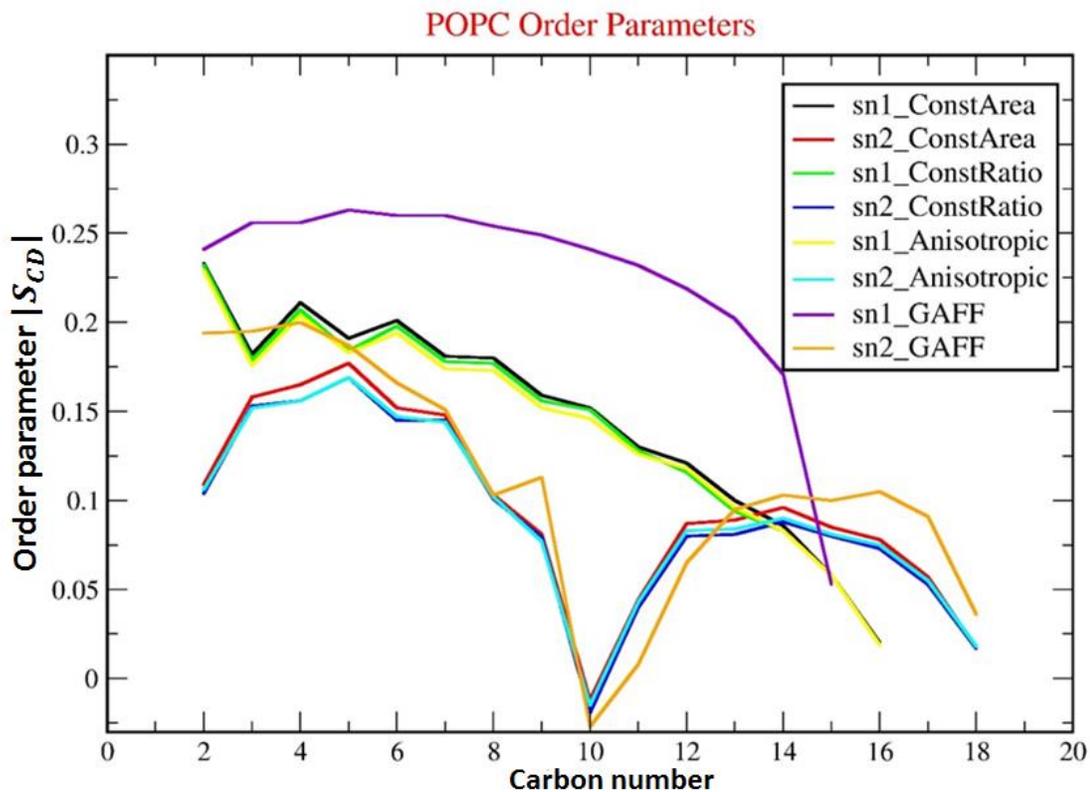


Figure 4.23: Deuterium order parameters compared between LIPID14 and GAFF lipid in CYP2C9-POPC membrane system. The time average order parameters are shown for Sn1 and Sn2 chain and different simulations methods.

4.4.10 Comparison with experiments and previous simulations

There is no experimental data characterizing full length CYPs and their interaction with the membrane in atomic detail. However, various experiments have been performed to study the membrane topology of CYPs and their interactions beyond the N-terminal transmembrane helix. Engineered CYP2C9 without an N-terminal helix remained membrane associated through catalytic domain as seen by atomic force microscopy

(AFM) (168). Experimental studies on different CYPs have reported the orientation in the membrane (116) from the heme-tilt angle which varies between 38 and 78°. The height of the catalytic domain above the membrane is reported as 35±9 Å using atomic force microscopy (169). However, based on AFM study, two possible models of CYP-membrane interactions are suggested (Figure 4.24). In model 1, the tip of the catalytic domain was proposed to interact with the membrane keeping the heme perpendicular to the membrane. In model 2, the distal side lies on the surface of the membrane while the heme remains parallel to the membrane. The height in model1 is suggested to be between 40 to 50 Å. Authors suggested model 1 to be closer to reality which is also consistent with site directed antibody-antipeptide studies and the surface hydrophobicity pattern in crystal structure of mammalian CYP2C5(27, 170), the first mammalian CYP to have structure determined. In our current study of CYP2C9 and 2C19, the membrane orientations and degree of insertions match well with model 1. However, the two proteins have distinct orientations above the membrane and form different heme-tilt angles.

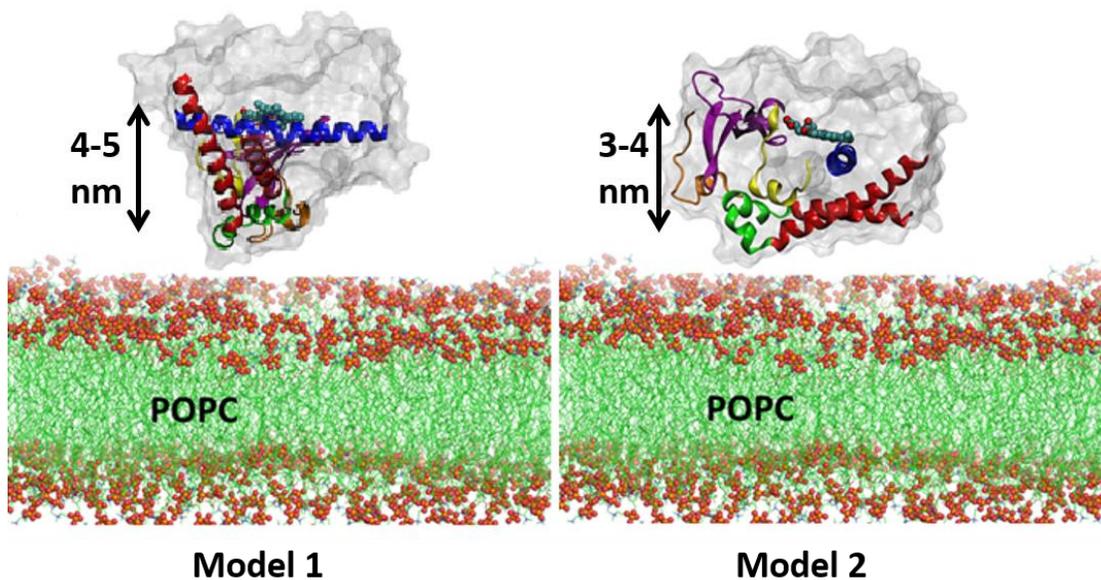


Figure 4.24: Orientations of CYP in the membrane suggested from AFM study (171). The CYP globular domain shown in white surface representation. The substrate recognition regions are shown in cartoon with different colors for different regions. Heme is shown in cyan color in VDW representation POPC lipid bilayer in green lines and phosphate headgroups are shown in red spheres

Insertion depth of catalytic domain in the membrane was also studied by tryptophan fluorescence scanning of CYP2C2, which suggested that residues L36 and F69 flanking A-helix and L380 located at beta strand 2 are inserted in the membrane while residue Y225 in the F'-G' helices remains outside (30). Primary sequence analysis of CYP2C2, CYP2C9 and CYP2C19 showed that CYP2C2 residues identified by tryptophan fluorescence scanning are conserved in all three CYPs. I observed similar interactions in CYP2C9 simulations where residues 36 and 380 shows interactions with the lipid tail region (100% occupancy), while residue 69 interacts with tail region for 40% of the simulation time. In CYP2C19, residue 36 is buried in the membrane while residue 69 shows interactions with the membrane tail region for 49% of the simulation time. Due to the orientation difference between the two proteins, beta strand 2 does not interact with membrane in CYP2C19 and therefore residue 380 remains outside the membrane. In both CYPs, F'-G' helices form strong anchoring point and residue 225 remains buried in the membrane. Furthermore, the linker re-orientation in two CYPs exposed the polar/charged residues on the cytoplasmic surface of endoplasmic reticulum, which agrees with site-specific antibody studies (peptide 2 region) ((170)). Linker region consists of a patch of polar to charged residues (20-30), a hydrophobic patch of proline rich region (30-40) and a patch of charged residues (40-50). Linker orientation and interactions in two CYPs, distribution of amino acids in the lipid bilayer and their propensity to reside in lipid head or tail region match well with the hydrophobicity scale for amino acids determined by various experiments and MD simulation studies (172) (173). During simulations, linker remains highly mobile and changes conformation to keep polar residues outside the membrane, in situation where polar/charged residues were buried in lipid hydrophobic core.

Table 4.10: Heme-tilt angle comparison with previously available data and current simulations

Source	PDB	Force-field	References	Heme tilt (°)
CYP2C9				
MD simulation	1R90	LIPID14	Current study	35±5-44±7
		GAFF lipid		25±5
	1R90	GAFF lipid	[31]	44±4 (1R901)
		GAFF lipid		41±4 (1R902)
	1OG5	Berger	(32)	55±5
		Berger	(174)	61±4
OPM database	1R90	OPM	(121)	59.8
	1OG5			71.9
CYP2C19				
MD simulation	4GQS	LIPID14	Current study	50±6-60±6
		GAFF lipid		46±1
OPM database	4GQS	OPM	(121)	74.0

Using multiscale simulation approach enables us to study CYP2C19-membrane interactions and orientations, starting with several different configurations. There is no reported experimental study focusing CYP2C19-membrane orientations. Therefore, comparison of CYP2C19 is done with CYP2C9 simulations. The orientations of CYP2C9 in membrane in current study using LIPID14 force-field match well with previously published work on CYP2C9 (31). CYP2C9 membrane-protein simulation studies by Berka et al (2001) have suggested different orientations and degree of insertion (Table 4.10). The authors have used a different structure, different procedure to model and generate

starting orientations. The Berger united atom force-field for lipids was used which could also contribute in different membrane-protein interactions and orientations. The orientation of CYPs in membrane has been reported in the OPM (orientation of proteins in membranes) database. The heme-tilt angle of CYP2C19 reported in OPM data base is close to the heme-tilt angle observed in MD simulations.

4.5 Conclusion

The two isoforms of CYP2C subfamily, CYP2C9 and CYP2C19, exhibit ~94% sequence identity, yet they show distinct substrate specificity. Since mammalian CYPs are anchored in the ER membrane by a N-terminal helix and secondary contacts from the catalytic domain, differences in the sequence and 3D structure in the membrane-interacting region in the catalytic domain can lead to different membrane protein interactions. As it has been hypothesized that lipophilic substrates enter into the binding pockets of CYPs from the membrane core, determining the orientation of CYPs in the membrane can provide insights into differences in substrate specificity. Here, I have used extended multiscale simulation methodology to understand the differences in primary sequence and 3D structure of two CYPs and their impact on the interactions and orientations in the membrane, which can elucidate the mechanism of drug selectivity and opening of ligand entrance tunnels through the membrane.

Several CG and AA simulations show consistency in the results and maintain the difference in the orientations of the two CYPs above the membrane. Based on angles calculated from simulation trajectories, orientations formed by two CYPs were classified into class A, B or A/B (intermediate orientation). CYP2C9 adopts mainly class A orientation, which has lower alpha, beta and heme-tilt angles, and CYP2C19 form class B orientation. The difference in the sequence and in conformations in the substrate recognition regions (SRS), near the membrane interface, resulted in different orientation and insertion depths in the membrane in two CYPs. Mutation of the key residues different at the membrane interface in CYP2C9 resulted in the similar orientations (50% of simulation results) to the wildtype CYP2C19 orientation. Mutation

also affected the insertion depth of linker and F'-G' regions, which was much lower than the wildtype CYP2C9. Whereas, similar mutations did not affect orientation of CYP2C19. However, an intermediate orientation (A/B) was formed by CYP2C9, when changing the starting structure, modeled or mutant. Whereas, in AA simulation using GAFF lipid force-field, similar orientation (A/B) was formed by CYP2C19, with lower angles compared to CYP2C19 orientations (B) observed in LIPID14 simulations. Therefore, I conclude that mutating key residues different in the linker, beta strand1 and F'-G' loop, changing protein conformations in these regions or use of different force-fields can have direct influence on the final orientation and interactions of CYP membrane-systems.

CHAPTER 5: INFLUENCE OF THE TRANSMEMBRANE HELIX-ANCHOR ON CYTOCHROME P450 17A1 MEMBRANE INTERACTIONS

5.1 Abstract

Mammalian CYPs are anchored in the ER membrane by an N-terminal transmembrane (TM) helix that is connected to the globular domain by a flexible linker sequence. The structural and functional importance of the TM-helix is unclear since it has been shown that the crystal structure CYPs are truncated to remove the TM-helix and the flexible linker associate with the membrane, still retain their enzymatic activity. Furthermore, modification of the N-terminal amino acids sequence of mammalian CYPs is found to increase the expression in bacteria. It has been reported that the modified construct of CYPs, like truncated N-terminal TM-helix, retained membrane association as well as enzymatic activity (175–177). In current study, which employs both coarse-grained and all-atom simulations, I investigate the effect of such modification of first 8 N-terminus residues of the TM-helix of CYP17 on the orientation and interactions of the globular domain with the membrane. Here I found that the heme-tilt angle of two CYPs mutant and wild-type (mtCYP17 and wtCYP17) remained same, having same orientation above the membrane. However, modified TM-helix sequence (mtCYP17), especially W2A and E3L, increased the likelihood of the TM-helix being pulled out of the membrane core which showed an amphipathic helix like characteristics, lying parallel to the membrane. In simulations, where modified TM-helix was buried in the membrane, it influenced on the linker position, which was trapped in the F'-G' loop region and formed polar contacts between linker, F-G region and C-terminal beta-sheet (residue 480-485). The trapped linker position in mtCYP17 could obstruct the tunnel 2d/2f which is opening through the membrane into binding pocket and may influence on the enzymatic activity.

5.2 Introduction

Mammalian CYPs are bitopic integral membrane proteins with a single N-terminal transmembrane (TM) helix spanning the endoplasmic reticulum (ER) membrane which is connected to the globular domain in the cytosol by a flexible linker. The structural and functional importance of the TM-helix is unclear since it has been shown that upon truncation of the TM helix and the flexible linker associated with the membrane, CYPs still retain their enzymatic activity (178). The experimental information on TM-helix interactions and orientation in the membrane, i.e., TM-helix tilt angle with respect to the membrane plane is scarce. To date, only full-length crystal structures of *Saccharomyces cerevisiae* lanosterol 14 α -demethylase of the CYP51 family, has been resolved (5). It consists of an amphiphatic helix (residues 1-25), the TM-helix (residues 27-50) and the globular domain connected to the TM-helix by a flexible linker. The TM-helix is connected to the amphiphatic helix via a short linker region (residues 24-26) containing a proline at position 26 (P26), which makes a kink at the beginning of the TM-helix. The TM-helix, which is 24 residues long (37.5 Å), has a pronounced tilt angle of 55° with respect to the membrane normal (5). The primary sequence and length of the TM-helix vary in different CYPs. Depending on the length of the hydrophobic sequence of the TM-helix, the distribution of amino acids, and the hydrophobic thickness of the membrane, the TM-helix can adopt different orientations and different tilt angles, or can develop kinks to avoid hydrophobic mismatch(9).

A solid state nuclear magnetic resonance (NMR) spectroscopy study of full length rabbit CYP2B4 and cytochrome-b5 showed a TM-helix tilt angle of 17° for cytochrome P450 (179) and a TM-helix tilt angle of 13° has been observed for NADPH-cytochrome P450 oxidoreductase (CYPOR)(180). The tilt angle for cytochrome oxidases (COX) with TM helix consisting of 20 residues is between 25-30° and between 28-32° for 21 residues long TM-helix (181). In our AAMD simulations of CYP3A4 with TM helix length of 24 residues (residues 3-26), the TM tilt angle was 48.9 \pm 3.9° which formed a kink on the N-terminal side (see Chapter 3 and Table 3.3 and Figure 3.3) (111). The higher TM tilt angle

observed in CYP3A4 is consistent with the TM tilt angle calculated for full length crystal structure of CYP51 (PDB 4LXJ). On the other hand, CYP2C9 and CYP2C19, with shorter TM-helix length of 19 residues, showed lower tilt angles of $11\pm 6^\circ$ - $25\pm 5^\circ$ and $14\pm 3^\circ$ - $20\pm 9^\circ$, respectively (see Chapter 4 and Table 4.8). In a molecular dynamics study of helix A of bacteriorhodopsin, different tilt and kink angle distributions were observed depending on the system size (POPC 84 vs 128 molecules)(182). An AAMD study of 25 TM-helices with various hydrophobic lengths (tryptophan-alanine-leucine repeats) and numbers of flanking tryptophan residues revealed that shorter TM-helices have smaller tilt angles (10 - 12°) due to negative hydrophobic mismatch, while larger peptide tilting (31 - 41°) is observed in the case of longer TM helices, due to a positive hydrophobic mismatch (157).

The TM-helix of CYPs generally contains polar, aromatic or charged residues on either side of the transmembrane domain flanking the hydrophobic residues. They are important for TM-helix stabilization in the bilayer, protein sorting and retention of CYPs in the ER membrane (183). It has been observed that retention of the TM-helix of CYP2C1 in the ER membrane is dependent on the hydrophobicity of the transmembrane domain (residues 3-20) and hydrophilic residues (21-23 KQS) in the start of linker region (residues 21-28). Mutation of linker, hydrophobic transmembrane domain or both regions resulted in the expulsion of CYP2C1 from ER membrane and affected direct retrieval from the pre-Golgi compartment(183).

A modified construct of CYP17 (pCWH17mod) was used for experiments performed by Ivan Lenov and Stefan Sligar (unpublished results) to study CYP17 membrane insertion and orientation in a POPC nanodisc. This modified construct of CYP17 was originally prepared by Professor M. R. Waterman (Vanderbilt University, Nashville, TN, U.S.A.) by mutating 5 residues in the N-terminal TM helix region for expression in the *Escherichia coli* and addition of 4 histidine residues at C-terminus (residue 509-5012) for purification (176). The five substitutions introduced in the N-terminal TM-helix are highlighted in red: wtCYP17: 1M**WELV**ALL8 to mtCYP17: 1M**ALLLAVF**8. Multiple

sequence alignment of CYP17 showed conserved W2 residues in different species (see result and discussion section 5.4.3 and Figure 5.1). As seen in the sequence comparison, after polar and charged residues on the N-terminal end (2WE3), two polar aromatic residues Y14 and W17 are found preceding P18 on the C-terminus, which act as boundary residues. Furthermore, the C-terminus of the TM-helix contains a patch of positively charged residues (19KRR21) which favor interactions with the lipid headgroup.

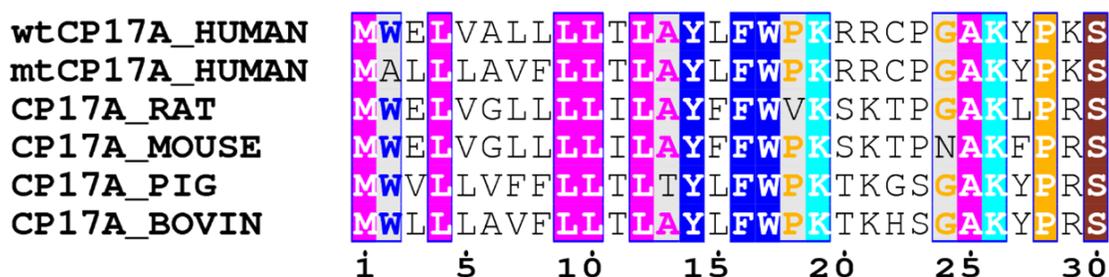


Figure 5.1: Multiple sequence alignment for first 1-30 residues, comparing human wtCYP17, mtCYP17, rat, mouse, pig and bovine sequences. The color codes differentiate the residues based on their physicochemical properties and conservation. The residues different in multiple sequences are shown with white background.

