

Computational Studies of Transmembrane Proteins

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte
Dissertation

von

M.Sc. Deepak Kumar
aus Bihar/Indien

1. Erster Gutachter: Dr. Tomáš Kubař
2. Zweiter Gutachter: Prof. Dr. Alexander Schug
Tag der mündlichen Prüfung: 20. Oktober 2022

To all unprivileged people around the world who do not have access to good education and health care.

“Education is the most powerful weapon which you can use to change the world.” - Nelson Mandela

Abstract

Interactions between membrane proteins or peptides play a central role in many biochemical processes such as signal reception, enzyme inhibition, transport of various solutes across the membrane, and protein activation by signal transduction. The purpose of the present study was to examine the structural, physical, and atomistic properties of biomolecules embedded in phospholipid bilayers with the aid of computer-based simulations.

The first part of this work aims to provide mechanistic details about water transfer and proton transfer reactions in the tetrameric charge zipper TisB assembly of peptides embedded in a bilayer. The formation of persister cells are among the most important strategies employed by a bacterial population to get rid of antibiotic treatment and external stress. *E. coli* overproduces the amphiphilic α -helical peptide TisB (toxic protein) into the inner membrane, which further leads to the formation of biofilms. Consequently, the electrochemical gradient on the membrane of the bacterial cell is obstructed. We have performed a long all-atom simulation of selected structures and employed enhanced sampling simulation to study the water transfer facilitated by the polar interface of the TisB assembly. Further, we have employed QM/MM simulations to study the detailed molecular mechanism and energetics of the transport of protons.

The second part of this work focuses on the structural analysis of the tetrameric and hexameric E5-PDGFR β complex. Through highly specific interactions between the trans-membrane helices, the platelet-derived growth factor receptor PDGFR β is activated by the oncoprotein E5. We performed 2 μ s long classical MD simulation with three different sets of protocols. Our MD simulation results were complimented by experimental findings from the group of Prof. Dr. Anne S. Ulrich from KIT. The proposed tetrameric and hexameric models were suggested to form E5-PDGFR β heterocomplexes, through specific interaction of Thr513-Gln17 and Lys499-Asp33, which is also supported by our MD simulation. Our simulation results match perfectly with the experimental observation.

The third part of this work focuses on the aquaporins derived from the Tonoplast Intrinsic Proteins family (TIP). These proteins are proposed to be responsible for facilitating the permeation of water and ammonia across the vacuolar membrane of plants. The experiment suggests a lower permeability of water and ammonia across gel phase lipid through aquaporin AtTip2;1 as compared with liquid phase lipid. We have performed 3 μ s long classical MD simulation in two phases of lipid. Each phase of lipid consists of three batches of simulations: (1) pure water system, (2) system containing 1000 NH₃, and (3) system containing 1000 NH₄⁺. This was also supported by our MD simulation, which was consistent with what we saw in the experimental data, showing a lower permeability of

NH₃ and H₂O in gel phase lipid. The permeability of NH₃ across AtTIP2;1 in gel phase lipid is substantially blocked in our simulation, which is also in line with the experiments. Though NH₄⁺ ions did not pass through the AtTIP2;1 our simulation shows a decrease in the permeability of the water by half in comparison to the permeability of water in the pure water system in the liquid phase, which is further supported by experiments.

Zusammenfassung

Wechselwirkungen zwischen Membranproteinen oder -peptiden spielen eine zentrale Rolle bei einer Vielzahl von biochemischen Prozessen wie Signalempfang, Enzyminhibition, dem Transport verschiedener gelöster Stoffe durch die Membran und der Aktivierung von Proteinen durch Signaltransduktion. Ziel der vorliegenden Studie war es, die strukturellen, physikalischen und atomistischen Eigenschaften von Biomolekülen, die in Phospholipid-Doppelschichten eingebettet sind, mit Hilfe von computergestützten Simulationen zu untersuchen.

Der erste Teil dieser Arbeit zielt darauf ab, mechanistische Details über Wassertransfer- und Protonentransferreaktionen in der tetrameren Ladungszipper-TisB-Assemblierung von Peptiden, die in eine Doppelschicht eingebettet sind, zu liefern. Die Bildung von Persister-Zellen gehört zu den wichtigsten Strategien von Bakterienpopulationen, um sich einer Antibiotikabehandlung und äußerem Stress zu entziehen. *E. coli* überproduziert das amphiphile α -helicale Peptid TisB (toxisches Protein) in die innere Membran, was wiederum zur Bildung von Biofilmen führt. Folglich wird der elektrochemische Gradient an der Membran der Bakterienzelle gestört. Wir haben eine lange All-Atom Simulation ausgewählter Strukturen durchgeführt und eine Enhanced Sampling Simulation eingesetzt, um den Wassertransfer zu untersuchen, der durch die polare Schnittstelle der TisB-Anordnung erleichtert wird. Außerdem haben wir QM/MM-Simulationen eingesetzt, um den detaillierten molekularen Mechanismus und die Energetik des Protonentransports zu untersuchen.

Der zweite Teil dieser Arbeit konzentriert sich auf die Strukturanalyse des tetrameren und hexameren E5-PDGFR β -Komplexes. Durch hochspezifische Wechselwirkungen zwischen den Transmembran-Helices wird der Platelet-Derived Growth Factor-Rezeptor PDGFR β durch das Onkoprotein E5 aktiviert. Wir haben 2 μ s lange klassische MD-Simulationen mit drei verschiedenen Protokollen durchgeführt. Unsere MD-Simulationsergebnisse wurden durch experimentelle Ergebnisse aus der Gruppe von Prof. Dr. Anne S. Ulrich vom KIT ergänzt. Die vorgeschlagenen tetrameren und hexameren Modelle bilden E5-PDGFR β -Heterokomplexe durch spezifische Wechselwirkung von Thr513-Gln17 und Lys499-Asp33, was auch durch unsere MD-Simulation bestätigt wird. Unsere Simulationsergebnisse stimmen perfekt mit den experimentellen Beobachtungen überein.

Der dritte Teil dieser Arbeit befasst sich mit den Aquaporinen aus der Familie der Tonoplast Intrinsic Proteins (TIP). Diese Proteine sollen für eine erleichterte Permeation von Wasser und Ammoniak durch die Vakuolenmembran von Pflanzen verantwortlich sein. Das Experiment deutet darauf hin, dass die Durchlässigkeit von Wasser und Ammoniak durch Aquaporin AtTip2;1 in der Gelphase geringer ist als in der Flüssigphase. Wir haben 3 μ s lange klassische MD-Simulationen in zwei Lipidphasen durchgeführt. Zu jeder Lipidphase

wurden drei Sätze von Simulationen durchgeführt: (1) reines Wassersystem, (2) System mit 1000 NH_3 und (3) System mit 1000 NH_4^+ . Im Gegensatz zu den experimentellen Beobachtungen zeigt unsere MD-Simulation auch eine geringere Durchlässigkeit von NH_3 und H_2O in der Gelphase des Lipids. Die Permeabilität von NH_3 durch AtTIP2;1 in der Gelphase des Lipids ist in unserer Simulation erheblich blockiert, was ebenfalls mit experimentellen Daten übereinstimmt. Obwohl NH_4^+ -Ionen das AtTIP2;1 nicht durchdringen, zeigt unsere Simulation eine Verringerung der Permeabilität des Wassers um die Hälfte im Vergleich zum reinen Wassersystem in der Flüssigphase, was auch durch Experimenten gestützt wird.

Acknowledgment

To begin with, I would like to express my deepest gratitude to Prof. Marcus Elstner and Dr. Tomáš Kubař for believing in me and giving me the opportunity to pursue my PhD at Karlsruhe Institute of Technology, Karlsruhe, and for giving me the opportunity to flourish as a researcher and individual. Thank you for giving me the opportunity to work on several interesting research projects throughout my PhD over the past three years.

I would also like to extend my thanks to my collaborator from Prof. Anne S. Ulrich's group from KIT for giving me the opportunity to work on some interesting interdisciplinary projects. I would especially like to thank Dr. Li Tian, Dr. Stephan Grage, Dr. Sebastian Otteni, Dr. Parvesh Wadhvani, and Dr. Torsten Walther for their support and contribution throughout the projects.

I would also like to extend my thanks to my collaborator from Johannes Kepler University, Linz, Dr. Andreas Horner, and Mr. Thomas Barta for their exciting, fruitful, and invaluable contributions. I would also like to acknowledge Dr. Kristyna Pluhackova from the University of Stuttgart for their help and suggestion in setting up the simulation and analysis.

All this would not have been possible without funding from the German Research Foundation (DFG). I am also extremely grateful to GRK2450 for its financial support. I would also like to acknowledge all the members of the GRK2450 for organising many interesting hands-on tutorials, seminar days, and workshops throughout the year. I would like to thank and acknowledge Deniz, Samaneh, and Leticia for their help and support in organising these activities.

Nothing is possible without paperwork and here comes one of the most important people in this entire journey, Sabine Holthoff. She has been constant support from the moment I got the offer for the PhD. Thank you for making my life easier and smooth with your help and assistance.

I would also like to extend my thanks to all the members of TCB group for creating a nice and friendly working atmosphere. It has been a pleasure seeing you all on campus west and chatting over lunch. The last six months have been the best time in the last three years. Thank you for making this place more fun and happening. Special thanks to the new members (Katharina Spies, Lena Eichinger, Lukas Petersen, Christian Schmidt) of TCB for invigorating a new life in the place. Thank you all for reading my thesis and giving important suggestions and helping in designing the content. Special thanks to Fathia, Denis, Philipp, David, Monja, Farhad, and Lukas for taking me out of your busy schedule.

Acknowledgment

It is also my pleasure to acknowledge Dr. Denis Maag for his helpful suggestions and advice throughout this project. I appreciate all the technical support you provided and the answers to my silly questions. I wish I could become a researcher like you and adopt your work ethic.

I would like to express my deepest love and respect for Katharina Spies for bringing new life and energy to this place. If it was a wonderful past six months, the main reason was you. You were not only friendly and talkative but someone who likes to sit together and eat. Thank you for all your time and long conversation. Thanks for always being there. Some people are unforgettable and you are one of them. If I am talking about food, how can I forget Lena, the shy girl for spending time over food and long conversations in the kitchen?

Apart from being a wonderful PI, how can I forget the special impact Dr. Tomáš Kubař has made in my personal and professional career! Loneliness is/was my biggest enemy and during bad COVID times, the situation worsened. Then there was one person who always wanted to make things normal. It was Tomáš Kubař, who is always there as a mentor, support system, and the one who tries to understand your problem. Thank you for all the hiking, biking, and trekking tours. Thank you for always listening to me and giving me the simplest solution or peace of your mind. Thank you for teaching me the value of speaking slowly in simple words. If Prof. Marcus Elstner was the reason to bring me here to Germany, then Dr. Tomáš Kubař is the reason to stay in Germany and keep me moving.

It would be unfair if I don't acknowledge the support given by my friends Michael Kasana, Luis I Vazquez, Prakash, Kajal, Tushar, Vicky, Navneet, Reetu, Sanjhal, Srishti, Ahsana for being there whenever needed. It must be very difficult without these wonderful people in your life.

In particular, I would like to thank Poorwa Kunwar, who has greatly influenced me in many ways. Thank you for being a support system in countries away from home and being so kind and caring to me. If someone asks me, what did you earn apart from my PhD in Germany, I would say Poorwa Kunwar.

I would also like to thank Gulshan Bhatia and Neelam Bhatia for their help and support and for taking care of me while I was writing my thesis. I would also like to thank my social media friend Ms. Nada Hzl, someone whom I have never met, but someone who cares and keeps motivating you.

Finally, I would also like to thank my family for their relentless support and motivation. This could not have been possible without their sacrifice, encouragement, and hunger to achieve and grow together. Thanks to each one of my family members for their contribution to making me a better human being and keeping me connected to my roots. The main credit for my entire education goes to my sisters (Aparajita and Mamta), who had no choice but to give up their education so that their brothers can study and grow. I am highly indebted to their sacrifice.