The substitutions of polar/charged residues (2WE3) on the N-terminus results in amphipathic helix-like characteristics, with hydrophobic residues at the N-terminal end and polar aromatic and charged residues at the C-terminus of the TM-helix. Amphiphatic helices, due to their segregation of hydrophobic and polar residues on two sides, are nonspanning and orient parallel to the membrane plane(184). In an amphiphatic helix, the central region interacts with membrane at the glycerol level, hydrophobic residues penetrate in the fatty acyl tail and polar residues remain at the lipid head-group region. With the replacement of the N-terminal residues of wtCYP17 TM helix, specially 2WE3 by 2AL3 (mtCYP17), the likelihood of mutant TM-helix to be pulled out of the membrane or to act as monotopic membrane helix or behave like an amphiphatic helix is increased. Therefore, I investigate the effect of mutation of the N-terminal polar/charged residues in the TM-helix in mtCYP17 on the globular domain interactions and orientation in the membrane and compared with wtCYP17-membrane

interactions. I also compare our results with the heme-tilt angle for mtCYP17 experimentally determined using linear dichroism and nanodisc method by Ivan Lenov and Stefan Sligar.

5.3 Material and methods

5.3.1 Preparation of all-atom models

The CYP17 X-ray crystal structure (PDB ID 3RUK) in complex with prostate cancer drug abiraterone was resolved after truncating the N-terminal residues (1 to 30)(185). The crystal structure has four protein copies (Chain A-D) in an asymmetric unit cell. First two chains A-B have further missing residues in the loop connecting helices H and I (residues 274-282 in chain A and 275-282 in chain B) and C-terminal residues 503-512. Chains C and D have missing N-terminal TM helix and part of linker region (residues 1-30) and C-terminal residues (503-512). Therefore, chain C was used after modeling missing TM-helix and linker region. For prediction of TM helix length, secondary structure prediction servers: PredictProtein (<https://www.predictprotein.org/>) (186) and PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) (187) were used. The TM helix length reported by Y.-L. Cui *et al.*(188) ranges from residues 1-19. Residues 3-19 were modeled as helical conformation and used the same initial helix length for both wtCYP17 and mtCYP17. The secondary structure prediction for TM helix length for wtCYP17 and mtCYP17 by various servers is discussed in the Results and discussion (section 5.4). After modeling the full-length structure, five different starting orientations of the globular domains wtCYP17 and mtCYP17 were generated which were used for preparation of coarse-grained models.

5.3.2 Preparation of coarse-grained systems and simulations

A similar procedure was used to generate wtCYP17 and mtCYP17 coarse-grained protein-membrane systems as described in Chapter 3 (111). For mtCYP17, three different CG simulation systems (S1-S3) were prepared varying in the length of flexible linker region as elastic network restraints were removed from the linker residues: 20-

29 (S1), 20-38 (S2) and 20-49 (S3). For wtCYP17 simulations, the linker residues 20-38 (S4) were kept flexible (Table 5.1). CG simulations (S5) using only globular domain, without TM and linker, were also performed to study the interactions and orientations of globular domain only with membrane. Additionally, TM helix self-assembly simulations were performed for both wildtype and mutant TM helices (residue 1-22). For self-assembly simulations, TM helix, POPC lipid and water molecules were randomly placed a box of size of 14 nm in each dimension (x, y, z). Elastic network restraints were applied on complete peptide (1-22 residues). Five self-assembly simulations were performed for wildtype and mutant TM helices.

Latest version of MARTINI force-field (MARTINI version 2.2) with standard water model (NPW) was used (78). The simulation procedure was same as described in Chapter 3. Before running CG simulations, short steepest descent energy minimization was followed by NPT equilibration with constant temperature of 310 K using a velocity rescale (v-rescale) thermostat. The temperature was coupled separately for the protein, the POPC and solvent by defining separate groups, with a coupling constant of 1 ps. Initial equilibration was performed with a Berendsen weak coupling scheme to maintain constant pressure with a coupling constant of 2.0 ps and a reference pressure of 1 bar. During production simulations, a Parrinello-Rahman barostat with a coupling constant of 12 ps was used. Semiisotropic pressure coupling was used with a compressibility of 3.0×10^{-5} . The time step was 20 fs. The non-bonded interactions were treated with RF (reaction field) and cut-off methods for calculations of Coulomb's and VDW interactions, respectively.

Table 5.1: CG simulations of mutant and wildtype CYP17 models and CYP17 globular domain only.

CG					
Simulation Identifier	Residues in regions				No of Simulations
	TM Helix	Linker region	Flexible linker	Globular domain	
mtCYP17					
S1	3-19	20-49	20-29	50-502	5
S2	3-19	20-49	20-38	50-502	5
S3	3-19	20-49	20-49	50-502	5
wtCYP17					
S4	3-19	20-49	20-38	50-502	5
S5	Globular Domain: 50-502				3
S6*	TM-helix self-assembly: residues 1-22				10

*Five TM helix self-assembly simulations for wtTM and five for mtTM helix

5.3.3 Trajectory analysis and back-conversion to all-atom models

The CG simulation trajectories (S1-S5) were analyzed for convergence or stable orientations of CYP in the membrane. For this purpose, the same angles and distances as mentioned in previous chapters (Chapter 3 and 4) were calculated (31, 111, 112). A representative frame was selected from each set of CG simulations and this was used for back-conversion to an all-atom model. The representative frame was selected to have angle and distance values within 1% of their mean value(112). The back-conversion procedure for the POPC bilayer was performed as described in Cojocaru *et*

al.(31). Protein back-conversion was done using the scripts `backward.py` and `initram.sh`, available at the MARTINI website (<http://cgmartini.nl>) (110). In previous studies, globular domain residues 50-490 (crystal structure) were superimposed on the back-mapped structure to preserve the side chain interactions within heme cofactor binding pocket. However, during CG simulations of mtCYP17, the position of the TM helix and linker region changed such that it developed strong interactions with the F'-G' loop, influencing the F'-G' loop conformation. Therefore, following our previous protocol with superimposition of the crystal structure on the back-converted structure resulted in intramolecular and intermolecular (between protein and membrane) steric clashes. The conformation of such regions showing close interactions with the membrane and within the protein were kept same as obtained from back-conversion of representative CG frame by `backward.py` script. These regions include the TM-helix, linker, A-helix, beta-strand1 (residues 1-76), F'-G' loop (residues 210-228), and beta-strand2 (residues 378-387). The conformation of the rest of the protein including heme binding pocket was extracted from the superimposed crystal structure.

5.3.4 All-atom molecular dynamics simulation

All-atom molecular dynamics (AAMD) simulations were performed using ff12SB and LIPID14 force field for protein residues (135), and POPC lipids(136), respectively. The detailed procedure for setting up a membrane protein and water system was described earlier in Chapter 3. With the LIPID14 force-field, two different pressure coupling methods i.e., anisotropic and semiisotropic (constant ratio) (Table 5.2) were used. Three AAMD simulations of mtCYP17 (SIM1-SIM3) were performed taking representative frame from CG simulation of mtCYP17 (S3). Two AA simulations of mtCYP17 (SIM1 and SIM2) have same starting velocities but different pressure coupling (anisotropic vs semiisotropic), whereas SIM3 is replica simulation with different starting velocities with semiisotropic pressure coupling (constant ratio). Three AAMD simulations of wtCYP17 (SIM4-SIM6) were performed with same starting structure but different pressure coupling method and different initial velocities. In SIM4 and SIM5,

same pressure coupling (anisotropic) but different velocities were used, while SIM6 is performed with same initial velocity as SIM5 but constant ratio instead of anisotropic pressure coupling.

TABLE 5.2. AAMD Simulations performed for CYP17.

AAMD Simulations	System	CG Simulation	Pressure coupling	Time (ns)
SIM1	mtCYP17	S3	Anisotropic	148
SIM2	mtCYP17	S3	Semiisotropic*	148
SIM3	mtCYP17	S3	Semiisotropic*	172
SIM4	wtCYP17	S4	Anisotropic	159
SIM5	wtCYP17	S4	Anisotropic	124
SIM6	wtCYP17	S4	Semiisotropic*	145

* semiisotropic pressure coupling with constant ratio

5.4 Results and discussion

5.4.1 Transmembrane helix membrane interactions

The MARTINI CG simulations were used to obtain the initial CYP-membrane interactions and orientations of the globular domain above the membrane. Several different starting orientations of globular domain were selected as mentioned in the Methods section 5.3. CG simulations of mtCYP17 (S1-S3) were performed which differ from the wtCYP17 (S4) in the first 8 residues in the N-terminal end of TM-helix. However, same length of the N-terminal helix (residues 3 to 19) was modeled for both mtCYP17 and wtCYP17 (Figure 5.2). Different secondary structure prediction servers were used, which predicted different TM helix lengths. For example: PredictProtein server suggested 20 residues long (residues 2-21) N-terminal α -helical conformation for wtCYP17 sequence, TM helix domain range from 4-17 residues, and buried region from 0-18 residues. For

mtCYP17 sequence, N-terminal α -helix length was predicted from residues 4-20 (17 residues), TM-helix region from 6-18 and buried region from 0-18 residues. The PSIPRED server suggested 16 residues long N-terminal α -helix (residue 2-17). Different TM helix lengths were suggested by different servers which also vary at different times of predictions. According to a recent prediction by PredictProtein, the N-terminal helical conformation is predicted from residues 2-20 for both wtCYP17 and mtCYP17, whereas, TM-helix residues 6-19 is only predicted for wtCYP17 but not for mtCYP17. However, when only first 1 to 60 residues of sequence of mtCYP17 is used, TM helix is predicted which range from residues 3-19. Furthermore, an extra TM-helix (residues 29-43) is predicted in the latter case. Although the initial TM helix length was modeled for residues 3-19 before running CG simulations, back-conversion of representative frames from CG to AA models resulted in the TM-helix length of 3-17 in both mtCYP17 and wtCYP17 models.

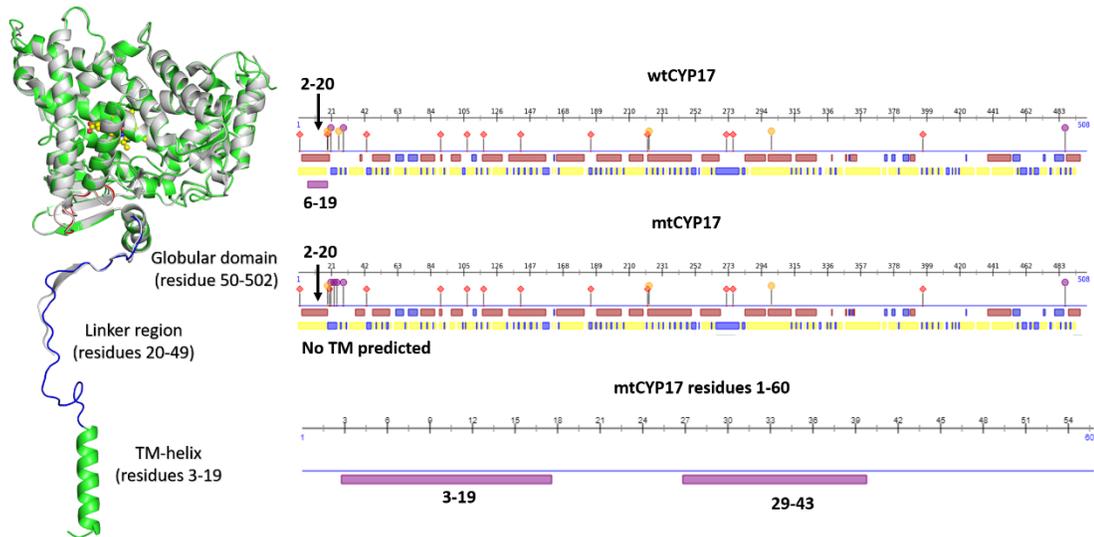


Figure 5.2: Left: Full length modeled structure of CYP17 superimposed on the x-ray crystal structure shown in grey (PDB 3RUK chain C). The cartoon representation of the secondary structures color coded by green alpha helices, linker in blue and F-G loop in red for modeled structure. Right: TM-helix length predicted for wildtype (top) and mutant (middle full length sequence and bottom only residues 1-60) are shown in the thick magenta line, predicted by PredictProtein server dated 06-22-2017 (186).

5.4.2 Convergence of coarse-grained simulations

The effects of mutations on the final orientation and interactions of the globular domain of CYP17 with the membrane were analyzed by comparing the full-length wtCYP17, mtCYP17 and globular domain only simulations (S1- S5). For mtCYP17 CG systems (S1- S3), different flexible lengths were used to check the effect on convergence and the orientation of the globular domain above the membrane. Only one flexible linker length (residue 20-38) was used for wtCYP17 CG-system S4.

5.4.2.1 mtCYP17 CG system: S1:

In mtCYP17 CG system S1, five different starting orientations (1-5) with the flexible linker residues 20-29 were used. In three out of five simulations (3-5), the globular domain did not develop stable interactions with the membrane (Figure 5.3). As shown in angles and distances graphs, mtCYP17 (3-5) remained fluctuating above the membrane surface and could not form stable orientation during 15 μ s of simulation time. In two CG simulations (3 and 5), the TM helix was pulled out of the membrane after ~ 6 μ s simulation time. The higher TM tilt angle value ($\sim 80^\circ$) corresponds to parallel positioning of the TM helix to the membrane plane (Figure 5.3). The pulling out of TM helix from membrane core disrupted globular domain interactions and orientations above the membrane, leading to higher CoM distances and unstable angles.

Although in simulation 4 (Figure 5.3 and S1 and row 4), the TM helix remained inside the membrane, the F'-G' loop formed unstable interactions with the membrane. The CoM distance of F'-G' loop to the membrane CoM showed continuous fluctuations. Thus, weaker interactions of F'-G' loop with the membrane resulted in the overall unstable orientation of the globular domain. As seen in previous studies of CYP2C9 and CYP2C19 membrane interactions, after TM helix and the linker region, secondary interactions are mainly developed by F'-G' helices/loop which stabilized the final orientation and CYP-membrane interactions (Chapter 4).

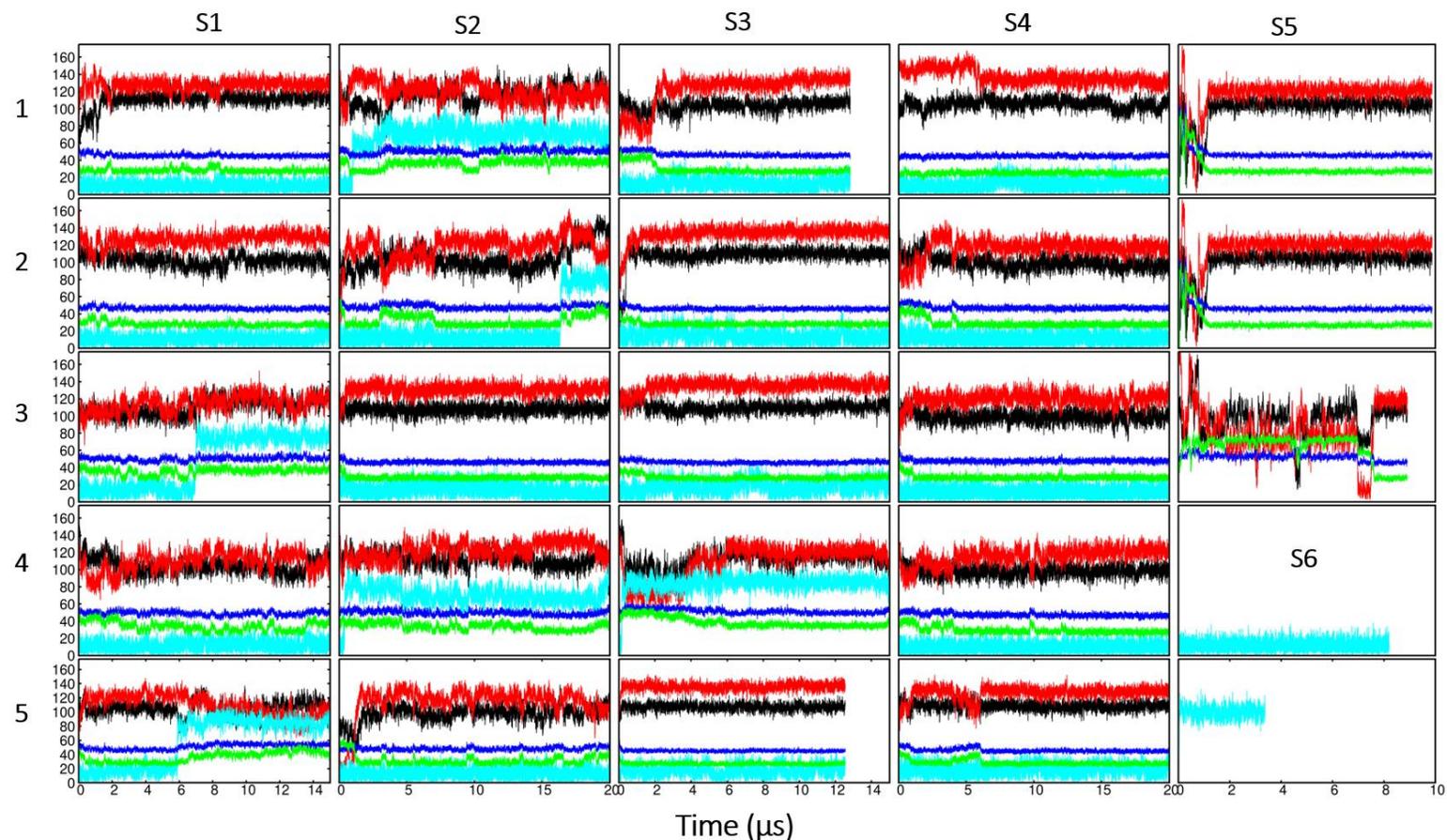


Figure 5.3: Evaluation of protein position with respect to the bilayer during CG simulations. Angle and Distance graphs for CG simulation systems (S1-S5). shown in columns 1-5 from left to right. X-axis shows time in μs and on Y-axis are shown angles ($^\circ$) and distances (\AA). Full length protein simulation systems are shown for mtCYP17 (S1-S3), wtCYP17 (S4) and globular domain only CG system (S5) and rows (1-3). Five initial orientations shown in rows 1-5 and TM helix only system S6 is shown in column 5 rows 4-5. Color scheme: Angles ($^\circ$) alpha, beta and TM tilt angle are shown in black, red and cyan, respectively. Distances (\AA) CoM of F'-G' loop (green) and globular domain (blue) to the membrane CoM are shown.

5.4.2.2 mtCYP17 CG system: S2

In mtCYP17 CG S2, flexible linker was set from residues 20-38. The starting orientations (1-5) were same as in S1 simulations. Due to unstable orientation of TM-helix and globular domain in the membrane CG simulations (S2) were extended to 20 microseconds (Figure 5.3). Fluctuations in the TM tilt angle and CoM distance of F'-G' loop resulted in the unstable globular domain orientation (1,2,4 and 5). In 3 out of 5 CG simulations (1,2 and 4), the TM-helix drifted out of the membrane and adopted an orientation parallel to the membrane plane, giving rise to a high TM tilt angle value ($\sim 80^\circ$). Only one simulation (3) converged to the final orientation, where globular domain interactions with the membrane remained stable and no fluctuations in angles and distances were observed. The angles and distance values matched with mtCYP17 (S1:1-2) simulations.