Contents

Abstract	v
Zusammenfassung	vii
Acknowledgment	ix
List of Figures	xv
List of Tables	xvii
I. Introduction	1
1. Introduction	3
1.1. Background	3
1.2. Categories of Membrane Proteins	4
1.2.1. α -Helix	4
1.2.2. β -Barrels	5
1.3. Function of Transmembrane Proteins	6
1.3.1. Transporters	6
1.3.2. Ion Channels	6
1.3.3. Receptors	7
II. Theoretical and Computational Background	9
2. Computational Chemistry	11
2.1. Introduction	11
2.2. Tools of Computational Chemistry	12
2.3. Combining all the Information	13
3. Quantum Chemistry	15
3.1. Schrödinger Equation	15
3.2. Born-Oppenheimer Approximation	17
3.3. Variational Principle	18
3.4. Hartree-Fock Theory	18
3.4.1. Hamiltonian	19
3.4.2. Energy Expression	19
3.4.3. Basis Set Approximation	20

3.4.4.	Electron Correlation	21
3.5.	Density Based Methods	21
3.5.1.	Density Functional Theory (DFT)	21
3.5.2.	The Hohenberg-Kohn Theorems	22
3.5.3.	The Kohn-Sham Approach	22
3.5.4.	Local Density Approximation	24
3.5.5.	Generalised Gradient Approximation	24
3.6.	Density-Functional Tight-Binding	25
3.6.1.	DFTB1	26
3.6.2.	DFTB2	27
3.6.3.	DFTB3	27
4.	Classical Mechanics and Molecular Dynamics Simulations	29
4.1.	Perspective	29
4.2.	Principles of Molecular Mechanics	30
4.2.1.	Bonding Interaction	31
4.2.2.	Non-bonding Interaction	32
4.2.3.	Complete Force Field Equation	33
4.3.	Propagation of a System	34
4.3.1.	Verlet Integration Method	34
4.3.2.	Leap Frog Integration Method	35
4.3.3.	Velocity Verlet Integration Method	35
4.4.	Periodic Boundary Conditions	36
4.5.	Temperature and Pressure Control	37
4.5.1.	Thermostat	37
4.5.2.	Barostat	37
4.6.	Intermolecular Interactions	38
4.7.	Hybrid QM/MM Simulations	39
4.7.1.	Mechanical embedding	40
4.7.2.	Electrostatic embedding	40
4.7.3.	Polarisation embedding	41
5.	Enhanced Sampling Techniques in Molecular Dynamics	43
5.1.	Introduction	43
5.2.	Collective Variable	45
5.3.	Potential of Mean Force	45
5.4.	Umbrella Sampling	46
5.5.	Metadynamics	47
5.5.1.	Standard Metadynamics	47
5.5.2.	Well-Tempered Metadynamics	48
5.5.3.	Multiple-Walker Metadynamics	49

III. Contributions	51
6. Water and Proton Transport Across Tetrameric Charge Zipper Protein	53
6.1. Introduction	53
6.1.1. Interaction between TisB Peptide	53
6.2. Methodology	55
6.2.1. Classical MD Simulation	55
6.2.2. QM/MM Simulation	56
6.2.3. Water Transfer – Classical Metadynamics	58
6.2.4. Proton Transfer – QM/MM Metadynamics	59
6.3. Results and Discussion	60
6.3.1. Water Transport	60
6.3.2. Proton Transfer or Proton Hole Transfer	65
7. Structural Analysis of E5-PDGFRβ Transmembrane Proteins	71
7.1. Introduction	71
7.1.1. Tyrosine Kinases Receptor	71
7.1.2. Oncoprotein E5	73
7.1.3. PDGF-Receptor β	74
7.1.4. E5-PDGFR Complex	74
7.2. Methodology	75
7.3. Results and Discussion	78
7.3.1. RMSD and RMSF	78
7.3.2. Helix Tilt and Rotational Angle	83
7.3.3. Order Parameter (S)	89
7.3.4. PISEMA Calculation	94
7.3.5. Distance Measurement	98
8. Water and Ammonia Selectivity Through Aquaporin AtTIP2;1	107
8.1. Introduction	107
8.1.1. Crystal Structure of AtTIP2;1	109
8.2. Methodology	110
8.2.1. Force field, simulation setup and equilibration of aquaporin AtTIP2;1 in liquid phase lipid	110
8.2.2. Simulation setup and equilibration of aquaporin AtTIP2;1 embedded in gel phase lipid	113
8.2.3. Simulations of AtTIP2;1 in the presence of ammonia or ammonium	114
8.3. Results	114
8.3.1. Water passage in liquid and gel phase lipid in pure water system	114
8.3.2. Ammonia passage in liquid and gel phase lipid in the system containing 1000 NH ₃ molecules	116
8.3.3. Water passage in liquid and gel phase lipid in the system containing 1000 NH ₃ molecules	117
8.3.4. Water passage in liquid and gel phase lipid in the system containing 1000 NH ₄ ⁺ molecules	119

8.4. Discussion	121
8.4.1. Water Permeability and Selectivity	121
8.4.2. Ammonia Selectivity	129
8.4.3. Ammonium Ion Blocking	132
8.5. Comparison with Experiment	134
9. Summary	135
Bibliography	139
IV. Appendix	153
10. Water and Proton Transport Across Tetrameric Charge Zipper Protein	155
10.1. Potential Gating mechanism at the center of the peptide	155
11. Structural Analysis of E5-PDGFR Transmembrane Proteins	159
11.1. Order parameter of N-H bond in tetrameric complexes	159
11.2. Order parameter of N-H bond in hexameric complexes	162
11.3. Distance measurements in tetrameric complexes	165
11.4. Distance measurements in hexameric complexes	167
12. Water and Ammonia Selectivity Through Aquaporin AtTIP2;1	169
12.1. Permeation of water (H ₂ O) and ammonia (NH ₃)	169
12.2. Distance measurement between important residues across the channels .	172
12.3. Time Evolution of Distances	173
12.4. Torsional angle analysis of the bulky residues	178

List of Figures

1.1.	Schematic representation of a membrane protein.	4
1.2.	Schematic representation of functions performed by integral proteins.	6
2.1.	Hierarchy of various computational chemistry methods	14
4.1.	Illustration of various contributions of force field parameters.	31
4.2.	Schematic representation of the periodic boundary conditions	36
5.1.	Graphical illustration of the potential of mean force	45
5.2.	Graphical illustration of umbrella sampling	46
5.3.	Graphical representation of metadynamics	48
6.1.	Representation of the anti-parallel arrangements of the TisB dimer	54
6.2.	Snapshots of TisB	55
6.3.	Insight view of QM region	57
6.4.	Graphical representation of the distance (d_z)	59
6.5.	Snapshots of the QM region in acidic and basic medium	60
6.6.	Water density of TisB	61
6.7.	Number of water molecules within the TisB assembly	62
6.8.	Hydrogen bonds within the subdimers in the TisB assembly	63
6.9.	Gibbs energy ΔG of water across the bilayer normal	64
6.10.	Free energy diagram of proton transfer in acidic medium (using ζ as CV)	66
6.11.	Free energy diagram of proton transfer in acidic medium (using z-component of ξ as CV)	67
6.12.	Free energy diagram of proton transfer in basic medium (ζ as CV)	68
6.13.	Free energy diagram of proton transfer in basic medium (z-component of ξ as CV)	68
7.1.	Schematic representation of receptor tyrosine kinases (RTK)	72
7.2.	Models for E5-PDGFR β complex	76
7.3.	Backbone RMSD analysis of 4-helix and 6-helix complexes - Starting structure	80
7.4.	Backbone RMSD analysis of 4-helix and 6-helix complexes - Average structure	81
7.5.	RMSF scan along all atoms in tetrameric complexes	82
7.6.	RMSF scan along all atoms in hexameric complexes	83
7.7.	Average structure of 4-helix and 6-helix complexes	84
7.8.	Schematic representation of tilt and roll angle	85
7.9.	Order parameter measurement from MD simulation	90
7.10.	N-H order parameter along a pair of E5 helices of the 4-helix complex	92
7.11.	N-H order parameter along two PDGFR β helices of the 4-helix complex	92

7.12. N-H order parameter along two pairs of E5 helices of 6-helix complex . . .	93
7.13. N-H order parameter along two PDGFR β helices of 6-helix complex . . .	93
7.14. Parameters used to calculate chemical shifts and dipolar splittings	95
7.15. Experimental measurements of the PISA wheel of E5 and PDGFR chain .	96
7.16. PISA wheel analysis of the 4-helix complex	97
7.17. PISA wheel analysis of the 6-helix complex	99
7.18. Labelling strategy of E5 and PDGFR β -TMD	100
8.1. Structure of <i>At</i> TIP2;1	107
8.2. Channel structure of <i>At</i> TIP2;1	108
8.3. Selective Filter composition of TIP2;1	110
8.4. Representative snapshots for the simulation setups	111
8.5. Disordered lipid in liquid and gel phase lipid	113
8.6. Average distance in liquid and gel phase lipid	123
8.7. Distribution of HSD131 dihedral angle	125
8.8. Distribution of HSD81 dihedral angle	126
8.9. Average distance – HSD81 and GLU24	128
8.10. Representative snapshots of the open and closed state of the channel . .	129
8.11. Selective filter region	130
8.12. Average distance – ALA134 and PHE192	131
8.13. Snapshots of NH $_4^+$ ions accumulation on the vascular surface	132
8.14. Snapshot of electrostatic interaction between the selective filter	133
10.1. Potential Gating mechanism – Q19 residues and D22	155
10.2. Potential Gating mechanism – Q19 residues	156
10.3. Potential Gating mechanism – Q19	157
11.1. N-H Order parameter along a pair of E5 helices – Tetramer and SP1 . . .	159
11.2. N-H Order parameter along two flanking PDGFR β – Tetramer and SP1 .	160
11.3. N-H Order parameter along a pair of E5 helices – Tetramer and SP2 . . .	160
11.4. N-H Order parameter along two flanking PDGFR β – Tetramer and SP2 .	161
11.5. N-H Order parameter along a pair of E5 helices – Hexamer and SP1 . . .	162
11.6. N-H Order parameter along two flanking PDGFR β – Hexamer and SP1 .	163
11.7. N-H Order parameter along a pair of E5 helices – Hexamer and SP2 . . .	163
11.8. N-H Order parameter along two flanking PDGFR β – Hexamer and SP2 .	164
12.1. Average distance – GLU24 and SER78	177

List of Tables

7.1.	Lengths of simulations and force constants of position and dihedral restraints	77
7.2.	Tilt and azimuthal angle for the tetramer	86
7.3.	Tilt and azimuthal angle for the hexamer	87
7.4.	Angles between the individual helices of the tetramer	88
7.5.	Angles between the individual helices of the hexamer	89
7.6.	Distance between pairs of amino acids side chains of the tetramer	102
7.7.	Distance between pairs of amino acids side chains of the hexamer	103
8.1.	NVTs equilibration steps, lengths of simulations and force constants	112
8.2.	Passage of water across the aquaporin <i>AtTIP2;1</i> – Pure water	115
8.3.	Passage of NH ₃ across the aquaporin <i>AtTIP2;1</i> – NH ₃ molecules	117
8.4.	Passage of water across the aquaporin <i>AtTIP2;1</i> – NH ₃ molecules	118
8.5.	Passage of water across the aquaporin <i>AtTIP2;1</i> – NH ₄ ⁺ molecules	120
8.6.	Average distance – HSD63 and ARG200	122
8.7.	Average distance – HSD131 and ARG200	124
8.8.	Average distance – GLU24 and HSD81	127
8.9.	Average distance – GLU24 and SER78	127
8.10.	Comparison of H ₂ O and NH ₃ passage	134
11.1.	Distance measurements of tetramer – SP1	165
11.2.	Distance measurements of tetramer – SP2	166
11.3.	Distance measurements of hexamer – SP1	167
11.4.	Distance measurements of hexamer – SP2	168
12.1.	Water passage across the aquaporin <i>AtTIP2;1</i>	169
12.2.	NH ₃ passage across the aquaporin <i>AtTIP2;1</i>	170
12.3.	Water passage across the aquaporin <i>AtTIP2;1</i>	170
12.4.	Water passage across the aquaporin <i>AtTIP2;1</i>	171
12.5.	Average distance – GLU146 and THR150	172
12.6.	Average distance – GLU146 and SER192	172
12.7.	Average distance – GLU146 and SER195	172
12.8.	Average distance – GLU146 and THR150	173
12.9.	Average distance – GLU24 and GLN180	173
12.10.	Average distance – ASN197 and ALA199	173
12.11.	Average distance – ASN197 and VAL83	173
12.12.	Average distance – ASN83 and THR87	174
12.13.	Average distance – ALA214 and ASP205	174
12.14.	Average distance – ALA134 and PHE192	174
12.15.	Average distance – ALA50 and TYR42	174

12.16.Average distance – ALA199 and SER112	175
12.17.Average distance – GLY80 and SER78	175
12.18.Average distance – GLY193 and HSD131	175
12.19.Average distance – LEU109 and THR113	175
12.20.Average distance – SER212 and ASP210	176
12.21.Average distance – TYR217 and SER192	176
12.22.Average torsion angle – ARG200	178
12.23.Average torsion angle – ASN83	178
12.24.Average torsion angle – ASN197	179
12.25.Average torsion angle – GLU24	179
12.26.Average torsion angle – GLU146	180
12.27.Average torsion angle – HSD63	180
12.28.Average torsion angle – HSD81	181
12.29.Average torsion angle – HSD131	181
12.30.Average torsion angle – HSD214	182
12.31.Average torsion angle – MET196	182
12.32.Average torsion angle – TYR42	183
12.33.Average torsion angle – TYR104	183
12.34.Average torsion angle – TYR158	184
12.35.Average torsion angle – TYR217	184

Part I.
Introduction

1. Introduction

1.1. Background

Cell membranes, also known as plasma membranes, are thin membranes that surround every living cell and form a barrier between the cell and its environment. There are two functions performed by the cell membrane. The first is to be a barrier that keeps the components of the cell inside and unwanted substances outside, and the second is to be a gate to allow the transport of essential nutrients into the cell and the removal of waste products from within the cell. Proteins and lipids, which are composed primarily of fatty acids, are the major components of the membranes of cells. Membrane lipids primarily consist of two types: phospholipids and sterols (generally cholesterol). The biomembranes also consist of glycolipids and glycoproteins, which contain carbohydrates. The carbohydrate component is external to the cell and plays an important role in cell identification. As one of the defining characteristics of phospholipids, the molecules are easily dissolved in organic solvents, but they also have a region by which they are attracted to and soluble in water, a characteristic that is common to both types.

Aqueous solutes cannot pass through the membrane because of the hydrophobic barrier created by lipids. In a semipermeable membrane, there can be a presence of proteins that allows the movement of aqueous solutes across the membrane. In addition to transport functions, membrane proteins have roles as receptors, enzymes, and structural components. Membrane proteins have a surface that is amphiphilic, meaning that they can interact with both the polar exteriors of membranes, as well as the internal nonpolar surfaces of membranes. Because membrane proteins need to maintain an amphiphilic structure, the process of purifying and handling them is much more difficult. In general, there are two types of membrane proteins. One type of protein referred to as extrinsic protein or peripheral protein, is loosely attached to the phosphoryl surface of the bilayer by ionic bonds or calcium bridges, making it possible to form electrical connections. Extrinsic proteins (also known as peripheral membrane proteins) can be washed off without disrupting the membrane. The second type of protein is called intrinsic protein [1]. It is evident from their names that intrinsic proteins are embedded firmly within the phospholipid bilayer. Integral membrane proteins interact extensively with membrane lipids and cannot be purified without disrupting the membrane using detergents or mechanical rupture. There is a higher proportion of proteins in membranes that are actively involved in metabolism compared to membranes that are not.

A number of crucial functions are performed by integral membrane proteins in all types of cells, making them of great importance. Their function is to transport polar and

non-polar solutes, metabolites and larger molecules, such as proteins and RNA, across the membrane. Furthermore, membrane proteins play a crucial role in sending and receiving chemical signals, propagating electrical impulses, attaching cells together, and anchoring other proteins to specific locations within the cell. In addition, they act as a signalling receptor for intracellular vesicular transport, regulate the composition of lipids in the membrane, and organise and maintain the shape of organelles and structures within the cell [1].

1.2. Categories of Membrane Proteins

Integral membrane proteins (see fig. 1.1) are capable of adapting to a variety of environments that differ from the cytosol in viscosity, dielectric constant as well as isotropy conditions, while also enduring gradients in pH, redox potential, and pressure. The integral proteins are sometimes regarded and described as transmembrane segments (TMSs). Further, two types of integral membrane proteins have been identified as α -helical and β -barrels. One of the most abundant and well-studied integral proteins is the α -helical membrane protein, which contains one or several α -helices.

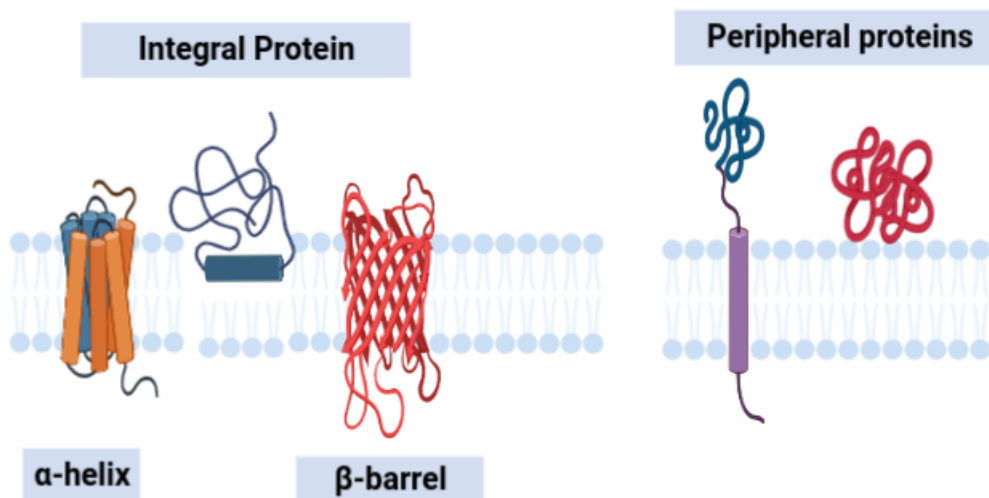


Figure 1.1.: Schematic representation of a membrane protein.

1.2.1. α -Helix

Most integral membrane proteins are composed of an α -helix of twenty or more predominantly nonpolar residues along with some polar or charged residues. It is common for charged amino acids to be positioned for functional purposes, to form electrostatic ion pairs, or to exist in their uncharged states. Alternatively, the charge or polar groups of charged or polar side chains can be orientated beyond the nonpolar core of the bilayer

by folding towards the membrane interface. Snorkeling (burying themselves with their aliphatic part facing the hydrophobic region of the lipid bilayer, while positioning the charged amino group in the more polar interface of the protein or solvent) is another phenomenon that is observed with lysine [2] and residues such as arginine, tyrosine, aspartate, glutamate, asparagine and glutamine. Additionally, polar side chains near the TMS ends form hydrogen bonds with other peptides to cap the ends of the helices [3]. Due to their electrostatic interactions with the hydrocarbon core, tryptophans and tyrosines are found at the interface of polar and nonpolar domains.

The first membrane protein observed to consist of an α -helical core was bacteriorhodopsin (bR), a light-driven proton pump from Halobacteria. BR consists of seven roughly parallel transmembrane membranes (TM) helices coupled to the retinal cofactor by covalent bonds [4]. Another example is aquaporin which is composed of a bundle of six transmembrane α -helices. Bacteriorhodopsin is quite regular and oriented fairly perpendicular to the bilayer, whereas, it is most common to find that α -helical bundle membrane proteins contain some tilted α -helices as well as TM helices with distortions due to kinks and bulges.

In this work, I have studied three such α -helix transmembrane proteins, which perform different functions for the cells. In my first research project, I worked on the tetrameric charge zipper TisB peptide, which is suggested to be involved in the proton transfer across the membrane, which I have covered in greater detail in chapter 7. My second project focused on the association of E5-PDGFR with peptides in membranes. The E5 protein plays a role in the activation of platelet-derived growth factor receptor (PDGFR β) in a ligand-independent manner by transmembrane helix-helix interactions. This part has been discussed in detail in chapter 8. In my third project, I worked on the water and ammonia selectivity through aquaporin *AfTIP2;1* in the liquid phase and gel phase lipid bilayer, which I have covered in great detail in chapter 9.

1.2.2. β -Barrels

Membrane-spanning barrels are found in outer membranes of bacteria and mitochondria as well as chloroplasts, within which they perform several functions such as transporters, phage binding structures, catalytic agents, and adhesion molecules. In spite of the wide variety of morphologies of monomers and oligomers as well as an open barrel to the tight packing of their interiors, most of them form pores that enable ions to disperse across their surfaces. β -barrels typically consist of 8 to 26 amphiphilic β -strands crossing the bilayer at $\sim 45^\circ$ and adjacent β -sheet are hydrogen bonded to each other. Each barrel consists of 9 to 11 residues hydrogen-bonded to adjacent strands. There can be an exception including a VDAC (voltage-dependent anion channel), a mitochondrial β -barrel, which has an odd number of strands [5]. It's been found that porins, a family of trimers with three open pores, vary in their selectivity towards ions and solutes due to the structural characteristics of their pores [6]. Larger barrels facilitate diffusion of specific large solutes containing an internal plug domain.

1.3. Function of Transmembrane Proteins

Numerous genetic, biochemical, and biophysical studies of membrane proteins have shed light on their molecular functions as seen in fig. 1.2. It is the architecture and the gating mechanisms of channels that illuminate the specificity of the integral proteins. Transporters from widely varying families share fundamental transport mechanisms. Mechanisms of membrane enzymes can be uncovered by capturing intermediates in their reaction cycles.

1.3.1. Transporters

In essence, membrane transport proteins are proteins that are involved with the movement of ions, small molecules, and macromolecules, for example another protein, across a biological membrane. There are two routes by which proteins may assist in the movement of substances, either through passive or active transport. There is a major difference between a channel protein and a carrier protein. In the case of carrier protein, both the extracellular and intracellular environments are not accessible simultaneously. Either the inner gate or the outer gate is open. A channel, on the other hand, can be open simultaneously to both environments, allowing molecules to diffuse unrestricted. In general, carrier proteins have binding sites, whereas channel proteins do not have binding sites.

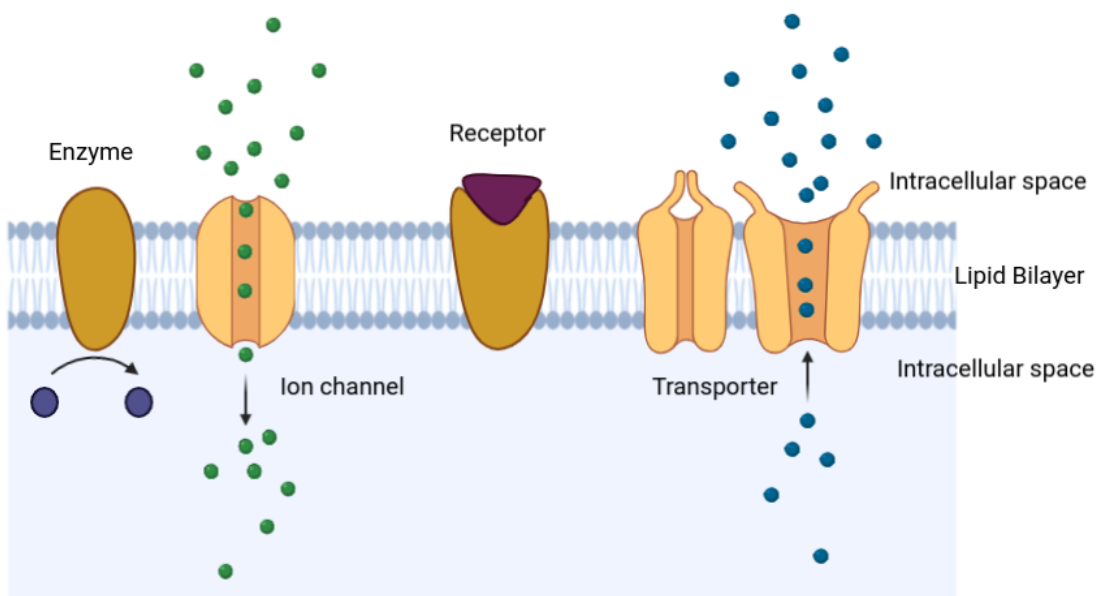


Figure 1.2.: Schematic representation of functions performed by integral proteins.