5.4.2.3 mtCYP17 CG system: S3

In mtCYP17 CG simulation S3, all residues in the linker (residues 20-49) were treated as flexible. All simulations (1-3 and 5) except 4, converged to same orientation and remained stable for a simulation time of 12 microseconds (Figure 5.3). In simulation 4, the TM-helix drifted out of the membrane core and remained parallel to the membrane plane at interface, as seen in previous earlier simulations (S1-S2). The converged orientations in mtCYP17 (S3) vary from mtCYP17 and wtCYP17 simulations, as higher alpha ($108 \pm 9^\circ$) and beta angles ($134 \pm 6^\circ$) are seen in S3 simulations than S1-S2 or S4 simulations (Table 5.3).

In all mtCYP17 simulations, in which the TM helix went out of the membrane core, the globular domain interactions with the membrane were disrupted. As shown in the mean and standard deviation table (Table 5.3) and angle and distance graphs (Figure 5.3), the CoM distance of F'-G' or CoM of protein globular domain to the membrane CoM, remained higher, resulting in unstable secondary contacts with the membrane. The secondary interactions by globular domain, especially by F'-G' loop, are crucial for maintaining stable orientation of CYPs in the membrane.

5.4.2.4 wtCYP17 CG system: S4

A set of five CG simulations of wtCYP17 (S4) with flexible linker (residue 20-38), same as for mtCYP17 CG simulation (S2), were performed. In all 5 wtCYP17 simulations (1-5), the globular domain converged to the same orientation and the TM helix remained embedded in the membrane (Figure 5.3). There were no large fluctuations in angles or distances or expulsion of TM-helix observed in wtCYP17 simulations, which were extended to 20 microseconds. The presence of polar/charged residues (2WE3) on the N-terminus stabilized the TM-helix interactions and orientation in the membrane and converged orientations in all 5 CG simulations. The orientations of wtCYP17 match with the converged orientations of mtCYP17 simulations (S1 and S2) but differ from mtCYP17 (S3) orientations.

5.4.2.5 Globular domain simulation: S5

The mean and standard deviation values for angles and distances are calculated for all CG simulations and given in (Table 5.3), which showed difference in the angle and distance values in S3 compared to S4 system. The differences in orientations between mtCYP17 and wtCYP17 could be due to fully flexible linker and unstable TM helix in mtCYP17. Therefore, three separate (1-3) globular domain only (residues 48-502) simulations (S5) were performed to study the interactions and orientations of the globular domain in the membrane independent of TM and linker interactions (Figure 5.4). The angle and distance values of globular domain were compared with the full-length simulation results (S1-S4). As shown in the angle and distance graphs (Figure 5.3 and column S5 rows 1-3) and Figure 5.4, the globular domain adopts various conformations and explores a wide conformational space, before F'-G' loop (green line in graph) distance CoM to the membrane CoM is decreased. The F'-G' loop interactions stabilized the orientation of the globular domain in the membrane, and no further changes in the orientations/conformation were observed (Figure 5.4). The final orientations of the globular domain in the membrane were stable. The mean and standard deviation values of angles alpha and beta, $105\pm 5.3^\circ$ and $123\pm 6^\circ$, in the S5

simulations, match well with the wtCYP17 simulations (S4) angles, alpha $101\pm 7^\circ$ and beta $125\pm 7^\circ$ and also with the mtCYP17 CG simulations (S1-S2) with TM in the membrane. The beta angle in mtCYP17 simulation (S3) is 10° higher compared to the wtCYP17 and globular domain. The mtCYP17 simulations (S1-S2), in which TM remained inside the membrane, showed similar orientations of the globular domain in the membrane as wtCYP17 and globular domain only simulations.

5.4.2.6 Transmembrane helix self-assembly simulations: S6

Five separate self-assembly simulations of mtCYP17 and wtCYP17 TM helices (residues 1-22) in POPC membrane (594 molecules) and MARTINI standard water model (NPW) were performed (TM1-5), to investigate the orientation of TM helix in the membrane. In one out of 5 wtCYP17 TM helix simulations (TM1-5), TM helix spanned the membrane like bitopic membrane-protein, and formed the TM tilt angle $12\pm 6^\circ$ which match with the TM tilt angle formed by full length wtCYP17 TM helix ($13\pm 6^\circ$) (Figure 5.5). In the rest of self-assembly simulations, wtCY19 and mtCYP17 TM helix could not span the bilayer or left the membrane and lay parallel to the membrane plane, as in full length mtCYP17, where the TM helix went out of the hydrophobic core of the membrane (Figure 5.5). The self-assembly simulation of wtCYP17 TM helix (TM1), in which the TM helix spanned membrane, was extended to 8 microseconds to observe any conformational change in the TM helix. As shown in the angles and distance graphs, column 5 and row 4, the TM helix remained embedded in the lipid bilayer, parallel to the membrane z-axis (bilayer normal), throughout the simulation time.

Table 5.3: Mean and standard deviations for angles and distances characterizing positioning of the protein in the phospholipid bilayer calculated for full length wtCYP17 vs mtCYP17 and globular domain of CYP17 CG simulations.

CG system	CG simulations	Angles (°)			Distances CoM (Å)			TM position
		Alpha	Beta	TM tilt (3-19)	Linker (20-49)	F'-G' loop (210-227)	Globular domain (50-502)	
mtCYP17								
S1	1,2,4	105±9	123±10	11±6	21±2	29±4	47±3	In
S1	3,5	111±10	110±13	78±6	27±2	39±4	53±3	Out
S2	3,5	104±8	125±11	14±9	22±2	29±4	47±3	In
S2	1,2,4	112±8	122±11	71±9	25±3	34±6	49±3	out
S3	1-3,5	108±9	134±6	15±8	22±2.1	28±2	46±2	In
S3	4	115±7	120±8	85±6	26±2	36±2	50±2	Out
wtCYP17								
S4	1-5	101±7.3	125±7	13±7	20±2	27±4	46±2	In
S5	1-3	105±5.3	123±6	--	--	27±2	46±2	--
S6	TM1	--	--	12±6				In

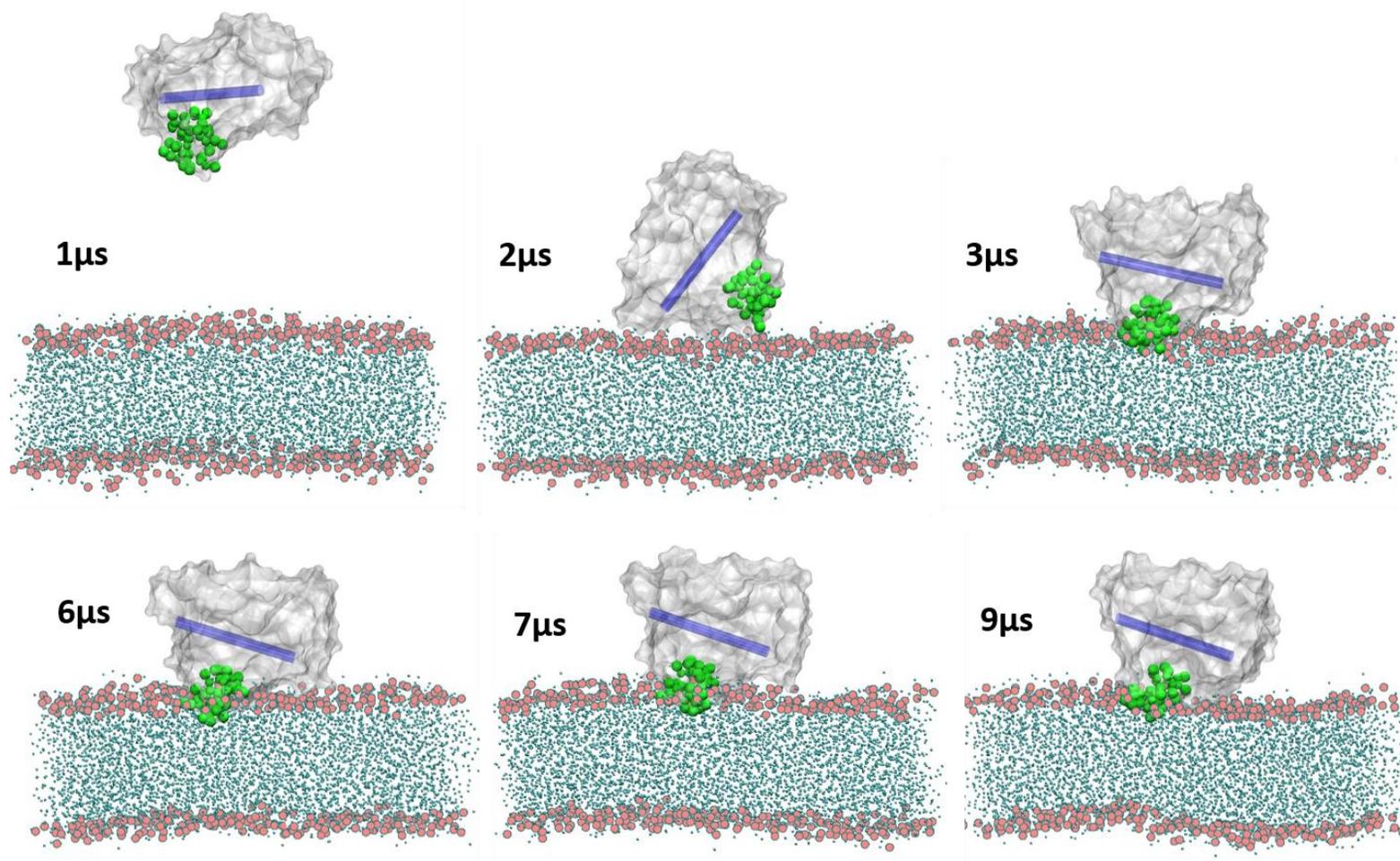


Figure 5.4: Snapshots at 4 different time steps showing different orientations of the globular domain, taken from S5:1 simulations. The globular domain is shown in a silver surface representation, the F'-G' loop in green VDW spheres, the I-Helix shown by blue cylinder. The phosphate atoms in the lipid-headgroup are shown by red spheres. The POPC membrane is shown in cyan color. The orientation of the globular domain remained stable after 3microsecond until the end of the simulation (9microsecond).

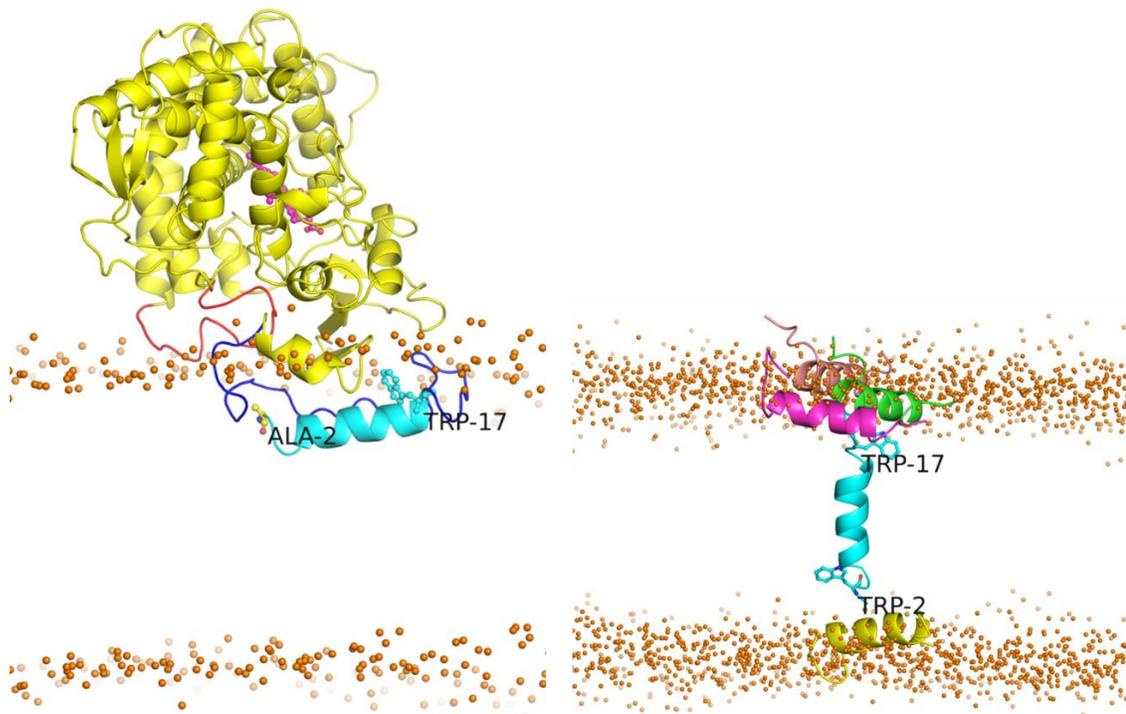


Figure 5.5 Position of TM-helices in the full-length mtCYP17 simulation (S1:3) on the left, compared with wtTM helix self-assembly simulations (S6: 1-5) shown in the right panel, after back-conversion of last frame from CG to AA model. In the full length mtCYP17, the TM helix is shown in cyan, and the linker (residues 18-49) and F-G loop (residues 210-228) are shown in blue and red, respectively. All five positions of wtCYP17 TM helices obtained from self-assembly simulations are shown relative to the phosphate atoms (orange spheres), representing lipid head-group region. The anchoring residues in the wtCYP17 and mtCYP17 TM helices are shown in stick representation.

5.4.3 Difference in the orientation and position of wtTM vs mtTM helix:

Compared to the TM-helices in CYP2C9 and CYP2C19 (19 residues) and CYP3A4 (24 residues), CYP17 has a short TM helix (17 residues) and longer linker region (residues 20-49), connecting the globular domain (residues 50-502) with the TM helix. As discussed in the introduction (section CHAPTER 5:), the shorter peptides or TM helices compensate the hydrophobic mismatch by forming aggregates, undergoing conformational change in the backbone residues or aligning parallel to the z-axis to avoid insertion of any charged residues in the hydrophobic core. Therefore, a lower TM tilt angle is observed for shorter TM helices (154). Due to negative mismatch in short TM helix, expulsion of the TM-helix from hydrophobic core could be seen (Figure 5.5).

Although short in length, the wtCYP17 TM-helix is flanked by polar aromatic and charged residues at both ends of the TM-helix: W2 and E3 on the N-terminal end and W17 followed by patch of positively charged residues 19KRR21 on the C-terminal end (Figure 5.1), which result in stable wtCYP17 TM helix interactions and orientation in the membrane.

Multiple sequence alignment of human wtCYP17 and mtCYP17 of first 1-30 residues was performed to check residue conservation in the region missing in the crystal structure (PDB 3RUK). For this CYP17 sequences from rat, mouse, pig and bovine (cow) were compared with the human CYP17 sequence. They have 63-70% sequence identity with human wildtype sequence. As shown in the sequence alignment in Figure 5.1, aromatic residues like W2, Y14, F16 and W17 are conserved in all species. Tryptophan (W) and Tyrosine (Y) residues in membrane proteins act as anchoring residues and are located between the polar headgroup and hydrophobic core of the lipid bilayer, most probably in the glycerol region of the lipid-bilayer (9). The presence of W2, W17 and charged residues on either side of the TM-helix of wtCYP17 keep the balance and stabilize the orientation of TM-helix in the membrane. Thus, wtCYP17 TM-helix showed bitopic characteristics during CG simulations (S4) and in self-assembly simulation of TM helix only (S6), which was fully spanning the hydrophobic core of the membrane. In mtCYP17, substitution of polar and charged residues 2WE3 by 2AL3 disrupted the balance and resulted in formation of amphiphatic helix characteristics which resulted in pulling out of the TM-helix from the membrane core, as discussed earlier (section 5.4.2). Trajectories where the TM-helix was unstable and left the membrane core, also showed unstable globular domain interactions with the membrane.

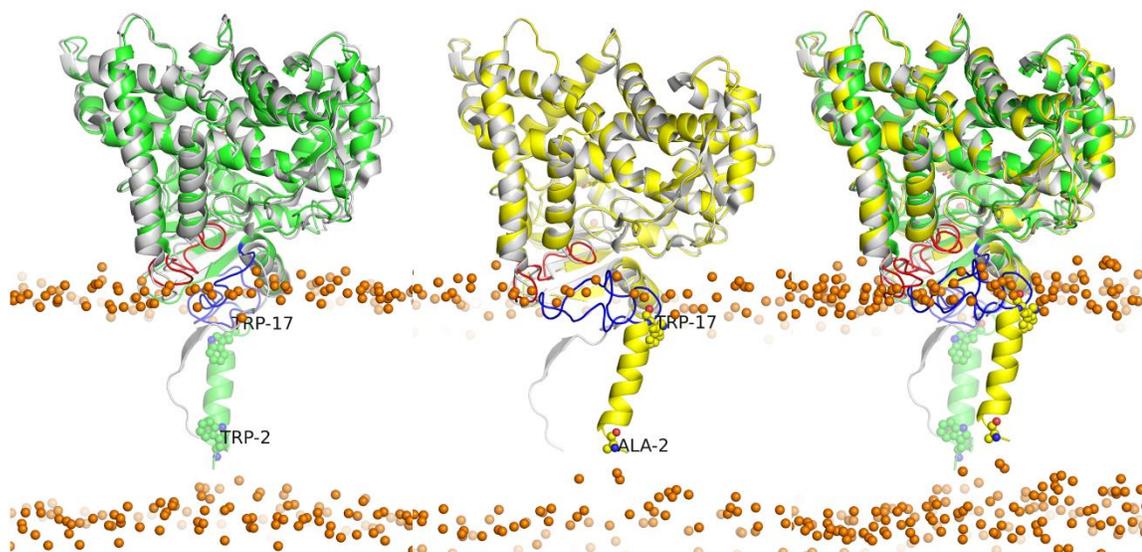


Figure 5.6: Starting configuration used for AAMD simulations of wtCYP17 (left) in green vs mtCYP17 (middle) in yellow superimposed on the x-ray crystal structure (silver) and wtCYP17, mtCYP17 and x-ray crystal structures (right), shown in cartoon representation. TM-helix and linker positions are compared. The linker residues 18-49 are shown in blue and F-G loop (residues 210-228) in red. The TM helix membrane spanning depth is compared between wtCYP17 and mtCYP17, with reference to the phosphate atoms (orange spheres) representing lipid head-group region. The anchoring residues W2 and W17 in the wtCYP17 and A2 and W17 in mtCYP17 TM helices are shown in ball and stick representation.

Furthermore, the position of TM helix in the membrane were predicted for wtTM and mtTM helices by prediction server, position of protein in membrane (PPM) (<http://opm.phar.umich.edu/server.php>) (189). For this, modeled 3D structure coordinates for TM helices region (N-terminal residues 1-22) were used. The orientation and position of the two TM helices differ from each other in the insertion depth and TM tilt angle. The TM tilt angle predicted by PPM server for mtCYP17 was $18 \pm 12^\circ$ which is closer to the TM angles observed in AAMD simulations (SIM1-SIM3) $19 \pm 11^\circ$ - $23 \pm 8^\circ$. A lower TM angle was predicted by PPM server for wtTM helix i.e., $13 \pm 6^\circ$, which is consistent with TM angle observed in CG ($13 \pm 7^\circ$) (Table 5.3) and AAMD simulations $13 \pm 8^\circ$ and $15 \pm 7^\circ$, respectively (Table 5.4). Similar TM tilt angle was observed in the TM-helix self-assembly simulation of wtTM helix ($12 \pm 6^\circ$) (Table 5.3). Besides TM angle, the insertion depth of wtTM and mtTM also differs in two helices. As shown in Figure 5.7,

wtTM span through hydrophobic core of the membrane (red and blue spheres indicating boundaries of hydrophobic core), like bitopic TM helices. mtTM helix showed monotopic TM helix characteristics.

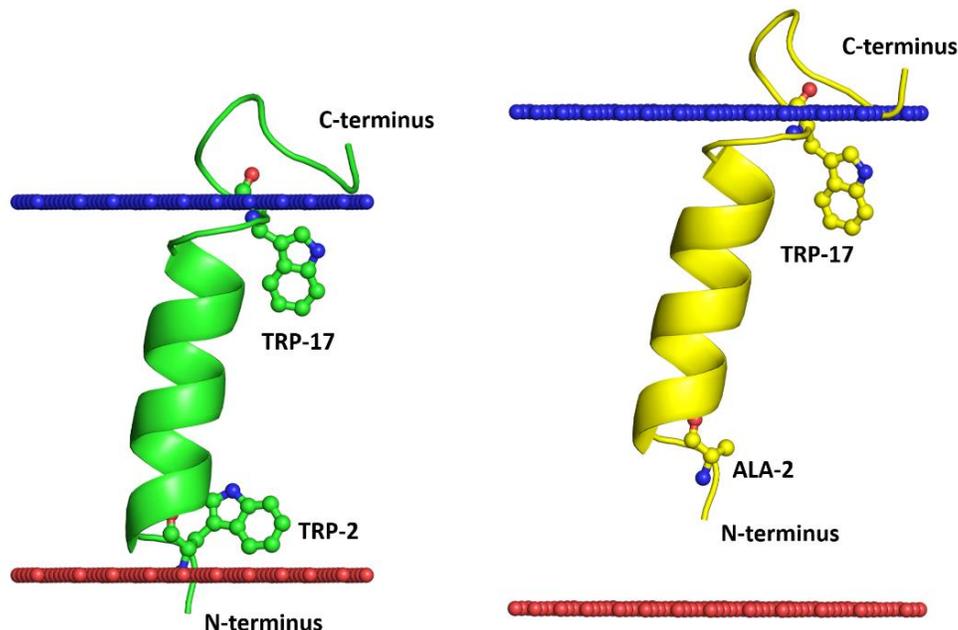


Figure 5.7: The TM helix orientation and position predicted by PPM server for all-atom model taken from back-conversion of wtTM helix (green) and mtTM helix (yellow) residues 1-22 shown in cartoon representation. mtTM helix conformation was same as wtTM helix, by mutating residues by Pymol mutagenesis option. The anchoring residues in the wtTM helix (W2 and W17) and mtTM helix (A2 and W17) are shown in stick representation. The hydrophobic boundaries are shown by blue and red spheres.

5.4.3.1 Different linker positions observed in wtCYP17 and mtCYP17:

The linker position was also checked after back conversion to AA models as it is affected by the TM-helix sequence and position in the membrane. The linker in CYP17 is comparatively long (32 residues) compared to other CYPs studied here, such as CYP2C9 and CYP2C19 (linker length 28 residues) and CYP3A4 with the shortest linker (22 residues). Different lengths of flexible linkers were used in mtCYP17 simulations (S1-S3) to obtain stable orientations of mtCYP17 in the membrane. The linker conformation in mtCYP17 differs from wtCYP17 linker position and develops strong polar interactions with F-G region in mtCYP17 (Figure 5.8 and Figure 5.9).

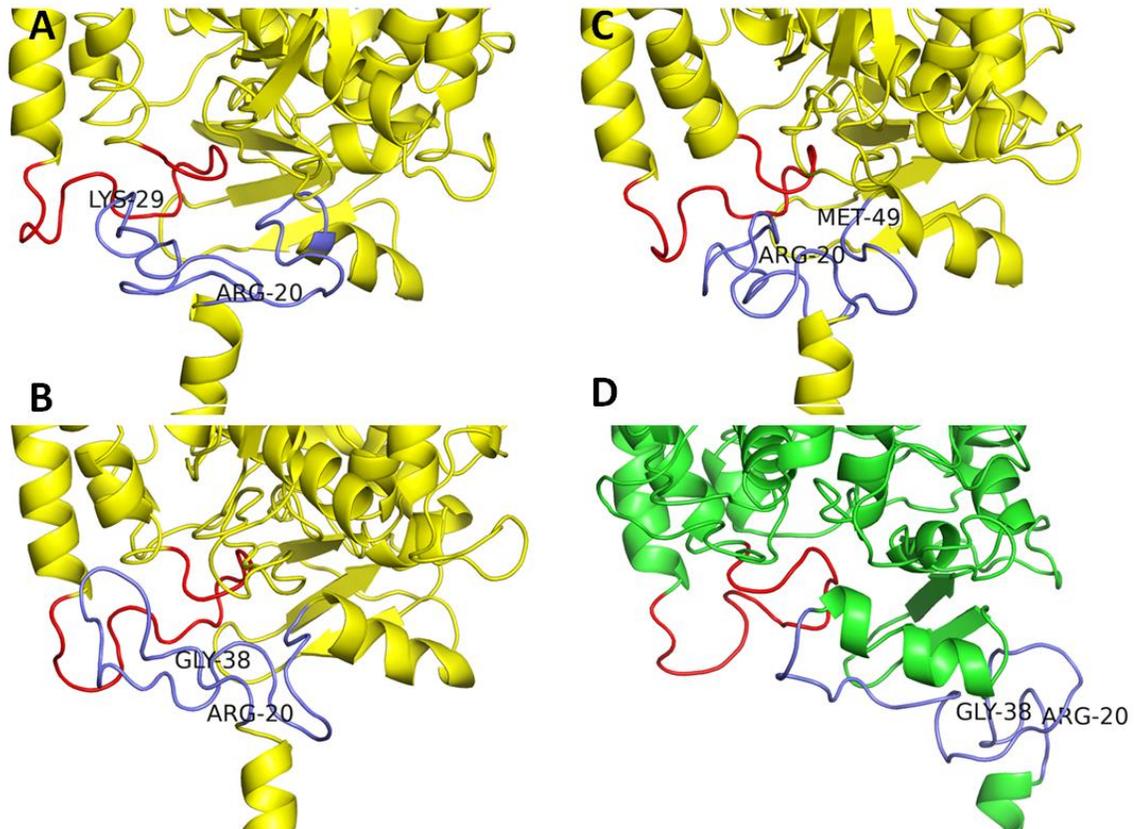


Figure 5.8: The different linker positions are compared between mtCYP17 CG simulations S1-S3 (A-C) and wtCYP17 simulation S4 (D), after back-conversion of representative frames taken from each CG simulation systems. The mtCYP17 globular domains shown in yellow (A-C) and wtCYP17 globular domain in green (D). The linker in blue and F-G loop is shown in red. The length of flexible linker is labelled in different CG simulations systems (S1-S4).

In CG simulation of mtCYP17, where TM-helix remained in the membrane, it was stabilized by the polar or charged interactions between linker and F-G region or by conformational changes in the TM helix backbone residues (extended conformation), resulting in the unwinding of TM helix conformation. The back-conversion of last frames obtained from mtCYP17 simulation (S1) trajectories resulted in the shorter TM-helix length which decreased from 3-19 residues (17 residues) to 4-13 (10 residues) in one trajectory and 4-15 (12 residues) in another trajectory, data not shown.

The linker position in wtCYP17 matches with the linker position observed in X-ray structure and did not show interactions with the F-G loop. While in mtCYP17

simulations, F-G loop region is blocked by the linker conformation, which also can interfere with substrate access via tunnel 2d/2f from the membrane to the active site. According to the tunnel definition (42), 2d tunnel egresses between the N-terminus and helices a/A' and A while tunnel 2f egresses between F'-helix/F-G loop and β 5 sheet. The 2f tunnel in CYP17 is formed by the aromatic gating residues (W220 and F224)(190), which are pushed away by linker position in mtCYP17 simulations. The side chain of F224 compared to crystal structure is flipped out and remains opposite to crystal structure F224 side chain (Figure 5.9).

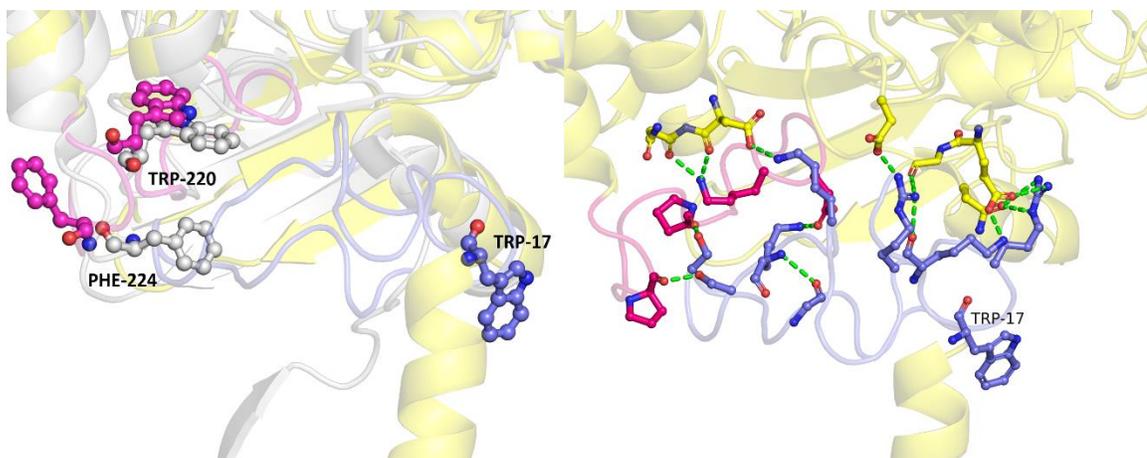


Figure 5.9: Initial representative frame after back-conversion of mtCYP17 (S3) (yellow) superimposed on X-ray crystal structure in grey (PDB ID 3RUK). Difference in the linker and F-G loop conformations between X-ray and CG frame is shown (A). The F-G loop orientation difference shown in pink for mtCYP17 after back-conversion and grey for X-ray crystal structure, and change in the side chain conformation of F224 is shown in stick representation. The polar interactions between the linker (blue), A helix, F helix (yellow), F'-G' loop region (pink) and beta-sheet (residue 480-485) shown by green dashed lines in the right panel. Side chains colored by atom-types.

5.4.4 All-atom MD simulation results

After achieving stable orientations of CYP17 in the membrane, AAMD simulations were performed to observe atomic level interactions between CYP17 and POPC membrane. The representative frames were taken from both wtCYP17 and mtCYP17 simulations. Three replica simulations for mtCYP17 and 3 wtCYP17 simulations were performed. In AAMD simulations (SIM1-6), the orientation of the globular domain above membrane

fluctuated throughout simulations in both mtCYP17 and wtCYP17 as shown in the angle and distance graphs (Figure 5.10). The angle alpha, TM and heme tilt angles show higher fluctuations, while, distances computed for CoM of linker, F'-G' and globular domain to the membrane CoM remained similar or decreased slightly from the distances observed in CG simulations. One reason for high fluctuations in CYP17 is that it has highly charged interface residues, with long flexible side chains, which develop transient polar/hydrogen bonds with phosphate head-group atoms in the membrane, resulting in the change in the orientation of globular domain above the membrane. In wtCYP17 simulation (SIM4), where angle alpha and distances varied from CG simulations, replica simulations with different initial velocity with anisotropic pressure (SIM5) and semiisotropic pressure (SIM6) were performed. The angles and distances values observed in mtCYP17 (SIM2-SIM3) and wtCYP17 (SIM5) match with each other. The heme-tilt angle values in mtCYP17 simulations are: $59 \pm 6^\circ$ (SIM2), $54 \pm 5^\circ$ (SIM3) and in wtCYP17 $57 \pm 5^\circ$ (SIM5). Lower heme-tilt angles are observed in mtCYP17 SIM1 ($45 \pm 5^\circ$) and wtCYP17 SIM4 ($47 \pm 5^\circ$) and SIM6 ($40 \pm 5^\circ$). A wide range of heme-tilt angle values (47° or 63°) are observed experimentally by Ohta et al (116) for CYP17 purified from guinea pig, which has 70% sequence identity with human CYP17. The heme-tilt angle observed in simulations by Cui et al. (190) varied between 40° and 67° , which suggests that CYP17 can adopt various orientations in the membrane. Overall, the heme-tilt angles observed in our AAMD simulations of wtCYP17 and mtCYP17 are consistent with the heme-tilt angle of CYP17 determined experimentally by Y. Ohta et al. (116), in previous simulations by Cui et al. and by Ivan Lenov ($64 \pm 4^\circ$) (unpublished data).

Compared to wtCYP17, mtCYP17 with linker trapped in the F'-G' loop region shows less changes in the globular domain orientation (Table 5.4). However, increased fluctuations in the TM tilt angle are observed in mtCYP17 compared to wtCYP17, which could be due to the freely floating monotopic helix in the absence of an anchoring W2 residue on the N-terminal end. TM tilt angle in wtCYP17 simulations (SIM4 and SIM6) corresponds to the previously reported lower tilt angles ($10-12^\circ$), observed in short TM-helices by AA simulation studies(157) as well as predicted by PPM server (discussed earlier). In the

presence of aromatic and charged flanking residues on both sides of the TM helix, wildtype TM helix remained stable in the membrane. However, due to the short TM helix length and the negative hydrophobic mismatch effect, a membrane curvature is observed near the N-terminal end of TM-helix, which is more pronounced in the mtCYP17 compared to wtCYP17 (Figure 5.11).

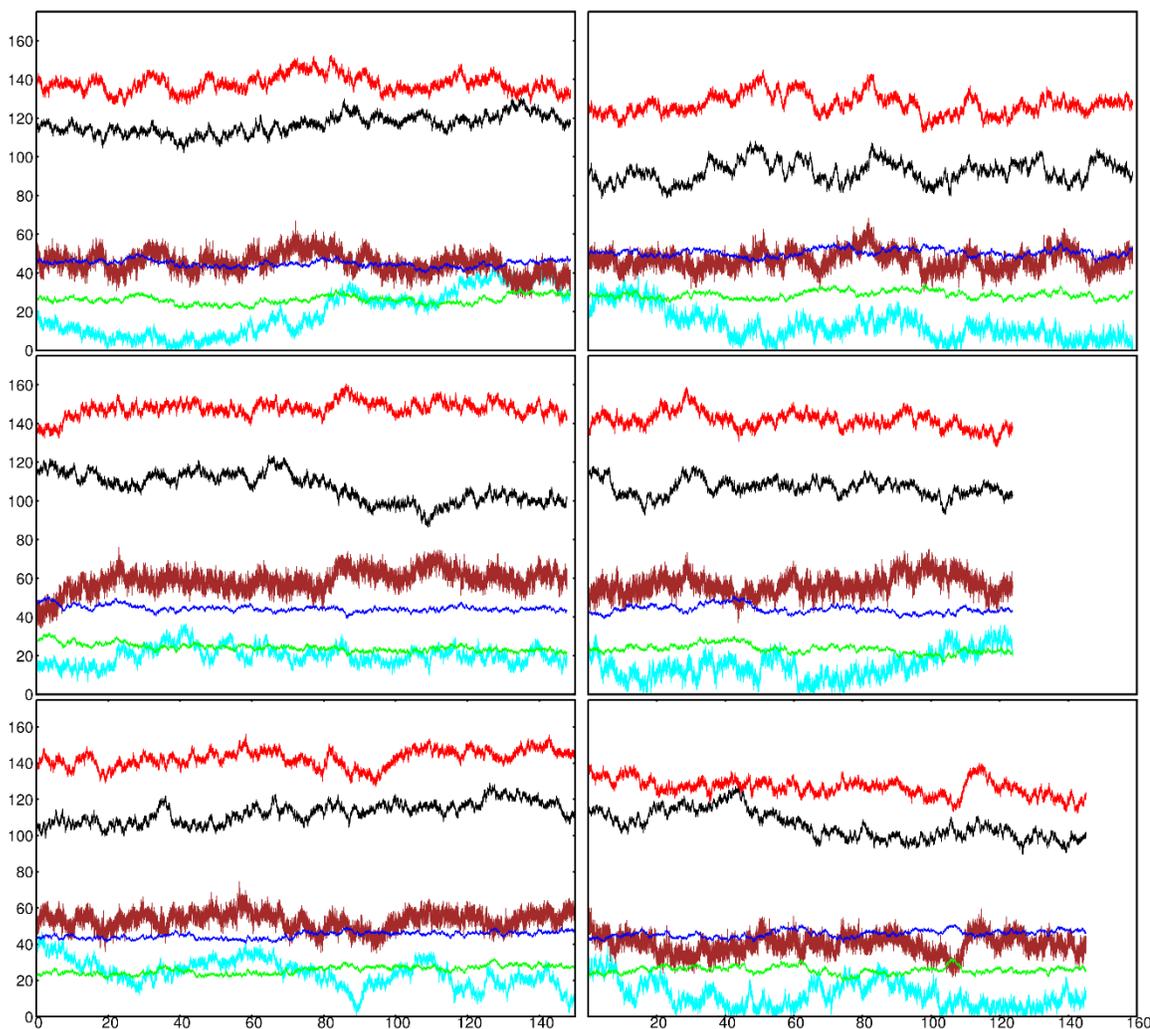


Figure 5.10: Evaluation of the position of the protein with respect to the phospholipid bilayer during AMD. The distance and angle plots for mtCYP17 (left column 1 SIM1-3) and wtCYP17 (column 2 SIM4-6). Angles: alpha (black), beta (red), heme-tilt (brown), TM-helix tilt (cyan) and distances: CoM distance of globular domain (blue) and F'-G' (green) CoM to the membrane CoM are shown.

Table 5.4: Mean and standard deviations calculated for angles and distances formed by mtCYP17 (SIM1-3) and wtCYP17 (SIM4-6) in AAMD simulations. CG simulation results are also shown for comparison.

AAMD Simulations	Angles (°)				Distances CoM (Å)		
	<i>Alpha</i>	<i>Beta</i>	<i>TM</i>	<i>Heme</i>	<i>Linker</i>	<i>F-G loop</i>	<i>Globular</i>
mtCYP17							
CG: S3	108±6	134±6	15±8	48.8*	22±2	28±2	46±2
SIM1	117±5	138±5	19 ± 11	45±5	20 ± 2	26 ± 2	45± 2
SIM2	105±7	136±4	20± 5	59±6	21 ± 2	24 ± 2	44± 2
SIM3	110±5	143±4	23 ± 8	54±5	20 ± 1	24 ± 1	44± 1
wtCYP17							
CG: S4	101±7	125±7	13±7	44.9*	20±2	27±2	46±2
SIM4	93±6	127±5	13±8	47±5	26±2	29±2	50±2
SIM5	108±5	142±4	15±7	57±5	22±2	23±2	44±2
SIM6	105±8	127±5	12±7	40±5	21±1	26±2	46±2

*heme-tilt angle of mtCYP17 and wtCYP17 was computed after conversion of representative CG frame to AA.