1.3.2. Ion Channels

Ion channels are proteins that construct pores in membranes and are responsible for allowing ions (such as Ca^{2+} , Na^{1+}) to pass through the channels. By regulating ion flow

across the membrane and controlling the flow of ions across secretory and epithelial cells, they establish a resting membrane potential and shape action potentials and other electrical signals. Ion channels differ from other types of ion transporters in two main ways [7]:

1. The rate of ion transport through the channel is very high.
2. Ions pass through channels down their electrochemical gradient.

These ion channels are often described as narrow water-filled tunnels that allow only selective ions to pass through. This characteristic is called selective permeability.

1.3.3. Receptors

There is a variety of chemical structures, consisting of protein molecules, that are involved in receiving and converting signals that may be integrated into biological systems through receptors [8]. As the name suggests, these signals are typically chemical messengers, which bind to a receptor. This triggers a response in a cell or tissue, such as an increase in its electrical activity, which leads to an increase in its metabolic activity. The action of the receptor can be classified into three main categories based on the manner in which it transmits signals, amplifies signals, and integrates signals.

Part II.

Theoretical and Computational Background

2. Computational Chemistry

Computational chemistry has a direct impact on advancing our understanding of chemistry as a manifestation of the behaviour of atoms and molecules, as well as seeing these entities as real entities rather than mere convenient models of human behaviour.

2.1. Introduction

The term computational chemistry (or molecular modelling; the two terms are basically the same) describes a collection of techniques that can be applied to the analysis and investigation of chemical problems using computers. With the help of computational chemistry, biochemical processes can be investigated at an atomistic level. A theoretical chemistry method is implemented into efficient computer programs to predict the structure and calculate a wide range of properties of small molecules or complex systems such as proteins and enzymes. These are some of the questions that are commonly investigated using computational methods:

- An understanding of molecular geometry involves understanding the shapes of molecules – bond lengths, angles, and dihedral angles.
- We can determine the equilibrium isomer by calculating the energies of the molecules and their transition states, as well as how fast the reaction will proceed (based on the energy of the transition state and the reactant energies).
- A comprehensive understanding of chemical reactivity, for instance, enables us to predict where various kinds of reagents will attack a molecule depending on where electrons are concentrating (nucleophilic sites) and where they desire to go (electrophilic sites).
- It is possible to design better drugs by observing how a molecule interacts with an enzyme's active site.
- Physical properties of substances are determined by the properties of individual molecules as well as how the molecules interact in the bulk materials. An example would be how well molecules fit together and how strong the forces between them are by determining the strength and melting point of a polymer (for example, a plastic).

2.2. Tools of Computational Chemistry

A wide range of methods is available that computational chemists can employ in studying these questions in order to gain a greater understanding of them. There are five broad categories of tools available:

1. **Ab initio:** Typically, *ab initio* calculations (*ab initio* is the Latin term that means "from the beginning", i.e. "from the first principle") are based on Schrödinger's equation. Modern physics relies heavily on this equation to explain, among other things, how electrons behave in molecules. We can calculate the energy and wave function (often written as wavefunction) of a molecule using the *ab initio* method by solving the Schrödinger equation. Electron distribution can be calculated using the wavefunction, a mathematical function. As a result of the electron distribution, we are able to determine for example the polarity of a molecule and the regions of the molecule that are likely to be attacked by nucleophiles and electrophiles [9].
2. **Semiempirical methods (SE):** In the same way as *ab initio* methods, semiempirical methods rely on Schrödinger's equation. The semiempirical method, however, uses more approximate solutions and does not evaluate the very complex integrals required for the *ab initio* method. In order to do this, the program uses a library of integrals which is searched for the best fit of some calculation entity like geometry or energy (such as heat of formation) to the experimental values in order to determine the best integrals. A semiempirical method is based on a combination of theory and experiment: the Schrödinger equation is used but parameterised according to experimental data (empirical means experimental). The process of plugging experimental values into a mathematical method in order to obtain the most accurate values is called parameterisation [10].
3. **Density functional theory (DFT):** Similarly to *ab initio* and semiempirical calculations, density functional calculations are also based on the Schrödinger equation, which is of course a fundamental component. There is one important difference between DFT and the other two methods – it does not utilise a conventional wavefunction but instead uses an electron density. This reduces the cost of computational calculations. Functionals are mathematical entities that relate to functions. Calculations based on density functionals are typically faster than *ab initio* calculations, but slower than those based on semiempirical methods. DFT usually scales with N^3 , so larger systems can be calculated over longer periods of time [11].
4. **Molecular mechanics (MM):** It is a theory based on the idea that molecules are composed of a collection of balls (atoms) connected by springs (bonds). In order to calculate the energy of any given collection of balls and springs, we need to know the spring lengths and angles between those, as well as the amount of energy used to stretch and bend these springs. It is not possible to simulate chemical reactions in which bonds are broken or formed, using molecular mechanics. By using molecular mechanics, potential functions are generated that are not absolute but rather can be used as a comparison tool between different configurations of molecules. As a result

of the efficiency of molecular mechanics, a fairly large molecule, such as a steroid, protein, or enzyme, can be optimised in seconds by a good personal computer.

5. **Molecular dynamics (MD):** The laws of motion are applied to molecules in molecular dynamics calculations. It is possible to simulate very large systems for several μs using force field methods because the computational effort is low. However, this comes at the expense of accuracy. The application of this method allows one to simulate the motion of an enzyme as it changes shape when binding to the substrate or the motion of water molecules surrounding a solute if it changes its shape. Molecular dynamics can also be used for simulating actual chemical reactions by combination with quantum mechanical principles [12].

2.3. Combining all the Information

In order to investigate the structure and behaviour of very large biological molecules, molecular mechanics is often employed, since other methods (quantum mechanical methods based on the Schrödinger equation) would take a long time. New molecules with unusual structures are preferably calculated with *ab initio* or DFT calculations due to the parameterisation inherent in MM or semiempirical methods, as it is difficult when dealing with molecules with very different structures compared to those used in the parameterisation. A separate parameterisation can be run under such situations.

A number of methods are available for calculating the energies of molecules, including MM, SE, *ab initio*, and DFT. Methods are selected based on the nature of the problem being addressed. Quantum mechanical methods are usually used to study reactivity, which is largely determined by electron distribution (SE, *ab initio*, or DFT). In order to calculate IR spectra accurately, *ab initio* or DFT methods are the most reliable, although SE spectra can be calculated in a useful manner with some MM programs. It is beneficial to use computational chemistry in order to investigate the properties of materials, i.e. in materials science. Computational chemistry has been utilised to investigate semiconductors, superconductors, plastics, and ceramics. A solid-state physics background and some specialised knowledge are usually necessary for such studies.

Computational chemistry is often used to see how a molecule fits into an enzyme's active site. The substrate can be docked using a mouse and a feedback device so that the forces acting on the molecule being docked can be felt. Automated docking is now standard practice. These docking experiments serve as a basis for the design of better drugs, small molecules that will interact more effectively with the target enzymes that are desired but will be ignored by other enzymes.

In fig. 2.1, there is a summary of the feasible system sizes and simulation times of the described methods. In chapter 3, a more detailed explanation of *ab initio* methods, DFT, and semiempirical DFTB methods will be provided. Chapter 4 describes in detail how the force field methods based on all atoms and molecular dynamics can be used. It is possible to accelerate rare events using enhanced sampling techniques when a process or

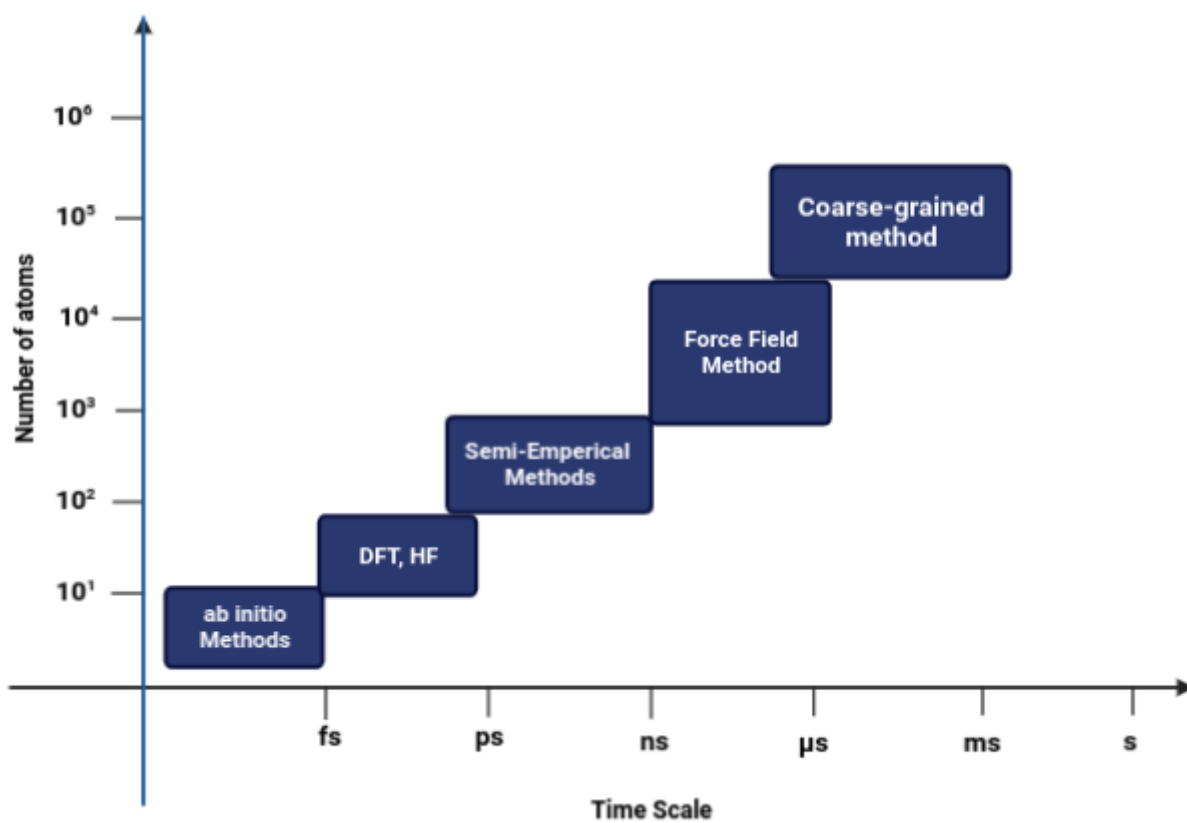


Figure 2.1.: Hierarchy of various computational chemistry methods based on system size and time scale.

mechanism exceeds the timescale of the available method.

3. Quantum Chemistry

Background

The concept of molecular mechanics is not always adequate to describe molecular systems in their proper context. It is necessary to use quantum mechanic calculations in order to simulate chemical reactions, e.g. bond breaking and bond formation, proton transfer, and many other biochemical processes. Through the use of computational methods, we are able to study biochemical processes at the atomic level. Biological processes involve a variety of spatial and temporal scales. From the absorption of light at femtoseconds, to the catalysis of an enzyme at seconds. Multiscale methods have been developed to link these scales, and thus provide a more accurate description of these processes. Biological processes require the use of a wide range of time scales to understand since from the absorption of light at femtoseconds to the catalysis of an enzyme at seconds, multiscale methods should be used for this purpose. To accurately describe biochemical reactions occurring in biological systems, *ab initio* methods are needed. Because of the high computational cost, *ab initio* methods are limited by the number of atoms (100-200). As a result, semiempirical methods are required when handling larger systems and when analysing structures that have been sampled.

Molecular dynamic (MD) simulations have been used over the years as a way to elucidate the structures and thermodynamic properties of proteins by sampling them based on force field methods, which have many benefits. There is a big advantage in using classical MD simulations in describing biological systems, a strength that comes from the fact that the computational effort involved is much lower than it is in *ab initio* simulations, which is particularly useful for large molecules. For example, a simulation of a system consisting of up to 100 000 atoms can be carried out for a period of several milliseconds. Several coarse-grained methods are available for the description of even larger systems or for longer periods of time. A range of methods, including *ab initio*, semiempirical, and MD simulations, will be discussed in this thesis.