The orientations and interactions of wildtype vs mutant CYP17 are shown in Figure 5.11. Different orientation of linker position with respect to A-helix and F'-G' loop is highlighted in mtCYP17 (SIM3) and wtCYP17 (SIM6) by transparent circles. Compared to wtCYP17, mtCYP17 is more tilted towards the membrane on the distal side, increasing alpha and beta-angles, which also resulted in increased heme-tilt angle. In mtCYP17, F'-G' loop is more inserted in the membrane than wtCYP17 which is also evident from the CoM distance of the F'-G' and globular domain from the membrane (Table 5.4). An important difference between two proteins (wtCYP17 vs mtCYP17) can be seen in the linker orientation and position of TM-helix, which is positioned in front of A-helix in mtCYP17 and behind the A-helix in wtCYP17. The polar/charged patch of linker region (19-30) developed contact with beta-strand1 (residue 61-74) and beta-strand2 (residue 370-380) in wtCYP17. Whereas, linker in mtCYP17 interact with A helix, F-G region (residues from F-helix, F'-G' loop) and beta-sheet on the C-terminal end (residue 470-480) (Figure 5.12). Thus, the linker position near F'-G' loop (mtCYP17) or away from F'-G' loop (wtCYP19) could give the reason for more changes in the orientation occurring in the wtCYP17 compared to the mtCYP17. The linker in the mtCYP17 changed its conformation during AAMD simulations (SIM3) and moved away from F'-G' loop region. However, large conformational changes are difficult to observe during AAMD simulation.

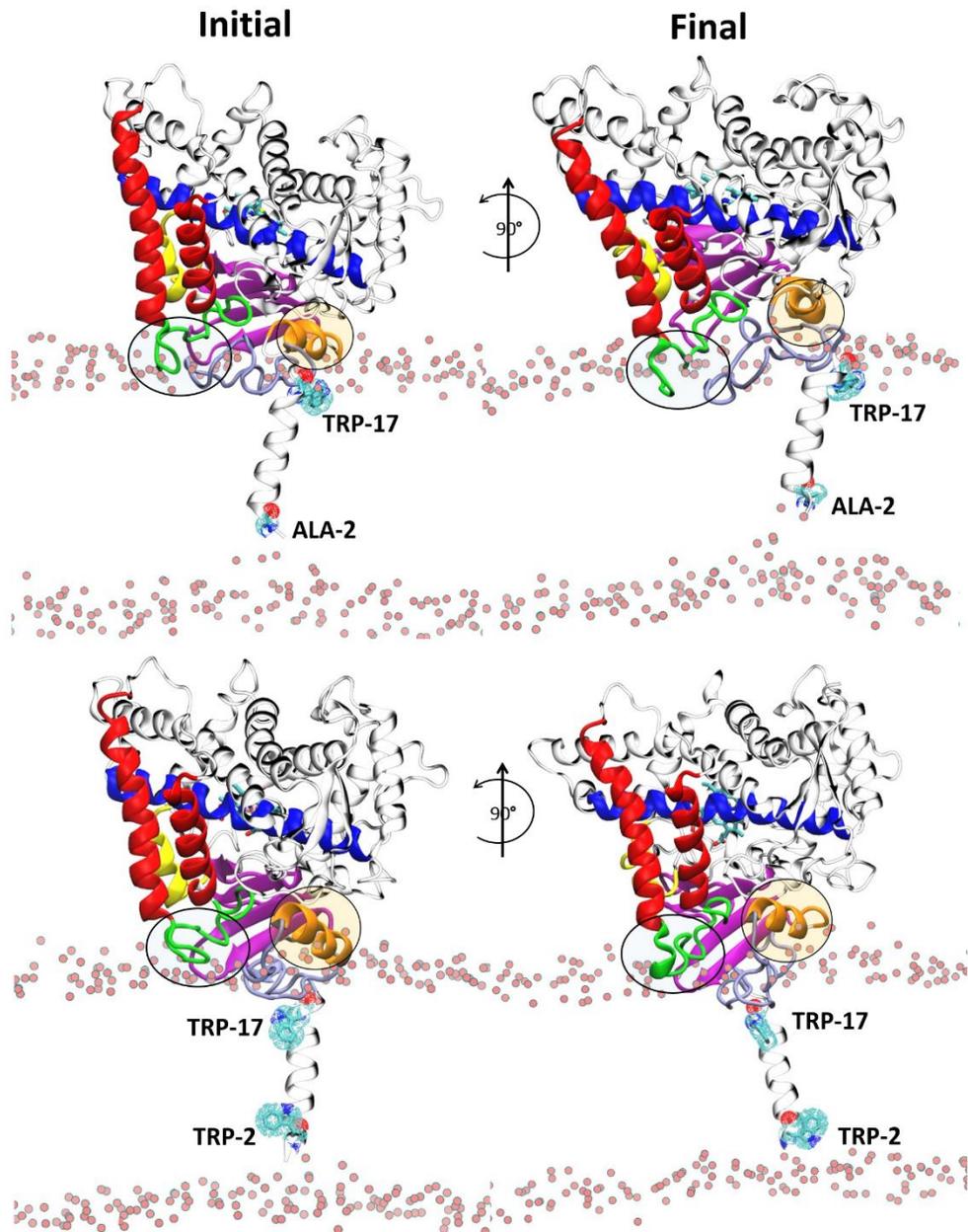


Figure 5.11: Last frame from *mtCYP17:SIM3* (top panel) and *wtCYP17:SIM6* simulations showing orientation of CYP17 in the membrane. Top panels show different views of the same frame rotated by $\sim 90^\circ$ around z-axis. The protein is shown in grey colored cartoon representation. Linker is shown by iceblue color, A-helix in orange, beta-strands in violet, BC-loop/helices in yellow, F-G helices in red and F'-G' loop in green. Central I-helix is colored blue, and heme is shown in cyan licorice representation. In the bottom panels is shown *wtCYP17*. Transparent circles indicate position of the linker with respect to the F'-G' loop (green) and A-helix (orange). The red spheres represent phosphate atoms in the POPC membrane.

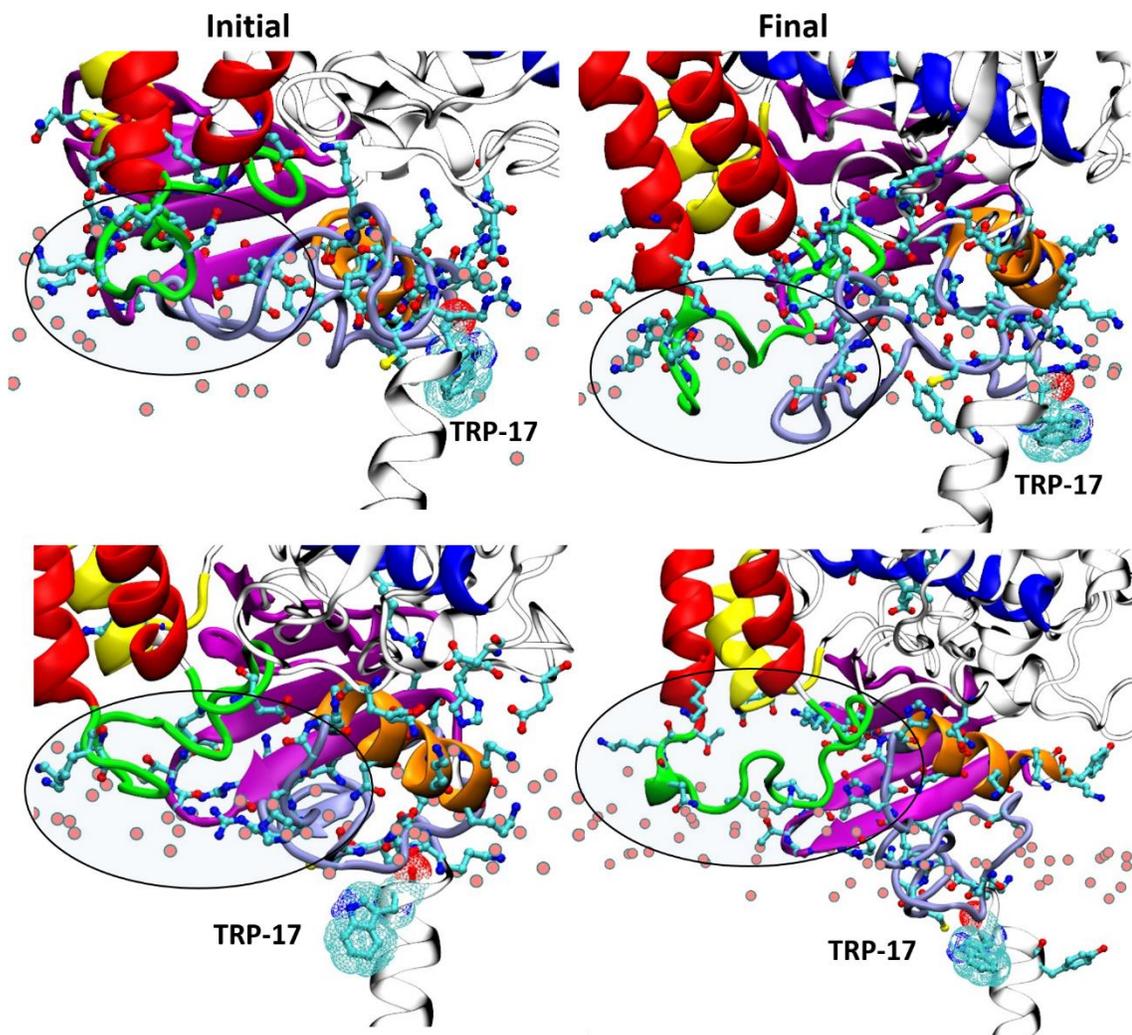


Figure 5.12: Comparison between the initial (left) and last frames (right) and between mtCYP17 (top) and wtCYP17 (bottom) is shown. The final frames are zoomed in view of Figure 5.11. Transparent circles indicate position of the linker (iceblue) with respect to the F-G loop (green). The red spheres represent phosphate atoms in the POPC membrane. The polar and charged residues within 5Å of phosphate atoms are shown in licorice representation.

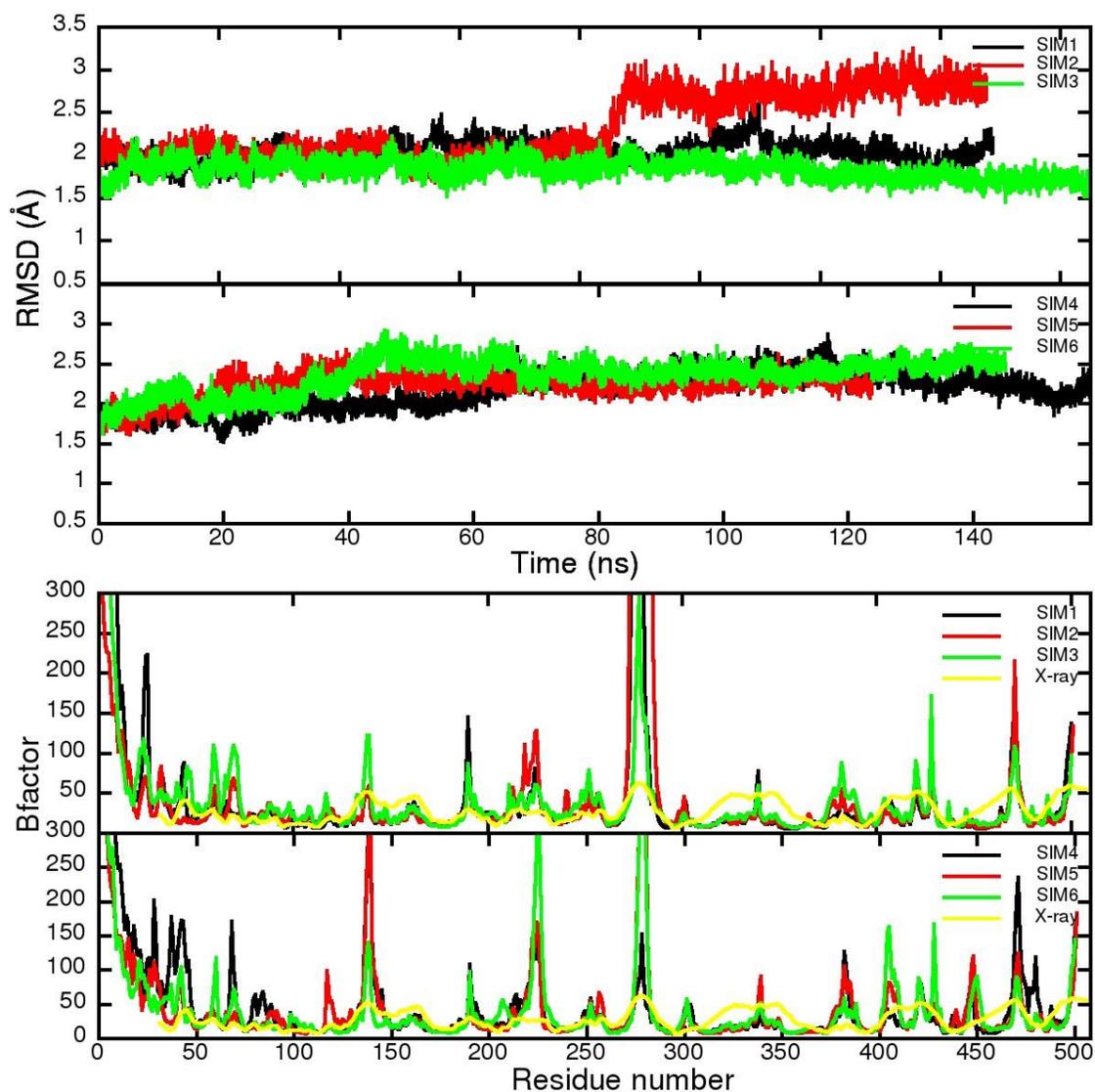


Figure 5.13: RMSD (top panel) and the average B-factor values (mean squared fluctuations multiplied by $8\pi^2/3$) Å² (bottom panel) are shown for wtCYP17 and mtCYP17 simulations. mtCYP17 simulation using anisotropic pressure control is shown in thick black line (SIM1) and replica simulations with semiisotropic pressure control red line (SIM2) and green (SIM3). wtCYP17 simulations shown in black (SIM4) and red (SIM5) using anisotropic pressure control while SIM6 using semiisotropic pressure control is shown in green lines. Crystallographic B-factor value shown in yellow line in bottom panel.

Furthermore, root mean squared deviation (RMSD) and B-factor values were calculated for AAMD simulations. The RMSD of the backbone atoms of the globular domain of the protein was calculated against the minimized structure. According to the RMSD graph for both mutant and wildtype simulations, the protein globular domain remained stable

throughout simulations (Figure 5.13). The average RMSD of the C α back-bone atoms of the globular domain (residues 50-502) with respect to the minimized structure was equal to or below 2.0 Å. However, in mtCYP17 SIM2, RMSD increased to 2.5 Å after ~80 ns, and then remained stable till 140 ns. The increased RMSD in SIM2 was observed due to highly flexible H-I loop as shown in the B-factor graph. The B-factor values obtained from AAMD simulations (SIM1-SIM6) were compared with the crystallographic B-factor values which showed similar pattern of fluctuation as seen in x-ray structure (Figure 5.13). The higher B-factor values are observed mainly in the TM-helix, linker regions and the different loops in contact with membrane or in the aqueous solvent. Compared to wtCYP17 simulations, lower B-factor is seen in mtCYP17 where linker is trapped in the F-G loop region

5.5 Conclusion

The current study, which involves CG and AA simulations to study CYP17-membrane interactions and to evaluate the effect of TM helix mutation on these interactions, reveals differences in the orientation and interactions of the TM-helix and linker region which influence the final orientations and interactions of the globular domain of mutant vs wildtype protein in the membrane. The change in N-terminal TM helix sequence leads to unstable interactions of TM helix in the membrane. In several mtCYP17 CG simulations (S1-S3), TM-helix is pulled out of the membrane interior and lies on the membrane interface, parallel to membrane plane. Expulsion of TM-helix in turn results in unstable orientations of CYP17 above the membrane. In mtCYP17 simulations, where TM-helix is seen in the membrane, it is stabilized due to the strong polar interactions developed by polar and charged linker residues (19-30) with the F-G region, or by distortion of TM helix length, change in the conformations of the backbone atoms of TM helix leading to unwinding of TM-helix. The interactions and orientation of the TM-helix in wtCYP17 is stabilized by aromatic and charged residues (W2 and E3) on the N-terminal end, which are mutated to hydrophobic residues (2A13) in mtCYP17. The TM tilt angle matches well with tilt angles reported for short TM helices (10-12°). The linker

position in wtCYP17 is similar to the linker position observed in the x-ray crystal structure. Due to mutation in the TM-helix, the mtCYP17 interactions with the membrane are disrupted due to expulsion of TM-helix from the hydrophobic core of the membrane. However, during AAMD simulations, the heme-tilt angles of mtCYP17 and wtCYP17 vary between $45 \pm 5^\circ$ to $59 \pm 6^\circ$ and $40 \pm 5^\circ$ to $57 \pm 5^\circ$, respectively. The heme-tilt angles observed in our simulations are closer to the experimental heme-tilt angle measure by Ivan Lenov for mtCYP17 ($64 \pm 4^\circ$), Ohta et al 1992 (47° to 63°) (116) and previous simulations results by Cui et al (190) who reported varying heme-tilt angles between 40° and 67° . Although the heme-tilt angles in mtCYP17 and wtCYP17 match with experimental values, the linker position and membrane interaction and degree of insertion in two CYPs differ from each other. In mtCYP17 where TM-helix is inside the membrane, the linker is trapped in the position which can obstructs the substrate entrance tunnels 2f/2d and could hinder substrate entrance in to the binding pocket through these tunnels. Although, mtCYP17 and wtCYP17 both are reported to be catalytically active (176), CG and AAMD simulations of two CYPs in the membrane showed different TM-helix and linker position. In wtCYP17 2d/2f tunnels are accessible for the hydrophobic substrates to enter the binding pocket through membrane, which are blocked by the linker position in mtCYP17. However, mtCYP17 TM-helix and linker position would not affect the binding of electron-transfer protein such as cytochrome-b5, rather CYP17- cytochrome-b5 intra-transmembrane interactions may stabilize the orientation and position of mutant TM-helix. A recent study using DNP-MAS-ssNMR spectroscopy has revealed such transmembrane-transmembrane interactions between full-length CYP2B4 and and cytochrome-b5 (191).

CHAPTER 6: CONCLUSION

Membrane-bound proteins (MPs) play important roles in cells as receptors, channels, transporters and enzymes. Around 26% of the human coding genome corresponds to MPs (1). Human cytochrome P450 (CYP) enzymes provide an example of bitopic integral membrane proteins (IMPs) with a single TM α -helix anchoring the catalytic domain with the membrane. CYPs play an important role in drug metabolism, steroid biosynthesis and xenobiotic degradation. Given the important role of MPs in regulating cellular function, there is increasing focus on studying the structure-function relationships of MPs. Experimental techniques like cryo-electron microscopy, NMR and X-ray crystallography have improved our understanding of the 3D-folds of MPs. However, there are practical difficulties with overexpression, solving high resolution crystal structures and studying the dynamic behavior of MPs in their physiological environment using experimental techniques. For example, the expression of eukaryotic CYPs in *E. coli* is increased after mutating residues in the TM-helix, truncating it, or even by mutating residues in the globular domain region that interact with the membrane. Moreover, the structures are usually determined without a lipid bilayer environment. Therefore, significant information on protein-membrane interactions and function is lost.

Experiments such as epitope analysis, mutagenesis, and tryptophan fluorescence scanning have been performed to understand the topology of CYPs, and to identify regions interacting with the membrane. The orientation of a few CYPs in the membrane has been determined experimentally by measuring the tilt angle of the active site heme by rotational diffusion and nanodisc experiments. However, these studies have been performed for specific CYPs. It is unclear how different CYPs interact with the membrane and what the key residue differences at the protein-membrane interface are that influence the orientations of CYPs above the membrane. Understanding these interactions and orientations in the membrane can help to elucidate the drug selectivity and specificity, as most of the substrates, prior to metabolism, are hydrophobic in nature and enter into the buried binding pocket of CYPs through the ligand access

tunnels which open into the membrane. For this purpose, computational modeling and simulations at various time and length scales have played a significant role in understanding the mechanistic behavior and dynamics of CYPs in their native environment.

Previously, multiscale simulations, employing CG and AA simulations, have been used to study the orientations of several CYPs (human CYP2C9, CYP3A4 and CYP51) in a bilayer (31, 134, 192). In this work, I have optimized the multiscale simulation protocol to study CYP-membrane interactions (193). For this, CYP3A4 was used as a test case for the insertion and orientation of CYP3A4 in a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) bilayer. The CYP3A4 orientation in the membrane has been predicted previously by experiments and MD simulations. Therefore, it provided us with an excellent test system for evaluating different protocols. We have identified an optimum procedure in terms of computational efficiency and the quality of the results for CYP3A4 which match with the experiments. This protocol was successfully applied to four other human drug metabolizing CYPs (1A1, 1A2, 2C9 and 2C19) and CYP17 and CYP19, which are involved in steroidogenesis. From simulations results, we conclude that, for a computationally efficient multiscale procedure to immerse a protein in a membrane, the CG simulations should be carried out with the MARTINI force field version 2.2 with the standard NPW water model and a RF treatment of long-range Coulombic interactions.

The optimized multiscale simulation protocol was applied to the CYP2C subfamily specifically CYP2C9 and CYP2C19, which contributes significantly to the hepatic clearance of several drugs. The two isoforms have high sequence conservation (94% sequence identity), yet, they have distinct substrate specificities. The main differences are observed in the linker, beta-strand1, the B-C loop, helices F, F'-G', G regions and their turns, which are the flexible substrate recognition sites (SRS1,2 and 3), located at the entrance to the active site. The sequence differences and conformational changes outside the binding cavity can be responsible for the differences in CYP-membrane

interactions and orientations and differential entry of drugs in to the binding pocket. Using our optimized multiscale simulation approach, I was able to observe differences in CYP2C9 and CYP2C19 orientations and insertion depths in the membrane. The key residue differences were identified in the protein-membrane interface that appear to be responsible for these differences. On mutation of the residues differing in the interface region, CYP2C9 showed similar orientations to CYP2C19 in 50% of CG simulations, as well decreased insertion depth. However, in this study, a critical issue is the selection of the starting structure of CYP2C9, as the crystal structure of CYP2C9 was determined either with missing linker and F'-G' loop regions, or with F'-G' region residues modified to those of the CYP2C5 sequence. Therefore, models with loop and helix conformations of the F'-G' regions of CYP2C9 have been simulated. Depending on the initial structure of CYP2C9, different orientations of CYP2C9 in the membrane were observed (31). However, both models of CYP2C9, with loop or helix conformations of the F'-G' region, differ in the orientation and insertion depth in the membrane from CYP2C19, the crystal structure of the latter was resolved with a helix conformation F'-G' region (23).

Furthermore, I investigated the effect of mutations in the TM-helix of CYP17 on the interactions and orientation of mtCYP17 in the membrane and compared with wtCYP17 employing both CG and AA simulations. The structural and functional importance of the TM-helix and the position of the linker is unclear, since, on truncation of the TM-helix, CYPs can still associate with the membrane and retain their enzymatic activity. Modification in the N-terminal helix residues was done to increase expression of CYP17 in *E. coli*. The heme-tilt angles observed in mtCYP17 and wtCYP17 match with the experimental heme-tilt angle measured by Ivan Lenov for mtCYP17 ($64 \pm 4^\circ$) (194), Ohta et al (47° to 63°) (116) and simulations results by Cui et al (195). However, in simulations, it was observed that mutations in the TM-helix affected the TM-helix orientation and interactions with the membrane. In mtCYP17 simulations, the TM-helix was pulled out from the hydrophobic core of the membrane and was lying parallel to

membrane plane. The unstable TM-helix in mtCYP17 disrupted the globular domain-membrane interactions, or in some cases, the linker was trapped in a position which blocked tunnel 2d/2f opening in the membrane. On the other hand, in wtCYP17 the 2d/2f tunnels are accessible for the hydrophobic substrates to enter the binding pocket through membrane and the linker position is similar to the linker position observed in X-ray structure. I assume that the mtCYP17 TM-helix and linker positions would not affect the binding of electron-transfer proteins such as cytochrome-b5, and therefore, not inhibit the catalysis at the enzyme active site but that enzyme turnover could be affected. Truncation of the TM-helix residues in CYP17 has shown 2.4-fold decreased catalytic activity which could be due to loss of hydrophobic interactions between CYP17 and its redox partner (196). Such hydrophobic interactions between TM helices of CYP2B4 and cytochrome-b5 have been reported by dynamic nuclear polarization (DNP) NMR spectroscopy under magic angle spinning (MAS)(191). Therefore, the unstable TM-helix lying parallel to the membrane plane in mtCYP17 could result in loss of inter-TM helix interactions between mtCYP17 and its redox partner, leading to decreased catalytic activity or affect membrane anchoring and thereby lead to reduced activity. The TM-helix orientation can be measured experimentally using solid state NMR spectroscopy (179, 180).

A complex biomolecular system, such as a CYP-membrane in explicit water solvent, constitutes a large system of more than 300,000 atoms and requires long computational times to equilibrate the system and achieve convergence of the dynamic properties. Increasing system size decreases the possible simulation time and therefore, sampling slow conformational changes in atomistic detail becomes difficult. In order to study mechanistic interactions in a CYP-membrane system and conformational transitions and the reorientation of CYP globular domain with respect to the membrane, current simulation time scales present a major bottleneck. However, using simplified models or enhanced sampling methods can overcome this limitation, but at the expense of losing atomic detail information and accuracy. MARTINI CG simulation is one of the

most widely used methods to obtain large conformational changes with reduced computational needs. Therefore, combining CG and AA simulations can increase the length of simulations and solve the sampling issue.

With the advancement of computational resources, MD simulation has increased the possibility to study biomolecular processes at spatial and temporal levels often inaccessible by experiments. However, biomolecular force-field accuracy is one of the major bottlenecks in MD simulation, when studying complex macromolecular systems. Force-fields are typically parameterized independently for each molecule type, i.e., small molecules, protein, lipid, DNA, carbohydrate etc. Therefore, combining different force-fields for protein-membrane systems correctly is critical as there is little experimental data for mixed systems against which to validate force-fields. In this study, I have evaluated the effects of different all-atom force-fields on CYP-membrane interactions. Two different lipid force-fields, GAFF lipid and LIPID14, showed different orientations and insertion depths of CYP2C9 and CYP2C19. GAFF lipid requires application of additional surface tension whereas LIPID14 is optimized without surface tension. The structural and dynamic properties of the membrane, such as area per lipid, membrane thickness and order parameters, matched well with experimental values in the case of LIPID14, while the GAFF lipid showed decreased membrane thickness, increased APL and higher order parameters. Furthermore, the effect of combining two force-fields on the conservation of protein secondary structures was also checked. I therefore concluded that LIPID14 in combination with the latest protein force-fields, ff12SB or ff14SB, gives a better prediction of CYP-membrane interactions, compared to the GAFF lipid and ff99SB force-fields.

In this work on CYP-membrane interactions, I have used a homogeneous lipid bilayer (POPC), while the native endoplasmic reticulum membrane is a mixture of different lipid molecules and cholesterol concentrations which may form different membrane microdomains, such as lipid-ordered domains enriched with saturated lipids and cholesterol, and lipid disordered domains containing unsaturated lipids. It has been

shown that different CYPs are localized in different domains of the ER membrane (197, 198). It could be possible that the length of the TM-helix plays a role in the localization of CYPs in different regions due to the hydrophobic mismatch effect. Also, inclusion of cholesterol in the membrane will not only affect the membrane properties, such as membrane thickness, lipid order parameter, gel to fluid phase transition and fluidity of membrane, but it will also affect the protein-membrane interaction, degree of insertion of the globular domain in the membrane, as well as the entry of substrates from the lipid bilayer into the active site by substrate access tunnels (143, 199). The strength of interaction and stability of the TM-helix, as well as globular domain orientation, may vary on changing the lipid bilayer composition. It would therefore be desirable to check the effect of cholesterol on CYP-membrane interactions and orientations. But, inclusion of several different lipid molecules will not only increase the complexity of system but it will also mean that longer simulation times are required to fully equilibrate such complex systems. Furthermore, CYPs are known to form homo- and heterooligomers in biological systems (200), and also to interact with their redox partners to carry out catalytic reactions. These reactions will influence the orientation of CYPs in the membrane. The orientation and interactions of the TM-helix in the membrane can also be stabilized in the oligomer state. Therefore, studying protein-protein and protein-membrane interactions can further improve our understanding about CYP orientations in the membrane and opening of ligand access tunnels, and, thereby the mechanism of drug selectivity and drug metabolism.

This work provides a mechanistic view of different CYP-membrane interactions at atomistic detail and sheds light on the mechanism of drug selectivity and specificity exhibited by different CYPs. In this work, I have identified important SRS regions located at the outer surface of CYPs, which contribute to the ligand selectivity and entrance into binding pocket as well as the CYP-membrane orientation and interactions. I have also identified the effect of mutations in the TM-helix on CYP-membrane interactions, which explains the decreased catalytic efficiency of the enzyme. Since CYPs are highly

polymorphic, this work can be extended to understand the mechanism of single nucleotide polymorphisms (SNP) on enzyme catalysis. The orientations of different CYPs (CYP2C9, CYP2C19, CYP3A4 and CYP17 and CYP19) in the membrane obtained in the current study can be used as starting point for studying protein-protein interactions, such as CYP and CPR (cytochrome P450 reductase) interactions to investigate the mechanism of electron transfer from CPR to different CYPs. The optimized multiscale protocol is suitable for use with different membrane compositions or with different CYP-oligomers in the membrane environment, which will give a more realistic picture and mechanistic insights into protein-protein-membrane interactions.

References

1. Fagerberg, L., Jonasson, K., Von Heijne, G., Uhlén, M., and Berglund, L. (2010) Prediction of the human membrane proteome. *Proteomics*. **10**, 1141–1149
2. Babcock, J. J., and Li, M. (2014) Deorphanizing the human transmembrane genome: A landscape of uncharacterized membrane proteins. *Acta Pharmacol. Sin.* **35**, 11–23
3. Lindahl, E., and Sansom, M. S. (2008) Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **18**, 425–431
4. Gábor E. Tusnády, Z. D. and I. S. (2013) PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* **41**, D524–D529
5. Monk, B. C., Tomasiak, T. M., Keniya, M. V, Huschmann, F. U., Tyndall, J. D. a., O’Connell, J. D., Cannon, R. D., McDonald, J. G., Rodriguez, A., Finer-Moore, J. S., and Stroud, R. M. (2014) Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. *Proc. Natl. Acad. Sci.* **111**, 3865–3870
6. Eeman, M., and Deleu, M. (2010) From biological membranes to biomimetic model membranes. *Biotechnol. Agron. Soc. Environ.* **14**, 719–736
7. Eze, M. O. (1991) Phase transitions in phospholipid bilayers: Lateral phase separations play vital roles in biomembranes. *Biochem. Educ.* **19**, 204–208
8. White, S. H., Ladokhin, A. S., Jayasinghe, S., and Hristova, K. (2001) How Membranes Shape Protein Structure*. 10.1074/jbc.R100008200
9. Lee, A. . (2003) Lipid protein interactions in biological membranes: a structural perspective. *Biochim. Biophys. Acta - Biomembr.* **1612**, 1–40
10. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H., and von Heijne, G. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*. **433**, 377–81
11. Wimley, W. C., and White, S. H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature*. **3**, 842–848
12. Engelman, D. M., Steitz, T. a, and Goldman, a (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321–353
13. Killian, J. A., and Von Heijne, G. (2000) How proteins adapt to a membrane-water interface. *Trends Biochem. Sci.* **25**, 429–434
14. Munro, A. W., Girvan, H. M., Mason, A. E., Dunford, A. J., and McLean, K. J. (2013) What makes a P450 tick? *Trends Biochem. Sci.* **38**, 140–150

15. Ortiz de Montellano, P. R. (2013) Cytochrome P450-activated prodrugs. *Future Med. Chem.* **5**, 213–228
16. Zanger, U. M., and Schwab, M. (2013) Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* **138**, 103–141
17. Danielson, P. B. (2002) The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metab.* **3**, 561–97
18. Furge, L. L., and Guengerich, F. P. (2006) Cytochrome P450 Enzymes in Drug Metabolism and Chemical Toxicology. *Biochem. Mol. Biol. Educ.* **34**, 66–74
19. Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak Vinkovi?, D., and Jhoti, H. (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature.* **424**, 464–468
20. Johnson, E. F., and Stout, C. D. (2013) Structural diversity of eukaryotic membrane cytochrome p450s. *J. Biol. Chem.* **288**, 17082–90
21. Johnson, E. F., and Stout, C. D. (2005) Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun.* **338**, 331–6
22. Sansen, S., Hsu, M.-H., Stout, C. D., and Johnson, E. F. (2007) Structural insight into the altered substrate specificity of human cytochrome P450 2A6 mutants. *Arch. Biochem. Biophys.* **464**, 197–206
23. Reynald, R. L., Sansen, S., Stout, C. D., and Johnson, E. F. (2012) Structural characterization of human cytochrome P450 2C19: active site differences between P450s 2C8, 2C9, and 2C19. *J. Biol. Chem.* **287**, 44581–91
24. Dong, D., Wu, B., Chow, D., and Hu, M. (2012) Substrate selectivity of drug-metabolizing cytochrome P450s predicted from crystal structures and in silico modeling. *Drug Metab. Rev.* **44**, 192–208
25. Otyepka, M., Skopalík, J., Anzenbacherová, E., and Anzenbacher, P. (2007) What common structural features and variations of mammalian P450s are known to date? *Biochim. Biophys. Acta.* **1770**, 376–89
26. Mustafa, G., Yu, X., and Wade, R. C. (2014) Structure and Dynamics of Human Drug-Metabolizing Cytochrome P450 Enzymes. in *Drug Metabolism Prediction* (KirchmairJohannes ed), pp. 77–102, Wiley-VCH, Weinheim
27. Williams, P. a, Cosme, J., Sridhar, V., Johnson, E. F., and McRee, D. E. (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell.* **5**, 121–131
28. Gotoh, O. (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2)

proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* **267**, 83–90

29. Zawaira, A., Ching, L. Y., Coulson, L., Blackburn, J., and Wei, Y. C. (2011) An expanded, unified substrate recognition site map for mammalian cytochrome P450s: analysis of molecular interactions between 15 mammalian CYP450 isoforms and 868 substrates. *Curr. Drug Metab.* **12**, 684–700
30. Ozalp, C., Szczesna-Skorupa, E., and Kemper, B. (2006) Identification of membrane-contacting loops of the catalytic domain of cytochrome P450 2C2 by tryptophan fluorescence scanning. *Biochemistry.* **45**, 4629–4637
31. Cojocar, V., Balali-Mood, K., Sansom, M. S. P., and Wade, R. C. (2011) Structure and Dynamics of the Membrane-Bound Cytochrome P450 2C9. *PLoS Comput. Biol.* **7**, e1002152
32. Berka, K., Hendrychová, T., Anzenbacher, P., and Otyepka, M. (2011) Membrane Position of Ibuprofen Agrees with Suggested Access Path Entrance to Cytochrome P450 2C9 Active Site. *J. Phys. Chem. A.* **115**, 11248–11255
33. Yu, X., Cojocar, V., and Wade, R. C. (2013) Conformational diversity and ligand tunnels of mammalian cytochrome P450s. *Biotechnol. Appl. Biochem.* **60**, 134–45
34. Denisov, I. G., Shih, a Y., and Sligar, S. G. (2012) Structural differences between soluble and membrane bound cytochrome P450s. *J. Inorg. Biochem.* **108**, 150–8
35. Wade, R. C., Winn, P. J., Schlichting, I., and Sudarko (2004) A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.* **98**, 1175–1182
36. Winn, P. J., Lüdemann, S. K., Gauges, R., Lounnas, V., and Wade, R. C. (2002) 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. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5361–6
37. Shen, Z., Cheng, F., Xu, Y., Fu, J., Xiao, W., Shen, J., Liu, G., Li, W., and Tang, Y. (2012) Investigation of indazole unbinding pathways in CYP2E1 by molecular dynamics simulations. *PLoS One.* **7**, e33500
38. Lüdemann, S. K., Lounnas, V., and Wade, R. C. (2000) 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. *J. Mol. Biol.* **303**, 797–811
39. Lüdemann, S. K., Lounnas, V., and Wade, R. C. (2000) How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J. Mol. Biol.* **303**, 813–830