3.1. Schrödinger Equation

Schrödinger's equation corresponds to Newton's second law of classical mechanics in quantum mechanics. In quantum mechanics, the Schrödinger equation allows us to predict the evolution over time of a wave function, which is a characterisation of an isolated physical system based on its quantum mechanical properties. For obtaining quantum mechanical information of stationary states, such as the total energy of a molecule, the non-relativistic time-independent Schrödinger equation is usually used.

$$\hat{H}(\mathbf{r})\Psi(\mathbf{r}) = E(\mathbf{r})\Psi(\mathbf{r}) \quad (3.1)$$

$$\hat{H}(\mathbf{r}) = \hat{T} + \hat{V}(\mathbf{r}) \quad (3.2)$$

Here, \hat{H} is a Hamiltonian operator consisting of the kinetic energy operator (\hat{T}) and the potential energy operator (\hat{V}). Their evolution in time is described by the time-dependent Schrödinger equation:

$$\hat{H}(\mathbf{r}, t)\Psi(\mathbf{r}, t) = i\hbar \frac{\partial \psi(\mathbf{r}, t)}{\partial t} \quad (3.3)$$

$$\hat{H}(\mathbf{r}, t) = \hat{T} + \hat{V}(\mathbf{r}, t) \quad (3.4)$$

All the terms have the usual meaning as described previously. \hbar is the reduced Planck constant and $\psi(\mathbf{r}, t)$ is a wave function, a function that assigns a complex number to each point \mathbf{r} at each time t . $V(\mathbf{r}, t)$ is the potential that represents the environment in which the particle exists and i is the imaginary unit. It is possible to describe the system as a function of time by solving the Schrödinger equation, which produces the wave function. Probability density functions are defined as the product of the square of the absolute value of the wave function at each point of the wave function.

$$\mathbf{P}(\mathbf{r}, t) = |\Psi(\mathbf{r}, t)|^2 \quad (3.5)$$

For a system of N nuclei and n electrons, the molecular Hamilton operator \hat{H} consists of five contributions:

$$\begin{aligned} \hat{H} = & \underbrace{-\sum_A^N \frac{\hbar^2}{2M_A} \nabla_A^2}_{\hat{T}_N} - \underbrace{\sum_i^n \frac{\hbar^2}{2m_e} \nabla_i^2}_{\hat{T}_e} \\ & + \underbrace{\sum_A^N \sum_{B>A}^N \frac{Z_A Z_B e^2}{4\pi\epsilon_0 \mathbf{r}_{AB}}}_{\hat{V}_{NN}} - \underbrace{\sum_A^N \sum_i^n \frac{Z_A e^2}{4\pi\epsilon_0 \mathbf{r}_{Ai}}}_{\hat{V}_{Ne}} + \underbrace{\sum_i^n \sum_{j>i}^n \frac{e^2}{4\pi\epsilon_0 \mathbf{r}_{ij}}}_{\hat{V}_{ee}} \end{aligned} \quad (3.6)$$

Z is the atomic number and M the mass of the nuclei, and r is the respective distance between particles. The first two terms \hat{T}_N and \hat{T}_e are the kinetic energy operators of the nuclei and electrons, \hat{V}_{NN} the potential energy operator for repulsive nucleus-nucleus interactions, \hat{V}_{Ne} the potential energy operator for attractive interactions between nuclei and electrons and \hat{V}_{ee} the potential energy operator for repulsive electron-electron interactions.

3.2. Born-Oppenheimer Approximation

It can be said that Born-Oppenheimer approximation is one of the most fundamental principles that underlies the description of quantum states of molecules in general. As a result of this approximation, the motions of the nuclei and the electrons can be separated from one another.

Using the Born-Oppenheimer approximation, the motions of electrons in a molecule are taken into account, while the motion of the atomic nuclei is neglected. The Born-Oppenheimer approximation has its physical basis in the fact that the mass of an atomic nucleus in a molecule is several thousand times larger than the mass of an electron. As a result, the nuclei move much more slowly than the electrons. Due to their opposite charges, the nucleus of an atom and an electron are also attracted to each other by a mutually attractive force as shown in equation 3.6. Both particles are accelerated because of the force between them. In general, the acceleration is inversely proportional to the mass of the particle ($a \propto \frac{1}{m}$), therefore the electron acceleration is much larger than the nuclear acceleration.

So in order to describe the electronic states of a molecule reasonably, it would be helpful to assume that the nuclei are stationary. In essence, electrons react almost instantly to changes in nuclear positions and are therefore considered to be moving in a field of fixed nuclei. This is known as the Born-Oppenheimer approximation. Because of the fixed positions of the nuclei, the kinetic energy of the nuclei is zero and the strength of the nucleus-nucleus repulsion for a given geometry is constant. In the case of a diatomic molecule, for example, the Hamiltonian operator is made up of three terms:

$$\hat{H}(\mathbf{r}, R) = \hat{T}_{nuc}(R) + \frac{e^2}{4\pi\epsilon_0} \frac{Z_A Z_B}{R} + \hat{H}_{elec}(\mathbf{r}, R) \quad (3.7)$$

$$\hat{T}_{nuc}(R) = -\frac{\hbar^2}{2m_A} \nabla_A^2 - \frac{\hbar^2}{2m_B} \nabla_B^2 \quad (3.8)$$

$$\hat{H}_{elec}(\mathbf{r}, R) = \frac{-\hbar^2}{2m} \sum_i \nabla_i^2 + \frac{e^2}{4\pi\epsilon_0} \left(-\sum_i \frac{Z_A}{r_{Ai}} - \sum_i \frac{Z_B}{r_{Bi}} + \frac{1}{2} \sum_i \sum_{j \neq i} \frac{1}{r_{ij}} \right) \quad (3.9)$$

An electron is approximately 2000 times lighter than a proton. Using the Born-Oppenheimer approximation, nuclear contributions can be ignored when calculating the electronic wave function and energy. Consequently the electronic wave function $\psi(\mathbf{r}, R)$ is found as the electronic Schrödinger equation.

$$\hat{H}_{elec} \Psi_{elec} = E_{elec} \Psi_{elec} \quad (3.10)$$

The total energy of the two-particle system can be calculated as follows:

$$E_{total} = E_{elec} + \frac{e^2}{4\pi\epsilon_0} \frac{Z_A Z_B}{R} \quad (3.11)$$

Although the Born-Oppenheimer approximation does not consider nuclear kinetic energy terms, the resulting electronic wavefunction is also influenced by variations in the positions of the nuclei R when determining the electronic energy.

3.3. Variational Principle

Unlike the Schrödinger equation for one-electron systems, such as H_2^+ , the electronic Schrödinger equation cannot be solved analytically for systems with many electrons. Using the variational principle, the many-body wave function Ψ is approximated best for systems with many electrons. This variational principle explains that depending on a wave function and a Hamiltonian operator, the expectation value for the binding energy of a system is always to be higher than or equal to the actual binding energy of the system, assuming an approximate wave function and an exact Hamiltonian operator are used. By adjusting one or more parameters in a wavefunction, we can find the best approximate wavefunction, also known as the trial wave function Ψ_{trial} .

$$\langle E_{trial} \rangle \geq E_{true} \quad (3.12)$$

$$\langle E_{trial} \rangle = \frac{\int \Psi_{trial}^* \hat{H} \Psi_{trial} d\tau}{\int \Psi_{trial}^* \Psi_{trial} d\tau} \quad (3.13)$$

Usually, expectation values and normalisation integrals can be calculated analytically.

3.4. Hartree-Fock Theory

An electronic wave function, Ψ , is a combination of one-electron wave functions Φ , also known as orbitals, and is made up of total electronic waves. It is necessary for the wave function to be antisymmetric with respect to the interchange of two electrons in order to comply with the Pauli principle, which states that two electrons differ in at least one quantum number.

The electronic wavefunction Ψ is constructed as a product of one-electron wavefunctions $\Phi_i(i)$ ($i=1,\dots,N$) to consider a many-electron system, which is known as Hartree product.

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \Phi_1(1)\Phi_2(2)\dots\Phi_N(N) \quad (3.14)$$

Hartree's product, however, is not in accordance with the antisymmetry principle. In order to achieve the antisymmetry, linear combinations of Hartree products are first calculated with the help of Slater determinants:

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \dots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \dots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \dots & \phi_N(N) \end{vmatrix} \quad (3.15)$$

$\frac{1}{\sqrt{N!}}$ is a normalisation factor. A Slater determinant can be used to describe the electrons through an antisymmetrised product. Using a trial wavefunction consisting of a single Slater determinant equates to the assumption that each electron moves independently except that it feels the Coulomb repulsion because of all other electrons. It is for that reason also known as the independent particle model or a mean-field theory. For the one-electron wave-functions, coordinates of electrons along the rows and columns follow orthonormality ($\langle \Phi_i | \Phi_j \rangle = \delta_{ij}$).

3.4.1. Hamiltonian

The energy of the Slater determinant is obtained by splitting the molecular Hamiltonian \hat{H} operator in the one-electron operator \hat{h}_i and the two-electron operator \hat{v}_{ij} .

$$\begin{aligned} \hat{H} &= \hat{T}_e + \hat{V}_{Ne} + \hat{V}_{ee} + \hat{V}_{NN} \\ &= \sum_i^N \hat{h}_i + \sum_{i>j}^N \hat{v}_{ij} + \sum_{A>B}^N \frac{Z_A Z_B}{r_{AB}} \end{aligned} \quad (3.16)$$

The one-electron operator \hat{h}_i and the two-electron operator \hat{v}_{ij} are given by:

$$\hat{h}_i = -\frac{1}{2} \nabla_i^2 - \sum_A \frac{Z_A}{\mathbf{r}_{iA}} \quad (3.17)$$

$$\hat{v}_{ij} = \frac{1}{\mathbf{r}_{ij}} \quad (3.18)$$

$$\hat{H}_{elec} = \sum_i \hat{h}_i + \sum_{i>j} \hat{v}_{ij} + V_{NN} \quad (3.19)$$

Since V_{NN} is constant, we can ignore it for now.

3.4.2. Energy Expression

By definition, the best possible one-electron wavefunctions will give the lowest possible total energy for a many-electron system when combined into a Slater determinant. This method can be used to calculate the expectation value of the total energy for a system when used with the complete Hamiltonian. This wavefunction of the system is called the Hartree-Fock wavefunction, and the total energy is the Hartree-Fock energy. Using the variational principle to minimise the total energy of a system, we can obtain the following set of Schrödinger-like equations that are called Hartree-Fock equations.

$$\hat{F}_i \Phi_i = \epsilon_i \Phi_i \quad (3.20)$$

\hat{F} is the Fock operator. A Hartree-Fock equation can be solved using the Fock operator, which is a one-electron operator and gives the energy for a single electron and the Hartree-Fock orbital attached to the electron. A fundamental feature of the Fock operator is, that it allows us to understand how Hartree-Fock (HF) or Self-Consistent Field (SCF) theory accounts for electron-electron interactions between atoms and molecules, while preserving the concept of atomic and molecular orbitals. It is necessary to write the full antisymmetrised wavefunction as a Slater determinant of spin orbitals in order to obtain the form of the Fock operator.

$$\hat{F} = \hat{H}^0 + \sum_{j=1}^N (\hat{J}_j - \hat{K}_j) = -\frac{\hbar^2}{2m} \nabla^2 - \frac{Ze^2}{4\pi\epsilon_0 \mathbf{r}} + \sum_{j=1}^N (\hat{J}_j - \hat{K}_j) \quad (3.21)$$

There are two terms in this equation that account for the kinetic energy of an electron and the potential energy of that electron when it interacts with the nucleus. The second term in the equation 3.21 represents the potential energy of one electron in an average field created by all the other electrons. In a many-electron system, the operators \hat{J} and \hat{K} come from electron-electron repulsion terms. The operator \hat{J}_j is known as the Coulomb operator, whereas the operator \hat{K}_j is known as the exchange operator.

$$\hat{J}_j(i) |\Phi_i(i)\rangle = \langle \Phi_j(j) | \frac{1}{r_{ij}} | \Phi_j(j)\rangle | \Phi_i(i)\rangle \quad (3.22)$$

$$\hat{K}_j(i) |\Phi_i(i)\rangle = \langle \Phi_j(j) | \frac{1}{r_{ij}} | \Phi_i(j)\rangle | \Phi_j(i)\rangle \quad (3.23)$$

Generally, the Coulomb operator \hat{J}_j is defined as the classical electron-electron repulsion (Coulomb repulsion), while the exchange operator \hat{K}_j takes into account the spin, which is based on quantum mechanics due to the antisymmetry of the Slater determinant. Increasing the flexibility of wavefunctions by adding additional parameters to the guess orbitals used in Hartree-Fock calculations is expected to result in better and better energies.

3.4.3. Basis Set Approximation

Due to the dependence of the HF equation on the MOs, the HF equation is solved by the self-consistent field (SCF) procedure, where the MOs are constructed from an initial guess and adjusted in each iteration until a convergence criterion of the total energy is met. Therefore, the molecular orbitals can be written as Linear Combination of Atomic Orbitals (LCAO) in order to solve it numerically.

$$|\Phi_i\rangle = \sum_{\alpha=1}^{M_{basis}} c_{i\alpha} |\chi_\alpha\rangle \quad (3.24)$$

where $c_{i\alpha}$ refers to the expansion coefficients and χ_α refers to the atomic spatial orbitals (AO). The set of AOs is called a basis set, which contains a Slater-type orbital (**STO**) or a

Gaussian-type orbital (GTO). The Hartree-Fock equations in the atomic orbital basis of the Roothan-Hall equation.

$$FC = SCE \quad (3.25)$$

Specifically, \mathbf{F} represents the Fock matrix elements, \mathbf{S} represents the overlap elements between the basis functions, and \mathbf{C} is the $\mathbf{N} \times \mathbf{N}$ matrix that contains the coefficients of the expansion and \mathbf{E} represents $\mathbf{N} \times \mathbf{N}$ which contains the orbital energies. In order to determine unknown MO coefficients, the Fock matrix is diagonalised. Roothan-Hall equations are iteratively solved starting with an estimation of the coefficients until the obtained set of coefficients is self-consistent within a certain threshold.

3.4.4. Electron Correlation

HF's main disadvantage is, that it fails to take into account electron correlation due to its mean-field approximation, which ignores the direct interaction between electrons. As the basis set size increases, the accuracy of the HF method improves. However, even with very large basis sets the HF method can not attain 100% accuracy due to the mean field approximation. In the mean-field approximation, each electron is treated independently, that is, it moves under the influence of the average electrostatic field created by the other electrons in the system. As a result, electron correlation is ignored, which is vital for describing chemical phenomena.

The correlation energy (E_{corr}) is given by

$$E_{corr} = E_{exact} - E_{HF} \quad (3.26)$$

Coulomb correlations are between electrons with opposite spins, while Fermi correlations are between electrons with the same spin. A linear combination of Slater determinants can be used for electron correlations.

There are various techniques developed for post-Hartree-Fock methods, that can improve the HF results by using wave functions containing more than one Slater determinant, i.e. coupled cluster or Møller–Plesset perturbation theory.

3.5. Density Based Methods

3.5.1. Density Functional Theory (DFT)

DFT is a computational method, that determines the properties of molecules by determining the electron density of the molecules. It is important to note that electron density is a physical characteristic of all molecules, in contrast to wavefunction, which is not a physical reality, but a mathematical construct. An electron density is an interpretation of ψ that it is equal to the square of the wavefunction. This system can be fully described by the electron density of the system since the integral of the electron density defines the electron number of the system, the cusps within the electron density define the position of nuclei in the system, and the heights of these cusps define the nuclear charges:

$$N = \int \rho(\mathbf{r}) d\mathbf{r} \quad (3.27)$$

3.5.2. The Hohenberg-Kohn Theorems

The core spirit of DFT is to substitute the complicated and thus hard-to-compute many-electron wavefunction, which contains $3N$ (N is the number of electrons) variables of electron density. The two Hohenberg-Kohn theorems establish a connection between $\rho(\mathbf{r})$ and the electronic molecular ground state. The theorems are as follows:

1. The first theorem provides a one-to-one mapping between the exact electron density $\rho(\mathbf{r})$ and the exact external potential $V_{\text{ext}}(\mathbf{r})$, therefore there cannot be two different V_{ext} , that yield the same ground state electron density $\rho(\mathbf{r})$. The exact external potential $V_{\text{ext}}(\mathbf{r})$ determines the exact ground state wavefunction and thus, the exact wavefunction is a functional of the ground electron density $\rho_0(r)$.

$$E_0[\rho_0(\mathbf{r})] = T[\rho_0(\mathbf{r})] + E_{ee}[\rho_0(\mathbf{r})] + E_{Ne}[\rho_0(\mathbf{r})] \quad (3.28)$$

In the case of molecules and solids, $V_{\text{ext}}(\mathbf{r})$ is the electrostatic potential of the nuclei. The ground state energy consists of the kinetic energy of the electrons, which is represented by $T[\rho_0]$, the energy of the electron-electron interaction, which is represented by $E_{ee}[\rho_0]$, and the electron-nuclei attraction, which is presented by $E_{Ne}[\rho_0]$.

2. The second Hohenberg-Kohn theorem says that a trial electron density results in higher or equal energy than an exact electron density, which is also analogous to the Raleigh-Ritz principle for wave functions, according to which the ground state wavefunction minimises the energy expectation value.

$$E[\tilde{\rho}] \geq E_0 \quad (3.29)$$

Thus, the variational principle can be used to approximate the ground state electron density $\rho_0(\mathbf{r})$.

3.5.3. The Kohn-Sham Approach

However, Hohenberg-Kohn based density calculation usually lacks accuracy due to the poor representation of the kinetic energy. To overcome this limitation, Kohn and Sham (KS) proposed splitting the functional parts of kinetic energy. The concept of a non-interacting reference system is introduced by Kohn and Sham in the form of an orbital-based system, where non-interacting electrons are placed in an effective potential of V_{eff} . It is possible to calculate a large part of the total energy using this approach. Furthermore, an approximated functional will be used to treat the non-classical contribution to the electron-electron repulsion. Thus equation 3.28 can be re-written as:

$$E[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] + E_{Ne}[\rho] \quad (3.30)$$

$$E_{XC}[\rho] \equiv T[\rho] - T_S[\rho] + E_{ee}[\rho] - J[\rho] \quad (3.31)$$

In equation 3.30, $J[\rho]$ refers to the classical Coulomb interaction between electrons and T_S refers to the kinetic energy of the non-interacting reference system, which is thus not equal to the exact kinetic energy of an interacting system ($T_S \neq T$). The remaining part of the kinetic energy and of the potential energy, the electron exchange, is transferred to the so-called exchange-correlation functional $E_{XC}[\rho]$. Further expanding the equation 3.30:

$$E[\rho(\mathbf{r})] = -\frac{1}{2} \sum_{i=1}^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle + \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{XC}[\rho] + \int V_{Ne}\rho(\mathbf{r})d\mathbf{r} \quad (3.32)$$

The electron density can be obtained from the Kohn-Sham orbitals:

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\varphi_i(\mathbf{r})|^2 \quad (3.33)$$

At this point, the effective potential V_{eff} as well as the orbitals have to be determined.

$$V_{\text{eff}}(\mathbf{r}) = V(\mathbf{r}) + \frac{1}{2} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{XC}(\mathbf{r}) \quad (3.34)$$

The effective potential V_{eff} contains the potential $V(\mathbf{r})$ of the nuclei, the Coulomb interaction between electrons and the mean-field created by all other electrons, and finally the so-far unknown potential $V_{XC}(\mathbf{r})$. E_{XC} represents the exchange-correlation energy, which is defined as the functional derivative of E_{XC} .

Finally, the one particle Kohn-Sham equation can be given by:

$$\underbrace{\left[-\frac{1}{2} \nabla^2 + V_{\text{eff}}(\mathbf{r}) \right]}_{\hat{H}_{KS}} \varphi_i = \epsilon_i \varphi_i \quad (3.35)$$

Similar to HF equations, the Kohn-Sham equations need to be solved iteratively to obtain orbitals and ground state energies, since $V_{\text{eff}}(\mathbf{r}_1)$ depends on the density due to the Coulomb part.

The V_{XC} is defined as the functional derivative of E_{XC} , with respect to the electron density, which is given by

$$V_{XC}(\mathbf{r}_1) = \frac{\partial E_{XC}}{\partial \rho(\mathbf{r})}. \quad (3.36)$$

Since both of the exchange-correlation terms are unknown, and thus have to be approximated, this is the main challenge to the DFT. Various approximations can be used in DFT to calculate the electron correlation by using the exchange-correlation functional, which is briefly explained in the next section.

3.5.4. Local Density Approximation

In DFT, local density approximations (LDA) depend solely upon the value of a point's electronic density at each point in space to approximate the exchange-correlation energy functional. LDA is the simplest method, which is based on the model of the uniform electron gas or homogeneous electron gas. The E_{XC} using LDA is given by

$$E_{XC}^{LDA}[\rho] = \int \rho(\mathbf{r})\epsilon_{xc}(\rho(\mathbf{r}))d\mathbf{r} \quad (3.37)$$

$$\epsilon_{XC}[\rho(\mathbf{r})] = \epsilon_X[\rho(\mathbf{r})] + \epsilon_C[\rho(\mathbf{r})] \quad (3.38)$$

ϵ_{xc} is split up as a sum of the exchange contribution and the correlation contribution. Using the free electron gas model, the exchange part can be calculated, while the correlation part can be determined using quantum Monte Carlo simulations. The exchange ϵ_X is given by:

$$\epsilon_x(\rho) = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \rho(\mathbf{r})^{1/3} \quad (3.39)$$

Geometries can be reproduced well by LDA, however, it fails when it comes to bond energies because of its tendency to overbind. LDA cannot adequately describe molecules due to their homogeneous electron density.

3.5.5. Generalised Gradient Approximation

LDA is best used for solids rather than molecules because it requires slowly varying densities in order to directly adopt the uniform electron gas. It is for this reason, that corrections have been developed that take into account fluctuations in density. The *generalised gradient approximation* (GGA) extends the LDA approach by introducing the concept of gradient of the density $\nabla\rho(\mathbf{r})$ to account for non-homogeneity of the electron density [13].

$$E_{XC}^{GGA}[\rho] = \int \rho(\mathbf{r})\epsilon_{xc}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))d\mathbf{r} \quad (3.40)$$

A number of gradient-corrected functionals were developed, most of which added a correction term to the LDA functional. In order to calculate the gradient expansion, several coefficients must be determined. Consequently, some functionals are created by fitting parameters to reference data. Some popular GGA parameters are B88 (by Becke) [14], LYP (by Lee, Yang and Parr) [15], and Perdew-Burke-Ernzerhof (PBE) functional [16].

Moreover, GGA functional works better than the LDA functional, but at the same time, it underestimates the energy barrier. Both GGA and LDA show a limitation due to self-interaction error. Thus, several other approximations were developed, such as hybrid functionals and long-range corrected functionals.

3.6. Density-Functional Tight-Binding

The DFT can treat systems up to 100 atoms, however, if the number of atoms increases, the computation cost also increases. In spite of the fact, that DFT is already an extremely efficient computational method compared to most wavefunction-based electronic structure methods, there are still many problems, that can be addressed using quantum mechanics at a lower cost. A typical example would be the simulation of large nanostructures or protein enzyme complexes with thousands of atoms each. As the relevant time scales can be in the order of nanoseconds, it is necessary to perform millions of quantum calculations, which are well beyond the capabilities of direct quantum computation. Various semiempirical and approximate methods are available for such applications, that are computationally less expensive but at the cost of some level of accuracy. Furthermore, it is important to note, that many of those methods are heavily dependent on fitting reference data parameters to them.

In many respects, Density-Functional Tight-Binding (DFTB) can be considered robust, because it is derived and parameterised from DFT. As the underlying assumptions of the model are usually satisfied, it tends to perform particularly well, when it comes to organic molecules, as it provides a kind of reliable foundation, that many empirical methods lack. In the first version of DFTB's use of a reference density equal to a fixed density, the procedure was non-self-consistent, resulting in the simplification of matrix elements and a reduction in computation time. Although this assumption works for non-polar systems, it fails for polarised systems, such as molecules with different electronegativities. The reference density ρ^0 is obtained from a superposition of pseudo-atomic electron densities previously determined with DFT calculations. Due to the tight-binding assumption, it is possible only to consider the valence electrons in calculations since the repulsive energy already includes the contributions of the core electrons. It is for the same reason, that the reference density ρ_0 is extended with the density fluctuation $\delta\rho$ around it, while the reference density ρ_0 is expressed as the sum of neutral atomic densities.