40. Wade, R. C., Motiejunas, D., Schleinkofer, K., Sudarko, Winn, P. J., Banerjee, A., Kariakin, A., and Jung, C. (2005) Multiple molecular recognition mechanisms. Cytochrome P450--a case study. *Biochim. Biophys. Acta.* **1754**, 239–244
41. Schleinkofer, K., Winn, P. J., Lüdemann, S. K., and Wade, R. C. (2005) Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep.* **6**, 584–589
42. Cojocaru, V., Winn, P. J., and Wade, R. C. (2007) The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta - Gen. Subj.* **1770**, 390–401
43. Rowland, P., Blaney, F. E., Smyth, M. G., Jones, J. J., Leydon, V. R., Oxbrow, A. K., Lewis, C. J., Tennant, M. G., Modi, S., Eggleston, D. S., Chenery, R. J., and Bridges, A. M. (2006) Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* **281**, 7614–22
44. Wang, A., Savas, U., Hsu, M.-H., Stout, C. D., and Johnson, E. F. (2012) Crystal structure of human cytochrome P450 2D6 with prinomastat bound. *J. Biol. Chem.* **287**, 10834–43
45. Yu, X., Nandekar, P., Mustafa, G., Cojocaru, V., Lepesheva, G. I., and Wade, R. C. (2016) Ligand tunnels in *T. brucei* and human CYP51: Insights for parasite-specific drug design. *Biochim. Biophys. Acta - Gen. Subj.* **1860**, 67–78
46. Cojocaru, V., Winn, P., and Wade, R. (2012) Multiple, ligand-dependent routes from the active site of cytochrome P450 2C9. *Curr. Drug Metab.* **13**, 143–54
47. Li, W., Shen, J., Liu, G., Tang, Y., and Hoshino, T. (2011) Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations. *Proteins.* **79**, 271–81
48. Fishelovitch, D., Shaik, S., Wolfson, H. J., and Nussinov, R. (2009) Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism. *J. Phys. Chem. B.* **113**, 13018–25
49. Cojocaru, V., Winn, P., and Wade, R. (2012) Multiple, ligand-dependent routes from the active site of cytochrome P450 2C9. *Curr. Drug Metab.*
50. Phillips, R., Ursell, T., Wiggins, P., and Sens, P. (2009) Emerging roles for lipids in shaping membrane-protein function. *Nature.* **459**, 379–85
51. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., and Shaw, D. E. (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* **41**, 429–52
52. Niemelä, P. S., Miettinen, M. S., Monticelli, L., Hammaren, H., Bjelkmar, P., Murtola, T., Lindahl, E., and Vattulainen, I. (2010) Membrane proteins diffuse as

dynamic complexes with lipids. *J. Am. Chem. Soc.* **132**, 7574–7575

53. Ramakrishnan, N., Sunil Kumar, P. B., and Radhakrishnan, R. (2014) Mesoscale computational studies of membrane bilayer remodeling by curvature-inducing proteins. *Phys. Rep.* **543**, 1–60
54. Pluhackova, K., Böckmann, R. A., Search, H., Journals, C., Contact, A., Iopsience, M., and Address, I. P. (2015) Biomembranes in atomistic and coarse-grained simulations. *J. Phys. Condens. Matter.* **27**, 323103
55. Freites, J. A., Tobias, D. J., von Heijne, G., and White, S. H. (2005) Interface connections of a transmembrane voltage sensor. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15059–15064
56. Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G., and Schulten, K. (2009) Discovery Through the Computational Microscope. *Structure.* **17**, 1295–1306
57. Marrink, S. J., and Tieleman, D. P. (2013) Perspective on the Martini model. *Chem. Soc. Rev.* **42**, 6801–6822
58. Chavent, M., Duncan, A. L., and Sansom, M. S. (2016) Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale. *Curr. Opin. Struct. Biol.* **40**, 8–16
59. Ingólfsson, H. I., Arnarez, C., Periole, X., and Marrink, S. J. (2016) Computational 'microscopy' of cellular membranes. *J. Cell Sci.* **129**, 257–268
60. Ingólfsson, H. I., Melo, M. N., van Eerden, F. J., Arnarez, C., Lopez, C. A., Wassenaar, T. A., Periole, X., de Vries, A. H., Tieleman, D. P., and Marrink, S. J. (2014) Lipid Organization of the Plasma Membrane. *J. Am. Chem. Soc.* **136**, 14554–14559
61. Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016) Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **116**, 7898–7936
62. Arinaminpathy, Y., Khurana, E., Engelman, D. M., and Gerstein, M. B. (2009) Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov. Today.* **14**, 1130–5
63. Koehler Leman, J., Ulmschneider, M. B., and Gray, J. J. (2015) Computational modeling of membrane proteins. *Proteins Struct. Funct. Bioinforma.* **83**, 1–24
64. Barth, P., Schonbrun, J., and Baker, D. (2007) Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc Natl Acad Sci U S A.* **104**, 15682–15687
65. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815

66. Kelm, S., Shi, J., and Deane, C. M. (2010) MEDELLER: Homology-based coordinate generation for membrane proteins. *Bioinformatics*. **26**, 2833–2840
67. Ebejer, J. P., Hill, J. R., Kelm, S., Shi, J., and Deane, C. M. (2013) Memoir: template-based structure prediction for membrane proteins. *Nucleic Acids Res.* **41**, 379–383
68. Hildebrand, A., Remmert, M., Biegert, A., and Söding, J. (2009) Fast and accurate automatic structure prediction with HHpred. *Proteins Struct. Funct. Bioinforma.* **77**, 128–132
69. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252-8
70. Koehl, P., and Delarue, M. (1995) A self consistent mean field approach to simultaneous gap closure and side-chain positioning in homology modelling. *Nat.Struct.Biol.* **2**, 163
71. Levitt, M. (1992) Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.* **226**, 507–533
72. Webb, B., and Sali, A. (2014) Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics*. **47**, 5.6.1-32
73. Guvench, O., and MacKerell, A. D. (2008) *Comparison of protein force fields for molecular dynamics simulations.*, 10.1007/978-1-59745-177-2_4
74. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 10.1007/s008940100045
75. Marrink, S. J., de Vries, A. H., and Mark, A. E. (2004) Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B.* **108**, 750–760
76. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007) The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B.* **111**, 7812–7824
77. Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., and Marrink, S.-J. (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **4**, 819–834
78. Jong, D. H. De, Singh, G., Bennett, W. F. D., Arnarez, C., Wassenaar, T. A., Scha, L. V, Periole, X., Tieleman, D. P., and Marrink, S. J. (2013) Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **9**, 687–697

79. Yesylevskyy, S. O., Schäfer, L. V, Sengupta, D., and Marrink, S. J. (2010) Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.* **6**, e1000810
80. Periole Xavier, M. S.-J. (2013) *Biomolecular Simulations* (Monticelli, L., and Salonen, E. eds), Methods in Molecular Biology, Humana Press, Totowa, NJ, 10.1007/978-1-62703-017-5
81. Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., and Marrink, S.-J. (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **4**, 819–834
82. Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., and Marrink, S.-J. (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **4**, 819–834
83. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* **22**, 2577–2637
84. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and Vries, A. H. De (2007) The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations. **111**, 7812–7824
85. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 10.1016/j.softx.2015.06.001
86. de Jong, D. H., Baoukina, S., Ingólfsson, H. I., and Marrink, S. J. (2016) Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **199**, 1–7
87. Periole, X., Cavalli, M., Marrink, S.-J., and Ceruso, M. A. (2009) Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **5**, 2531–2543
88. and, S. A. A., and McCammon*, J. A. (2006) Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. 10.1021/CR040426M
89. Verlet, L. (1967) Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103
90. Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem.*

Phys. **76**, 637

91. Allen, M. (2004) Introduction to molecular dynamics simulation. *Comput. Soft Matter From Synth. Polym. to* [online] http://www.physics.iitm.ac.in/~iol/lecture_notes/allen.pdf (Accessed January 9, 2015)
92. Hockney, R. W., Goel, S. P., and Eastwood, J. W. (1974) Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **14**, 148–158
93. Petrenko, R., and Meller, J. (2010) Molecular Dynamics. in *Encyclopedia of Life Sciences*, pp. 1–8, John Wiley & Sons, Ltd, Chichester, UK, 10.1002/9780470015902.a0003048.pub2
94. Monticelli, L., and Tieleman, D. P. (2013) Force Fields for Classical Molecular Dynamics. in *Methods in molecular biology (Clifton, N.J.)* (Monticelli, L., and Salonen, E. eds), pp. 197–213, Methods in Molecular Biology, Humana Press, Totowa, NJ, **924**, 197–213
95. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089
96. Anézo, C., de Vries, A. H., Höltje, H.-D., Tieleman, D. P., and Marrink, S.-J. (2003) Methodological Issues in Lipid Bilayer Simulations. *J. Phys. Chem. B.* **107**, 9424–9433
97. Scott, R., Allen, M. P., and Tildesley, D. J. (1991) Computer Simulation of Liquids. *Math. Comput.* **57**, 442
98. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, a, and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690
99. Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* 10.1063/1.2408420
100. Hoover, W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A.* **31**, 1695–1697
101. Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613
102. Parrinello, M. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182
103. Kandt, C., Ash, W. L., and Peter Tieleman, D. (2007) Setting up and running molecular dynamics simulations of membrane proteins. *Methods.* **41**, 475–488

104. Levitt, M. and A. W. (1975) Computational Simulation of protein folding. *Nat. Publ. Gr.*
105. Kar, P., and Feig, M. (2014) Recent advances in transferable coarse-grained modeling of proteins. *Adv. Protein Chem. Struct. Biol.* **96**, 143–180
106. Bradley, R., and Radhakrishnan, R. (2013) Coarse-Grained Models for Protein-Cell Membrane Interactions. *Polymers (Basel)*. **5**, 890–936
107. Shih, A. Y., Arkhipov, A., Freddolino, P. L., and Schulten, K. (2006) Coarse Grained Protein -- Lipid Model with Application to Lipoprotein Particles. *J. Phys. Chem. B.* **110**, 3674–3684
108. Bond, P. J., and Sansom, M. S. P. (2006) Insertion and Sssembly of Membrane Proteins via Simulation. *J. Am. Chem. Soc.* **128**, 2697–2704
109. Kar, P., and Feig, M. (2014) Recent Advances in Transferable Coarse-Grained Modeling of Proteins, pp. 143–180, 10.1016/bs.apcsb.2014.06.005
110. Wassenaar, T. A., Pluhackova, K., Böckmann, R. A., Marrink, S. J., and Tieleman, D. P. (2014) Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **10**, 676–690
111. Mustafa, G., Nandekar, P. P., Yu, X., and Wade, R. C. (2015) On the application of the MARTINI coarse-grained model to immersion of a protein in a phospholipid bilayer. *J. Chem. Phys.* **143**, 243139
112. Yu, X., Cojocaru, V., Mustafa, G., Salo-Ahen, O. M. H., Lepesheva, G. I., and Wade, R. C. (2015) Dynamics of CYP51: implications for function and inhibitor design. *J. Mol. Recognit.* **28**, 59–73
113. Yano, J. K., Wester, M. R., Schoch, G. a., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05 resolution. *J. Biol. Chem.* **279**, 38091–38094
114. Frey, B., and David, D. (1987) Determination of the Membrane Topology of the Phenobarbital-inducible Rat Liver Cytochrome P-450 Isoenzyme 4 Using Site-specific Antibodies. *J. Cell Biol.* **104**, 209–219
115. Black, S. D., Martin, S. T., and Smith, C. A. (1994) Membrane Topology of Liver Microsomal Cytochrome P450 2B4 Determined via Monoclonal Antibodies Directed to the Halt-Transfer Signal. *Biochemistry.* **33**, 6945–6951
116. Ohta, Y., Kawato, S., Tagashira, H., Takemori, S., and Kominami, S. (1992) Dynamic structures of adrenocortical cytochrome P-450 in proteoliposomes and microsomes: protein rotation study. *Biochemistry.* **31**, 12680–12687
117. Baylon, J. L., Lenov, I. L., Sligar, S. G., and Tajkhorshid, E. (2013) Characterizing

the membrane-bound state of cytochrome P450 3A4: Structure, depth of insertion, and orientation. *J. Am. Chem. Soc.* **135**, 8542–8551

118. Dror, R. O., Arlow, D. H., Borhani, D. W., Jensen, M. O., Piana, S., and Shaw, D. E. (2009) Identification of two distinct inactive conformations of the 2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci.* **106**, 4689–4694
119. Periolo, X., Huber, T., Marrink, S.-J., and Sakmar, T. P. (2007) G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers. *J. Am. Chem. Soc.* **129**, 10126–10132
120. Vanni, S., Neri, M., Tavernelli, I., and Rothlisberger, U. (2009) Observation of “Ionic Lock” Formation in Molecular Dynamics Simulations of Wild-Type β 1 and β 2 Adrenergic Receptors. *Biochemistry.* **48**, 4789–4797
121. Lomize, M. a., Lomize, A. L., Pogozheva, I. D., and Mosberg, H. I. (2006) OPM: Orientations of proteins in membranes database. *Bioinformatics.* **22**, 623–625
122. Gábor E. Tusnády, Z. D. and I. S. (2004) Transmembrane proteins in the Protein Data Bank: Identification and classification. *Bioinformatics.* **20**, 2964–2972
123. Ash, W. L., Zlomislic, M. R., Oloo, E. O., and Tieleman, D. P. (2004) Computer simulations of membrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1666**, 158–189
124. Spijker, P., van Hoof, B., Debertrand, M., Markvoort, A. J., Vaidehi, N., and Hilbers, P. a J. (2010) Coarse grained molecular dynamics simulations of transmembrane protein-lipid systems. *Int. J. Mol. Sci.* **11**, 2393–2420
125. Noid, W. G. (2013) Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **139**, 90901
126. Bradley, R., and Radhakrishnan, R. (2013) Coarse-Grained Models for Protein-Cell Membrane Interactions. *Polymers (Basel).* **5**, 890–936
127. Ward, A. B., Guvench, O., and Hills, R. D. (2012) Coarse grain lipid-protein molecular interactions and diffusion with MsbA flippase. *Proteins Struct. Funct. Bioinforma.* **80**, 2178–2190
128. Li, S., Zhang, X., and Wang, W. (2009) Coarse-grained model for mechanosensitive ion channels. *J. Phys. Chem. B.* **113**, 14431–14438
129. Carpenter, T., Bond, P. J., Khalid, S., and Sansom, M. S. P. (2008) Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel. *Biophys. J.* **95**, 3790–3801
130. Thorpe, I. F., Zhou, J., and Voth, G. a. (2008) Peptide folding using multiscale coarse-grained models. *J. Phys. Chem. B.* **112**, 13079–13090

131. Smeijers, a. F., Pieterse, K., Markvoort, a. J., and Hilbers, P. a J. (2006) Coarse-grained transmembrane proteins: Hydrophobic matching, aggregation, and their effect on fusion. *J. Phys. Chem. B.* **110**, 13614–13623
132. Scott, K. a., Bond, P. J., Ivetac, A., Chetwynd, A. P., Khalid, S., and Sansom, M. S. P. (2008) Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly. *Structure.* **16**, 621–630
133. Sansom, M. S. P., Scott, K. a, and Bond, P. J. (2008) Coarse-grained simulation: a high-throughput computational approach to membrane proteins. *Biochem. Soc. Trans.* **36**, 27–32
134. Lonsdale, R., Rouse, S. L., Sansom, M. S. P., and Mulholland, A. J. (2014) A Multiscale Approach to Modelling Drug Metabolism by Membrane-Bound Cytochrome P450 Enzymes. *PLoS Comput. Biol.* **10**, e1003714
135. Tan, Z., Spasic, A., and Mathews, D. H. (2015) 96 Benchmark of Amber ff12SB force field by comparison of estimated hairpin loop folding stabilities to experiments. *J. Biomol. Struct. Dyn.* **33**, 61–62
136. Dickson, C. J., Madej, B. D., Skjevik, Å. a., Betz, R. M., Teigen, K., Gould, I. R., and Walker, R. C. (2014) Lipid14: The amber lipid force field. *J. Chem. Theory Comput.* **10**, 865–879
137. Harris, D. L., Park, J.-Y., Gruenke, L., and Waskell, L. (2004) Theoretical study of the ligand-CYP2B4 complexes: effect of structure on binding free energies and heme spin state. *Proteins.* **55**, 895–914
138. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341
139. Mori, T., Ogushi, F., and Sugita, Y. (2012) Analysis of lipid surface area in protein-membrane systems combining voronoi tessellation and monte carlo integration methods. *J. Comput. Chem.* **33**, 286–293
140. Steffen, C., Thomas, K., Huniar, U., Hellweg, A., Rubner, O., and Schroer, A. (2010) GridMAT-MD: A Grid-Based Membrane Analysis Tool for Use With Molecular Dynamics. *J. Comput. Chem.* **31**, 2967–2970
141. Kučerka, N., Nieh, M. P., and Katsaras, J. (2011) Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochim. Biophys. Acta - Biomembr.* **1808**, 2761–2771
142. Kučerka, N., Tristram-Nagle, S., and Nagle, J. F. (2006) Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Membr. Biol.* **208**, 193–202

143. Navrátilová, V., Paloncýová, M., Kajšová, M., Berka, K., and Otyepka, M. (2015) Effect of Cholesterol on the Structure of Membrane-Attached Cytochrome P450 3A4. *J. Chem. Inf. Model.* **55**, 628–635
144. Navrátilová, V., Paloncýová, M., Berka, K., and Otyepka, M. (2016) Effect of Lipid Charge on Membrane Immersion of Cytochrome P450 3A4. *J. Phys. Chem. B.* **120**, 11205–11213
145. Wiener, M. C., Suter, R. M., and Nagle, J. F. (1989) Structure of the fully hydrated gel phase of dipalmitoylphosphatidylcholine. *Biophys. J.* **55**, 315–325
146. Sevrioukova, I. F., and Poulos, T. L. (2010) Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18422–18427
147. Sevrioukova, I. F., and Poulos, T. L. (2017) Structural basis for regiospecific midazolam oxidation by human cytochrome P450 3A4. *Proc. Natl. Acad. Sci.* **114**, 486–491
148. Furuta, T., Ohashi, K., Kamata, T., Takashima, M., Kosuge, K., Kawasaki, T., Hanai, H., Kubota, T., Ishizaki, T., and Kaneko, E. (1998) Effect of genetic differences in omeprazole metabolism on cure rates for *Helicobacter pylori* infection and peptic ulcer. *Ann. Intern. Med.* **129**, 1027–1030
149. Wester, M. R., Yano, J. K., Schoch, G. a, Yang, C., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) The Structure of Human Cytochrome P450 2C9 Complexed with Flurbiprofen at 2.0-Å Resolution. *J. Biol. Chem.* **279**, 35630–35637
150. Nugent, T., Ward, S., and Jones, D. T. (2011) The MEMPACK alpha-helical transmembrane protein structure prediction server. *Bioinformatics.* **27**, 1438–1439
151. Ibeanu, G. C., Ghanayem, B. I., Linko, P., Li, L., Pedersen, L. G., and Goldstein, J. A. (1996) Identification of residues 99, 220, and 221 of human cytochrome P450 2C19 as key determinants of omeprazole hydroxylase activity. *J. Biol. Chem.* **271**, 12496–12501
152. Tsao, C.-C., Wester, M. R., Ghanayem, B., Coulter, S. J., Chanas, B., Johnson, E. F., and Goldstein, J. A. (2001) Identification of Human CYP2C19 Residues that Confer S-Mephenytoin 4'-Hydroxylation Activity to CYP2C9 †. *Biochemistry.* **40**, 1937–1944
153. Attia, T. Z., Yamashita, T., Hammad, M. A., Hayasaki, A., Sato, T., Miyamoto, M., Yasuhara, Y., Nakamura, T., Kagawa, Y., Tsujino, H., Omar, M. A., Abdelmageed, O. H., Derayea, S. M., and Uno, T. (2014) Effect of Cytochrome P450 2C19 and 2C9 Amino Acid Residues 72 and 241 on Metabolism of Tricyclic Antidepressant Drugs. *Chem. Pharm. Bull.* **62**, 176–181