The starting point of DFTB is the DFT total energy, which is expanded in a Taylor series around a chosen density $\rho(\mathbf{r})$.

$$\rho(\mathbf{r}) \equiv \rho = \rho_0 + \delta\rho = \sum_a \rho_a^0 + \delta\rho \quad (3.41)$$

A Taylor series is then used to expand the exchange-correlation functional around the reference density:

$$\begin{aligned} E_{\text{XC}}[\rho_0 + \delta\rho] &= E_{\text{XC}}[\rho_0] + \int \left. \frac{\partial E_{\text{XC}}[\rho]}{\partial \rho} \right|_{\rho_0} \delta\rho \, d\mathbf{r} \\ &+ \frac{1}{2} \iint \left. \frac{\partial^2 E_{\text{XC}}[\rho]}{\partial \rho \partial \rho'} \right|_{\rho_0, \rho_0'} \delta\rho \delta\rho' \, d\mathbf{r} d\mathbf{r}' \\ &+ \frac{1}{6} \iiint \left. \frac{\partial^3 E_{\text{XC}}[\rho]}{\partial \rho \partial \rho' \partial \rho''} \right|_{\rho_0, \rho_0', \rho_0''} \delta\rho \delta\rho' \delta\rho'' \, d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' + \dots \end{aligned} \quad (3.42)$$

Rearranging equation 3.42, the total energy of the system can be rewritten as:

$$\begin{aligned}
 E^{\text{DFTB}} &= \sum_i^{\text{occ}} \langle \varphi_i | \underbrace{-\frac{\nabla^2}{2} + V_{\text{ext}} + \int \frac{\rho_{0'}}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{XC}}[\rho_0]}_{\hat{H}^0} | \varphi_i \rangle \\
 &\quad - \frac{1}{2} \iint \frac{\rho_0 \rho_{0'}}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' - \int V_{\text{XC}}[\rho_0] \rho_0 d\mathbf{r} + E_{\text{XC}}[\rho_0] + E_{\text{NN}} \\
 &\quad + \frac{1}{2} \iint \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} + \frac{\partial^2 E_{\text{XC}}[\rho]}{\partial \rho \partial \rho'} \bigg|_{\rho_0 \rho_{0'}} \right) \delta \rho \delta \rho' d\mathbf{r} d\mathbf{r}' \\
 &\quad + \frac{1}{6} \iiint \frac{\partial^3 E_{\text{XC}}[\rho]}{\partial \rho \partial \rho' \partial \rho''} \bigg|_{\rho_0 \rho_{0'} \rho_{0''}} \delta \rho \delta \rho' \delta \rho'' d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' + \dots \\
 &= E^0[\rho_0] + E^1[\rho_0, \partial \rho] + E^2[\rho_0, (\partial \rho)^2] + E^3[\rho_0, (\partial \rho)^3] + \dots
 \end{aligned} \tag{3.43}$$

Different methods of DFTB correspond to different truncations of the Taylor series. It is the original DFTB method, that results from the truncation after the first-order term [17, 18]. Second-order terms are included in DFTB2 [19], whereas third-order terms are included in DFTB3 [20, 21].

3.6.1. DFTB1

If only the zero-order and first-order terms are considered from equation 3.43, the standard (non-self-consistent) DFTB1 method is obtained. In this DFTB scheme, the superposed density of the atoms is taken as starting point of ρ_0 [22]. It is important to note that only the valence electrons are explicitly considered in the DFTB formalism, and the KS orbitals are expanded in a minimal manner:

$$\rho_0(\mathbf{r}) = \sum_A \rho_A(\mathbf{r}) \tag{3.44}$$

Based on a minimal basis set of Slater-type for the KS orbitals, the first term of equation 3.43 represents the Hamiltonian matrix elements as follows:

$$E^{(1)} = \sum_i^{\text{occ}} \langle \varphi_i | \hat{H}_0 | \varphi_i \rangle = \sum_i^{\text{occ}} \sum_{\mu} \sum_{\nu} c_i^{\mu} c_i^{\nu} H_{\mu\nu}^0 \tag{3.45}$$

The diagonal Hamiltonian matrix $H_{\mu\mu}^0$ element leads to the approximation of KS eigenvalues, which are calculated by an atomic DFT calculation using the PBE exchange-correlation functional. For off-diagonal elements, $H_{\mu\nu}^0$, a two-centre approximation is applied and the matrix elements are listed in tables by calculating different internucleic distances. Collecting the terms, that only depend on ρ_0 and thus the repulsive energy contribution, is summarised as a sum of two-body potentials. The repulsive energy contribution is given by:

$$E^{\text{rep}} = \frac{1}{2} \sum_A \sum_B V_{AB}^{\text{rep}} [\rho_A^0, \rho_B^0, \mathbf{r}_{AB}] \quad (A \neq B), \quad (3.46)$$

where A and B are two atoms. It performs well with systems, in which there is no charge transfer or a complete charge transfer. It is necessary to include higher-order terms in systems, that are sensitive to charge fluctuations.

3.6.2. DFTB2

The inclusion of the second-order term in DFTB1 will lead to the self-consistent charge (SCC) DFTB method, which is also known as DFTB2 [23]. The total energy contribution is given by:

$$E^{\text{DFTB2}} = \underbrace{\sum_{iAB} \sum_{\mu \in A} \sum_{\nu \in B} n_i c_{\mu i} c_{\nu i} H_{\mu\nu}^0}_{E^{\text{H0}}} + \underbrace{\frac{1}{2} \sum_{AB} \Delta q_A \Delta q_B \gamma_{AB}}_{E^{\gamma}} + \underbrace{\frac{1}{2} \sum_{AB} V_{AB}^{\text{rep}}}_{E^{\text{rep}}} \quad (3.47)$$

This equation consists of three terms: the first term E^{H0} contains the DFTB matrix elements, the second is an approximation of the second order term of the DFT Taylor series expansion, and the third represents the DFTB repulsive potential. Due to the fact that DFTB2 does not differentiate between the chemical hardness of anions, cations, and neutral atoms, highly charged systems can have some limitations. A further characteristic of atoms is, that their shape is fixed, as it is defined by the initial reference density.

3.6.3. DFTB3

Further extension of the DFTB2 method by the inclusion of third-order terms will lead to the DFTB3 formalism [21]. By introducing E^{Γ} , these restrictions were removed as the Γ_{AB} function contains the derivative of the Hubbard parameter, which takes into account the chemical hardness. As a result, charged systems can also exhibit the expected behaviour using DFTB3 formalism.

$$E^{\text{DFTB3}} = \underbrace{\sum_{iAB} \sum_{\mu \in A} \sum_{\nu \in B} n_i c_{\mu i} c_{\nu i} H_{\mu\nu}^0}_{E^{\text{H0}}} + \underbrace{\frac{1}{2} \sum_{ab} \Delta q_A \Delta q_B \gamma_{AB}}_{E^{\gamma}} + \underbrace{\frac{1}{3} \sum_{AB} \Delta q_A^2 \Delta q_B \Gamma_{AB}}_{E^{\Gamma}} + \underbrace{\frac{1}{2} \sum_{AB} V_{AB}^{\text{rep}}}_{E^{\text{rep}}}. \quad (3.48)$$

DFTB3 parameters should also be refined to improve the accuracy of this semiempirical method. DFTB3 parametrisation 3OB was developed by optimising electronic and repulsive parameters [24].

3. *Quantum Chemistry*

The DFTB method is about 2 to 3 orders of magnitude faster than a DFT calculation, as it is based on a large number of readily available parameters, which do not have to be determined during the simulation. However, it maintains a reasonable level of accuracy.

4. Classical Mechanics and Molecular Dynamics Simulations

The fundamental concept of molecular mechanics (MM) is based on the uses of classical mechanics to model molecular systems. Molecular potential energy can be expressed through bonding and nonbonding interactions. In order to formulate force fields, one must give explicit mathematical forms to these terms, and in order to parameterise the field, one must give actual numbers to the constants in the force fields. MM has a lot of potential applications, but one of the most important ones is providing a powerful computational tool for the investigation of biomolecules in order to use in pharmaceutical industries and drug design. MM is used in molecular dynamics to simulate the motion of molecules by computing forces acting on them.

4.1. Perspective

It can be said that molecular mechanics (MM) [25, 26, 27, 28] is the application of mathematical techniques to the representation of molecules as a collection of balls (representing the atoms) held together by springs (corresponding with the bonds). This model proposes that the energy of a molecule is affected by the relationship between geometry and its shape, because neither springs easily stretch or bent away from some "natural" length or angle, nor balls can be pushed together too tightly. The mathematical model is therefore conceptually very close to the intuitive feeling one has when dealing with molecular models and trying to describe molecular energetics conceptually. In space-filling models, atoms cannot be pushed too close together due to the model's ability to resist distortions from the "natural" geometry.

According to the principle of MM, the energy of a molecule is estimated by the resistance of its bonds to stretching, bending, and crowding. By applying the energy equation to the minimum-energy geometry, the corresponding bond lengths, angles, and dihedrals are found. A force field is composed of the mathematical expression for the potential energy, and the parameters contained in that expression. It is referred to as force field methods when applied to molecular mechanics. According to this definition, force is the amount of force exerted on a particle with respect to displacement along spatial coordinates. It is a result of the negative of the first derivative of the potential energy. In order to determine the force on each atom, we can differentiate $E(x, y, z)$, x, y, z being coordinates of atoms.

As the method does not take electrons into account, it cannot provide information about electronic properties such as charge distributions or nucleophilic and electrophilic

behaviour. Due to the fact that MM implicitly uses the Born-Oppenheimer approximation, only the nuclei experience a static attraction from electrons and/or springs. MM emphasises bonding as a fundamental concept, but it is not essential – though often useful – when calculating electronic structures. Atoms and bonds are considered the defining characteristics of a molecule by MM, where the bonds are viewed as springs connecting atoms.

Molecular mechanics evolved from an attempt to provide quantitative information about chemical reactions at a time when performing quantum mechanical calculations on systems much larger than the hydrogen molecule seemed very unlikely. Westheimer and Meyer [29] and Hill [30, 31] formulated the principles of MM as a potentially general method for examining how the energy and geometry of molecular systems are affected by the configuration or geometry in 1946. It is important to note that all of these papers deal with the conformations of organic molecules. According to Westheimer and Mayer, steric repulsion can be described by exponential terms. It was clearly demonstrated through approximate calculations that MM can be helpful when solving organic chemistry problems, but further extensive computations and parameters are required to achieve greater accuracy. The choice of parameters for MM calculations was challenging. Nevertheless, it was clear that they were not only useful tools for rationalising experimental observations, reproducing, and predicting the structure and energy characteristics of organic molecules of medium size with experimental accuracy but also for the development of synthesis pathways. Force field emerged as a transformation of various mathematical expressions and parameters into ready-to-use computer programs, i.e. complete and verified sets of formulas and constants. The paper by Allinger and Sprague [32] resulted in MM1, as one of the first force fields and programs to be developed, followed by MM2 [27], MM3, and MM4, which are force fields for organic molecules. Until last year, new terms of energy and atom types were included, as well as parameter refinement.

4.2. Principles of Molecular Mechanics

The energy function, in general, consists of two terms describing two kinds of interactions: bonded interactions between atoms and non-bonded interactions. Both are based on the distance between the interacting atoms.

$$E_{total} = E_{bonded} + E_{non-bonded} \quad (4.1)$$

Three forms of bonded interaction are involved in chemical bonds: stretching the bonds between atoms, bending the angles (where three atoms are involved), and dihedral angles (when four atoms are involved). Interactions between non-bonded molecules are pairwise and are dependent on distance. The non-bonded interaction consists of two types of Lennard-Jones interactions, short-range repulsion, and long-range attraction of dispersion. Coulomb interactions are used to describe particles with charge or partial charge. Thus the potential energy of a molecule can be written as

$$E = \underbrace{\sum_{\text{bonds}} E_{\text{stretch}} + \sum_{\text{angles}} E_{\text{bend}} + \sum_{\text{dihedrals}} E_{\text{torsion}}}_{E_{\text{bonded}}} + \underbrace{\sum_{\text{pairs}} E_{\text{LJ}} + \sum_{\text{pairs}} E_{\text{coulomb}}}_{E_{\text{non-bonded}}} \quad (4.2)$$

where E_{stretch} , E_{bend} , E_{torsion} , $E_{\text{non-bonded}}$ are energy contributions from bond stretching, angle bending, torsional motion (rotation) around single bonds, and interactions between atoms or groups that are not directly bonded (non-bonded) respectively. The contributions are demonstrated in fig. 4.1. These terms constitute a particular force field based on their mathematical form and parameters.