154. Lee, A. G. (2004) How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1666**, 62–87
155. Spijker, P., van Hoof, B., Debertrand, M., Markvoort, A. J., Vaidehi, N., and Hilbers, P. A. J. (2010) Coarse Grained Molecular Dynamics Simulations of Transmembrane Protein-Lipid Systems. *Int. J. Mol. Sci.* **11**, 2393–2420
156. Scott, K. A., Bond, P. J., Ivetac, A., Chetwynd, A. P., Khalid, S., and Sansom, M. S. P. (2008) Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly. *Structure.* **16**, 621–630
157. de Jesus, A. J., and Allen, T. W. (2013) The determinants of hydrophobic mismatch response for transmembrane helices. *Biochim. Biophys. Acta - Biomembr.* **1828**, 851–863
158. Kučerka, N., Nagle, J. F., Sachs, J. N., Feller, S. E., Pencer, J., Jackson, A., and Katsaras, J. (2008) Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophys. J.* **95**, 2356–2367
159. Nagle, J. F., and Tristram-Nagle, S. (2000) Structure of lipid bilayers. *Biochim. Biophys. Acta - Rev. Biomembr.* **1469**, 159–195
160. Leekumjorn, S., and Sum, A. K. (2007) Molecular characterization of gel and liquid-crystalline structures of fully hydrated POPC and POPE bilayers. *J. Phys. Chem. B.* **111**, 6026–6033
161. Caracciolo, G., Pozzi, D., and Caminiti, R. (2007) Hydration effect on the structure of dioleoylphosphatidylcholine bilayers. *Appl. Phys. Lett.* **90**, 5–8
162. Pabst, G., Rappolt, M., Amenitsch, H., and Laggner, P. (2000) Structural information from multilamellar liposomes at full hydration: Full q-range fitting with high quality x-ray data. *Phys. Rev. E.* **52**, 4000–4009
163. Sevcsik, E., Pabst, G., Jilek, A., and Lohner, K. (2007) How lipids influence the mode of action of membrane-active peptides. *Biochim. Biophys. Acta - Biomembr.* **1768**, 2586–2595
164. Vermeer, L. S., De Groot, B. L., Réat, V., Milon, A., and Czaplicki, J. (2007) Acyl chain order parameter profiles in phospholipid bilayers: Computation from molecular dynamics simulations and comparison with ²H NMR experiments. *Eur. Biophys. J.* **36**, 919–931
165. Perly, B., Smith, I. C., and Jarrell, H. C. (1985) Acyl chain dynamics of phosphatidylethanolamines containing oleic acid and dihydrosterculic acid: ²H NMR relaxation studies. *Biochemistry.* **24**, 4659–4665
166. Seelig, J., and Waespe-Sarcevic, N. (1978) Molecular order in cis and trans unsaturated phospholipid bilayers. *Biochemistry.* **17**, 3310–3315

167. Jójárt, B., and Martinek, T. A. (2007) Performance of the general amber force field in modeling aqueous POPC membrane bilayers. *J. Comput. Chem.* **28**, 2051–2058
168. Nussio, M. R., Voelcker, N. H., Miners, J. O., Lewis, B. C., Sykes, M. J., and Shapter, J. G. (2010) AFM study of the interaction of cytochrome P450 2C9 with phospholipid bilayers. *Chem. Phys. Lipids*. 10.1016/j.chemphyslip.2009.11.003
169. Bayburt, T. H., and Sligar, S. G. (2002) Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6725–6730
170. De Lemos-Chiarandini, C. (1987) Determination of the membrane topology of the phenobarbital-inducible rat liver cytochrome P-450 isoenzyme PB-4 using site-specific antibodies. *J. Cell Biol.* **104**, 209–219
171. Bayburt, T. H., and Sligar, S. G. (2002) Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks. *Proc. Natl. Acad. Sci.* **99**, 6725–6730
172. MacCallum, J. L., Bennett, W. F. D., and Tieleman, D. P. (2008) Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **94**, 3393–3404
173. Morita, M., Katta, A. M., Ahmad, S., Mori, T., Sugita, Y., and Mizuguchi, K. (2011) Lipid recognition propensities of amino acids in membrane proteins from atomic resolution data. *BMC Biophys.* **4**, 21
174. Berka, K., Paloncýová, M., Anzenbacher, P., and Otyepka, M. (2013) Behavior of Human Cytochromes P450 on Lipid Membranes. *J. Phys. Chem. B.* **117**, 11556–11564
175. Barnes, H. J., Arlotto, M. P., and Waterman, M. R. (1991) Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **88**, 5597–5601
176. Imai, T., Globerman, H., Gertner, J. M., Kagawa, N., and Waterman, M. R. (1993) Expression and purification of functional human 17 alpha-hydroxylase/17,20-lyase (P450c17) in *Escherichia coli*. Use of this system for study of a novel form of combined 17 alpha-hydroxylase/17,20-lyase deficiency. *J. Biol. Chem.* **268**, 19681–9
177. Shukla, A., Huang, W., Depaz, I. M., and Gillam, E. M. J. (2009) Membrane integration of recombinant human P450 forms. *Xenobiotica.* **39**, 495–507
178. Pernecky, S. J., Larson, J. R., Philpott, R. M., and Coon, J. (1993) Expression of truncated forms of liver microsomal P450 cytochromes 2B4 and 2E1 in *Escherichia coli*: Influence of NH₂-terminal region on localization in cytosol and

membranes. *Proc Natl Acad Sci U S A.* **90**, 2651–2655

179. Yamamoto, K., Gildenberg, M., Ahuja, S., Im, S.-C., Pearcy, P., Waskell, L., and Ramamoorthy, A. (2013) Probing the Transmembrane Structure and Topology of Microsomal Cytochrome-P450 by Solid-State NMR on Temperature-Resistant Bicelles. *Sci. Rep.* **3**, 2556
180. Huang, R., Yamamoto, K., Zhang, M., Popovych, N., Hung, I., Im, S.-C., Gan, Z., Waskell, L., and Ramamoorthy, A. (2014) Probing the Transmembrane Structure and Dynamics of Microsomal NADPH-cytochrome P450 oxidoreductase by Solid-State NMR. *Biophys. J.* **106**, 2126–2133
181. Lundeen, M., Chance, B., and Powers, L. (1987) The Transmembrane Helices of Beef Heart Cytochrome Oxidase. *Biophys. J.* **51**, 693–695
182. Ulmschneider, M. B., Tieleman, D. P., and Sansom, M. S. P. (2004) Interactions of a transmembrane helix and a membrane: Comparative simulations of bacteriorhodopsin helix A. *J. Phys. Chem. B.* **108**, 10149–10159
183. Szczesna-Skorupa, E., and Kemper, B. (2000) Endoplasmic Reticulum Retention Determinants in the Transmembrane and Linker Domains of Cytochrome P450 2C1. *J. Biol. Chem.* **275**, 19409–19415
184. Drin, G., and Antonny, B. (2010) Amphipathic helices and membrane curvature. *FEBS Lett.* **584**, 1840–1847
185. DeVore, N. M., and Scott, E. E. (2012) Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. *Nature.* **482**, 116–119
186. Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N., and Rost, B. (2014) PredictProtein - An open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* **42**, 337–343
187. Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, 349–357
188. Cui, Y., Xue, Q., Zheng, Q., Zhang, J.-L., Kong, C., Fan, J., and Zhang, H. (2015) Structural features and dynamic investigations of the membrane-bound cytochrome P450 17A1. *Biochim. Biophys. Acta - Biomembr.* **1848**, 2013–2021
189. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., and Lomize, A. L. (2012) OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, 370–376

190. Cui, Y. L., Xue, Q., Zheng, Q. C., Zhang, J. L., Kong, C. P., Fan, J. R., and Zhang, H. X. (2015) Structural features and dynamic investigations of the membrane-bound cytochrome P450 17A1. *Biochim. Biophys. Acta - Biomembr.* **1848**, 2013–2021
191. Yamamoto, K., Caporini, M. A., Im, S.-C., Waskell, L., and Ramamoorthy, A. (2017) Transmembrane Interactions of Full-length Mammalian Bitopic Cytochrome-P450-Cytochrome-b5 Complex in Lipid Bilayers Revealed by Sensitivity-Enhanced Dynamic Nuclear Polarization Solid-state NMR Spectroscopy. *Sci. Rep.* **7**, 4116
192. Yu, X., Cojocaru, V., Mustafa, G., Salo-Ahen, O. M. H., Lepesheva, G. I., and Wade, R. C. (2015) Dynamics of CYP51: implications for function and inhibitor design. *J. Mol. Recognit.* **28**, 59–73
193. Mustafa, G., Nandekar, P. P., Yu, X., and Wade, R. C. (2015) On the application of the MARTINI coarse-grained model to immersion of a protein in a phospholipid bilayer. *J. Chem. Phys.* 10.1063/1.4936909
194. IVAN LENKOV LENOV (2016) *METHODOLOGIES FOR THE ANALYSIS OF MEMBRANE SYSTEMS USING LIPID NANODISCS*. Ph.D. thesis, University of Illinois at Urbana-Champaign, 2016
195. Cui, Y.-L., Zheng, Q.-C., Zhang, J.-L., Xue, Q., Wang, Y., and Zhang, H.-X. (2013) Molecular dynamic investigations of the mutational effects on structural characteristics and tunnel geometry in CYP17A1. *J. Chem. Inf. Model.* **53**, 3308–3317
196. Pechurskaya, T. a, Lukashovich, O. P., Gilep, a a, and Usanov, S. a (2008) Engineering, expression, and purification of “soluble” human cytochrome P45017alpha and its functional characterization. *Biochem. Biokhimiia.* **73**, 806–11
197. Park, J. W., Reed, J. R., Brignac-Huber, L. M., and Backes, W. L. (2014) Cytochrome P450 System Proteins Reside in Different Regions of the Endoplasmic Reticulum. *Biochem. J.* **464**, 241–249
198. Brignac-Huber, L. M., Reed, J. R., Eyer, M. K., and Backes, W. L. (2013) Relationship between CYP1A2 localization and lipid microdomain formation as a function of lipid composition. *Drug Metab. Dispos.* **41**, 1896–1905
199. Yamada, A., Shimizu, N., Hikima, T., Takata, M., Kobayashi, T., and Takahashi, H. (2016) Effect of Cholesterol on the Interaction of Cytochrome P450 Substrate Drug Chlorzoxazone with the Phosphatidylcholine Bilayer. *Biochemistry.* **55**, 3888–3898
200. Liu, J., Tawa, G. J., and Wallqvist, A. (2013) Identifying cytochrome P450 functional networks and their allosteric regulatory elements. *PLoS One.* **8**, 1–11
201. Mustafa, G., Yu, X., and Wade, R. C. (2014) Structure and Dynamics of Human

Drug-Metabolizing Cytochrome P450 Enzymes. in *Drug Metabolism Prediction* (Kirchmair, J. ed), pp. 77–102, Wiley-VCH, Weinheim

202. Hendrychová, T., Anzenbacherová, E., Hudeček, J., Skopalík, J., Lange, R., Hildebrandt, P., Otyepka, M., and Anzenbacher, P. (2011) Flexibility of human cytochrome P450 enzymes: molecular dynamics and spectroscopy reveal important function-related variations. *Biochim. Biophys. Acta.* **1814**, 58–68
203. Zhang, T., Liu, L. A., Lewis, D. F. V, and Wei, D.-Q. (2011) Long-range effects of a peripheral mutation on the enzymatic activity of cytochrome P450 1A2. *J. Chem. Inf. Model.* **51**, 1336–46
204. Tu, Y., Deshmukh, R., Sivaneri, M., and Szklarz, G. D. (2008) Application of Molecular Modeling for Prediction of Substrate Specificity in Cytochrome P450 1A2 Mutants. *Drug Metab. Dispos.* **36**, 2371–2380
205. Skopalík, J., Anzenbacher, P., and Otyepka, M. (2008) 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.* **112**, 8165–8173
206. Li, D., Huang, X., Han, K., and Zhan, C.-G. (2011) Catalytic mechanism of cytochrome P450 for 5'-hydroxylation of nicotine: fundamental reaction pathways and stereoselectivity. *J. Am. Chem. Soc.* **133**, 7416–27
207. Xu, Y., Shen, Z., Shen, J., Liu, G., Li, W., and Tang, Y. (2011) Computational insights into the different catalytic activities of CYP2A13 and CYP2A6 on NNK. *J. Mol. Graph. Model.* **30**, 1–9
208. Hendrychova, T., Berka, K., Navratilova, V., Anzenbacher, P., and Otyepka, M. (2012) Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. *Curr. Drug Metab.* **13**, 177–89
209. Banu, H., Renuka, N., and Vasanthakumar, G. (2011) Reduced catalytic activity of human CYP2C9 natural alleles for gliclazide: molecular dynamics simulation and docking studies. *Biochimie.* **93**, 1028–36
210. Bikádi, Z., and Hazai, E. (2008) In silico description of differential enantioselectivity in methoxychlor O-demethylation by CYP2C enzymes. *Biochim. Biophys. Acta.* **1780**, 1070–9
211. Seifert, A., Tatzel, S., Schmid, R. D., and Pleiss, J. (2006) Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: The molecular basis of substrate binding and regioselectivity toward warfarin. *Proteins Struct. Funct. Bioinforma.* **64**, 147–155
212. Sano, E., Li, W., Yuki, H., Liu, X., Furihata, T., Kobayashi, K., Chiba, K., Neya, S., and Hoshino, T. (2010) Mechanism of the decrease in catalytic activity of human

- cytochrome P450 2C9 polymorphic variants investigated by computational analysis. *J. Comput. Chem.* **31**, 2746–58
213. Zhou, Y.-H., Zheng, Q.-C., Li, Z.-S., Zhang, Y., Sun, M., Sun, C.-C., Si, D., Cai, L., Guo, Y., and Zhou, H. (2006) On the human CYP2C9*13 variant activity reduction: a molecular dynamics simulation and docking study. *Biochimie.* **88**, 1457–65
214. Berka, K., Anzenbacherová, E., Hendrychová, T., Lange, R., Mašek, V., Anzenbacher, P., and Otyepka, M. (2012) Binding of quinidine radically increases the stability and decreases the flexibility of the cytochrome P450 2D6 active site. *J. Inorg. Biochem.* **110**, 46–50
215. Li, J., Wei, D.-Q., Wang, J.-F., Yu, Z.-T., and Chou, K.-C. (2012) Molecular dynamics simulations of CYP2E1. *Med. Chem.* **8**, 208–21
216. Yuki, H., Honma, T., Hata, M., and Hoshino, T. (2012) Prediction of sites of metabolism in a substrate molecule, instanced by carbamazepine oxidation by CYP3A4. *Bioorg. Med. Chem.* **20**, 775–83
217. Shahrokh, K., Cheatham, T. E., and Yost, G. S. (2012) Conformational dynamics of CYP3A4 demonstrate the important role of Arg212 coupled with the opening of ingress, egress and solvent channels to dehydrogenation of 4-hydroxy-tamoxifen. *Biochim. Biophys. Acta.* **1820**, 1605–17
218. Teixeira, V. H., Ribeiro, V., and Martel, P. J. (2010) Analysis of binding modes of ligands to multiple conformations of CYP3A4. *Biochim. Biophys. Acta.* **1804**, 2036–45

APPENDIX I

Table A 1 A list of *in silico* studies on human drug metabolizing CYPs is taken from the book chapter on “Drug Metabolism Prediction” (201). The data presented in the table was compiled in July 2014.

CYP	PDB	RESOLUTION (Å)	LIGAND [#]	TUNNELS	METHOD ⁺	REFERENCE	YEAR
1A2	2HI4	1.95	----- ^b	2, S	UV/VIS	(202)	2011
					MD		
	Model		7-ethoxy 7-methoxyresorufin	1, 2a, 2b, 2e, 2f, 5, S	MD	(203)	2011
2A6	1Z10	1.9	Coumarin	2e, S	Modeling	(204)	2008
					Docking		
	1Z11	2.05	Methoxsalen (S)-(-)-nicotine		MD MM-PBSA QM/MM	(206)	2011

	1Z10	1.90					
	1Z11	2.05					
	2FDU	1.85			Docking		
	2FDV	1.65	NNK		MD	(207)	2011
	2FDW	2.05			MM-GBSA		
	2FDY	1.95					
	1Z10	1.90	Coumarin	2c, 6, 3, 2a, 2e, S	MD, RAMD SMD	(47)	2011
	1Z10	1.90	Coumarin	2b, 2e	MD	(208)	2012
	1OG2	2.60	----- ^b	2e, S	MD	(205)	2008
2C9	1OG2	2.6	----- ^b	2, S	UV/VIS MD	(202)	2011
	1R90	1.9	Gliclazide		Modeling MD	(209)	2011

	1R90	1.9	Methoxychlor		MD	(210)	2008
	1R90	2.0	----- ^b	2a, 2c, S	Modeling MD ^{mem}	(31)	2011
	1OG2	2.6	----- ^b	2b, 2c, S	Modeling MD ^{mem}	(32)	2011
	1OG5	2.7					
	1OG2	2.6	WAR, WAR-OH	2a/2b, 2c, 2e,	MD	(46)	2012
	1R90	2.0	FLU, FLU-OH	2d, 2f, S, 3	RAMD		
	1OG2	2.6	----- ^b	2a, 2b, 2e, S	MD	(208)	2012
	1OG2	2.6					
	1OG5	2.7	Warfarin		MD	(211)	2006
			Diclofenac				
2C9*2			Mefnamic acid		Modeling		
*3	1OG2	2.6			MD	(212)	2010
*5 ^A			Phenytoin				
			Piroxicam		Docking		

			S-Warfarin				
			Tenoxicam				
			Tolbutamide				
					Modeling		
2C9*13 ^A	1R90	1.9	Diclofenac	S	Docking	(213)	2006
			Lornoxicam		MD		
	2F9Q	3.0	----- ^b	2, S	UV/VIS	(202)	2011
					MD		
2D6	2F9Q	3.0	----- ^b		UV/VIS	(214)	2012
			Quinidine		MD		
	2F9Q	3.0	----- ^b	S	MD	(208)	2012
	3E6I	2.2	----- ^b	2ac, 2e	MD	(208)	2012
			Indazol				
2E1	3E6I	2.2		2a, 2b, 2c, 2f, 3, S, W, 1, 2d	MD RAMD SMD	(37)	2012

2E1	3E4E	2.20	4-methylpyrazole BNZ, ANI, AAP, CLZ & THP		Modeling Docking MD	(215)	2012
	1TQN	2.05			Modeling		
	1W0E	2.80	----- ^b		Docking	(216)	2012
	1W0F	2.65	CBZ		MD		
	1W0G	2.73					
3A4	1TQN	2.05	----- ^b	2, S	UV/VIS MD	(202)	2011
	1TQN	2.05			Modeling		
	1W0E	2.08	4-OHT	2b, 4, S	Docking	(217)	2012
					MD, QM		
	1TQN	2.05	----- ^b	2e, S	MD	(205)	2008
	1W0F	2.6	VARIOUS LIGANDS ^c		DOCKING MD	(218)	2010

a, CYP2C9*2,*3,*5 and *13 are the allelic variants of wild type CYP2C9 with single-point mutations R144C, I359L, D360E and L90P, respectively.

b, In the ligand column, “----” is used to show ligand free structures of CYPs, either from removal of the ligand from original structure or from solving the crystal structure without a ligand (apo form).

c, Various ligands are docked which are given in Table 1 of Teixeira et al., 2010.

#, The ligand names are abbreviated as follows: ANF (α -naphthoflavone), NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butane), CBZ (carbamazepine), TMZ (temazepam), TST-OH (hydroxylated testosterone), WAR and WAR-OH (warfarin and its hydroxylated product), FLU and FLU-OH (flurbiprofen and its hydroxylated product), BNZ (benzene), ANI (aniline), AAP (acetaminophen), CLZ (chlorzoxazone), THP (theophylline), 4OH-T (4-hydroxy-temoxifen).

+, In the methods column, abbreviations for the experimental and computational methods are given which indicate, UV/VIS: Ultra-violet/Visible and resonance Raman spectroscopy, MD: molecular dynamics simulations, Modeling: either homology modeling or modeling of missing residues in the crystal structure, Docking: ligand docking, RAMD: random accelerated molecular dynamics, SMD: steered molecular dynamics, QM; quantum mechanics, and MM-PBSA/GBSA: Molecular Mechanics/Poisson-Boltzmann Surface Area/Generalized Born Surface Area calculation of free energies. The superscript ‘mem’ denotes that the simulations have been performed with a bilayer present. All other simulations were performed for the proteins in water without any membrane bilayer.