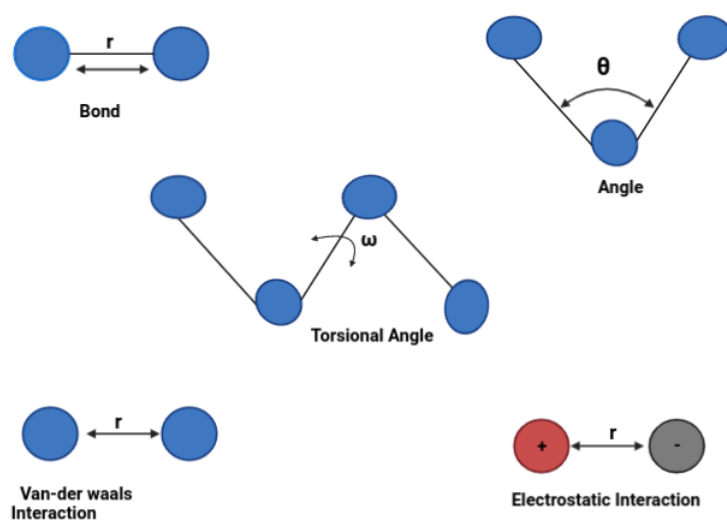


Figure 4.1.: Illustration of various contributions of force field parameters.

4.2.1. Bonding Interaction

4.2.1.1. The Bond Stretching Term

The Morse potential can describe the dissociation of a bond, which is given by

$$E(r) = D_e(1 - \exp[-\alpha(\mathbf{r} - \mathbf{r}_0)])^2 \quad (4.3)$$

Here, D_e is the dissociation energy needed to break the bond, \mathbf{r}_0 is the equilibrium nucleus-nucleus distance (lowest potential energy), \mathbf{r} the displacement from \mathbf{r}_0 and k the spring constant of the bond, α is the width of the potential determined by

$$\alpha = \sqrt{\frac{k}{2D_e}} \quad (4.4)$$

But, this potential describes a chemical bond well. Due to the harmonic approximation, bonds cannot be broken or formed with molecular mechanics. A Taylor series expansion up to the second order is performed and a harmonic approximation for the Morse potential is obtained. By truncating after the quadratic term, the result is what is known as the harmonic approximation in force fields.

$$E_{stretch} = \sum_{bonds} \frac{1}{2} \cdot k_{stretch} \cdot (\mathbf{r} - \mathbf{r}_0)^2 \quad (4.5)$$

Here, $k_{stretch}$ is the force constant. The larger the value of $k_{stretch}$ is, the stiffer the bond and the more it resists being stretched.

4.2.1.2. The Angle Bending Term

Also, a harmonic approximation for the angle deformation is performed. The analogous equation is given by:

$$E_{bend} = \sum_{angles} \frac{1}{2} \cdot k_{bend} \cdot (\theta - \theta_0)^2 \quad (4.6)$$

Here, k_{bend} is the force constant, and θ is the angle when it is distorted from the equilibrium angle, that is θ_0 .

4.2.1.3. The Torsional Term

The dihedral angle or torsional angle of the system is the angle between the A–B bond and the C–D bond as viewed along the B–C bond, for a system of four atoms bonded sequentially as A–B–C–D. It is evident that the energy function exhibits a rotational barrier and that there must be a periodicity in angle ω , which must remain the same after a 360° rotation. The torsional potential energy curve is more complicated, but a combination of sine or cosine functions will sufficiently reproduce the curve:

$$E_{torsion} = \sum_{dihedrals} V_n \cdot \cos(n \cdot \omega) \quad (4.7)$$

where V_n is the barrier for rotation, n is the rotation periodicity, and ω is the torsional angle.

4.2.2. Non-bonding Interaction

4.2.2.1. Lennard-Jones Potential

Essentially, the Lennard-Jones potential is a way to express the repulsion and attraction of particles that are not directly bonded with each other but in close proximity. For large distances, the potential is zero, for short and intermediate distances there is a slight interaction (because of electron correlation). The Lennard-Jones potential combines the none charge Pauli repulsion (r^{-12}) and attraction (r^{-6}) derived from the London dispersion.

$$E_{LJ} = \sum_{pairs} \left\{ 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^6 \right) \right\} \quad (4.8)$$

where ϵ_{ij} is the depth of the potential well and σ_{ij} is the atomic distance and r is the distance.

4.2.2.2. Coulomb Potential, Charged Interactions

The Coulomb potential is applied to charged atoms and partially charged atoms. As electrons are redistributed within the molecule, partial charges are introduced in parts of the molecule:

$$E_{coulomb} = \sum_{pairs} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (4.9)$$

The Coulomb potential describes the electrostatic interactions of point charges q_i and q_j divided by the distance r_{ij} , which is the distance between these two charged particles. ϵ_0 is the vacuum dielectric constant. It is important to note that non-bonded interactions are calculated only between atom pairs that are separated by at least three covalent bonds. Interactions between neighbouring atoms and atoms that are separated by two and three bonds are already covered by the bonded energy terms.

We can now consider putting actual numbers, all the force constants for a bond, angle, torsion angle, equilibrium positions for distances, and angles, to give expressions that we can actually use. The process of finding these numbers is called parameterising the force field. In short, these values are often empirically derived from experiments or calculated with quantum chemical methods.

4.2.3. Complete Force Field Equation

The total potential energy is obtained by combining all the bonding and non-bonding energy functions described previously.

$$E_{total} = \underbrace{\sum_{bonds} \frac{1}{2} \cdot k_{stretch} \cdot (r - r_0)^2 + \sum_{angles} \frac{1}{2} \cdot k_{bend} \cdot (\theta - \theta_0)^2 + \sum_{dihedrals} V_n \cdot \cos(n \cdot \omega)}_{E_{bonded}} + \underbrace{\sum_{pairs} \left\{ 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^6 \right) \right\} + \sum_{pairs} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}}_{E_{non-bonded}} \quad (4.10)$$

4.3. Propagation of a System

To a good extent, nuclei are heavy enough to behave as classical particles to a good extent, and therefore the dynamics of the nucleus can be modelled by solving Newton's second equation, $F = ma$, which offers a good approximation. It can be written as:

$$-\frac{\partial V}{\partial \mathbf{r}} = m \frac{\partial^2 \mathbf{r}}{\partial t^2} \quad (4.11)$$

Here, V is the potential energy at position \mathbf{r} . The vector \mathbf{r} consists of all the coordinates for all the particles (in Cartesian coordinates, it is a vector of length $3N_{atom}$). Negative potential energy gradients are also called forces (F) on particles. By integrating the equations of motion, a trajectory is obtained that provides information about particle positions, velocities, and accelerations over time.

The problem cannot be solved analytically, so it must be solved numerically. Taylor expansions are used to find the positions of particles at small time steps Δt later for a set of particles with positions r_i .

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + v(t) \cdot \Delta t + \frac{1}{2} a(t) \cdot \Delta t^2 \quad (4.12)$$

$$v(t + \Delta t) = v(t) + \frac{1}{2} (a(t) + a(t + \Delta t)) \cdot \Delta t \quad (4.13)$$

$$a(t + \Delta t) = a(t) + b(t) \cdot \Delta t \quad (4.14)$$

where \mathbf{r} is the position, v is the velocity which is the first derivative of coordinates with respect to time and a is acceleration which is the second derivative of coordinates with respect to time. The period of the fastest motion is used as the lower boundary of the integration step in order to keep the integration step as large as possible. For biomolecular simulations, the X-H bond vibration has a period of 10 fs. In view of this, the time step has been set to 1 fs, which is stable with numerical integration. It was necessary to design several constraints for bond lengths so that the time step could be increased from 1 to 2 fs. For molecular dynamics simulation and in this study, the most frequently used constraints are LINCS (Linear Constraint Solver for Small Molecules) [33], SHAKE (bond geometry restraint for large molecules) [34] and SETTLE (to constrain rigid water) [35].

4.3.1. Verlet Integration Method

The Verlet algorithm uses position and acceleration at time t and time $t - \Delta t$ to calculate the new position $t + \Delta t$. An expansion of the Taylor series is used here to derive the result. Velocities were not used in the calculation, but they can be calculated from $t - \Delta t$ and $t + \Delta t$. But, the velocities at time $t + \Delta t$ must be calculated before the velocities at time t .

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + v(t) \cdot \Delta t + \frac{1}{2} a(t) \cdot \Delta t^2 \quad (4.15)$$

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - v(t) \cdot \Delta t + \frac{1}{2} a(t) \cdot \Delta t^2 \quad (4.16)$$

The sum of these two equations is given by

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) + \mathbf{r}(t - \Delta t) + a(t) \cdot \Delta t^2 \quad (4.17)$$

This is the Verlet algorithm for solving Newton's equation numerically [36]. A trajectory is generated by propagating atomic positions in time and evaluating acceleration from the forces at each time step. With decreasing step size Δt , the trajectory becomes closer and closer to the "true" trajectory, until finite numerical problems arise. A small time step, however, implies that a greater number of steps is required to propagate the system in a given amount of time, which means that the computational effort increases in an inverse proportion to the size of the time step. As a result of the truncation problem and the lack of explicit velocities, the Verlet algorithm has a numerical disadvantage in the context of generating ensembles with constant temperatures.

4.3.2. Leap Frog Integration Method

Leap frog can deal with the numerical aspect and the lack of explicit velocities in the Verlet algorithm [37]. The velocities, in this case, were first calculated at time $t + 1/2\Delta t$. Based on these data, positions \mathbf{r} were calculated at time $t + \Delta t$. Afterwards, velocities leap over positions, then positions leap over velocities, and so on will be calculated. Within this approach, velocities are explicitly calculated, but not at the same time as the position. However, they can be approximated for a particular time t :

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + v(t + \frac{1}{2}\Delta t) \cdot \Delta t \quad (4.18)$$

$$v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + a(t) \cdot \Delta t \quad (4.19)$$

The theoretical accuracy of this algorithm is also of the third order, same as the Verlet's, but the numerical accuracy is better. A further benefit is the direct display of the velocities, which facilitates coupling to an external heat bath. A disadvantage of this method is that positions and velocities are not known simultaneously. They are always out of phase by half a time step.

4.3.3. Velocity Verlet Integration Method

The velocity Verlet algorithm [38] can eliminate the latter abnormality by propagating atoms using the following equations:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + v(t) \cdot \Delta t + \frac{1}{2} \cdot a(t) \cdot \Delta t^2 \quad (4.20)$$

$$v(t + \Delta t) = v(t) + \frac{1}{2} \cdot [a(t) + a(t + \Delta t)] \cdot \Delta t \quad (4.21)$$

4.4. Periodic Boundary Conditions

For systems with around 10^6 atoms, the modelled boundary surface has a large effect on the system properties, which requires huge computing time. It is necessary to keep the molecular dynamics simulation system as small as possible in order to save computing resources. Meanwhile, a bulk solvent should be used instead of a surface to simulate the molecule. In order to replicate the box around the molecule and its solvent across all spatial directions, periodic boundary conditions, which are schematically represented in fig. 4.2, must be introduced.

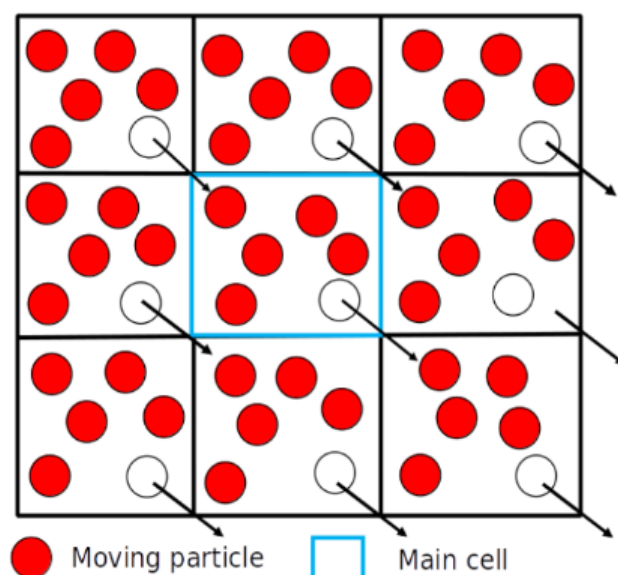


Figure 4.2.: Schematic representation of the periodic boundary conditions to minimise the error due to abrupt environmental changes at the boundary. Whenever a particle moves out of the main cell, a particle moves in from the neighbouring box.

In biomolecular simulations, water plays a very important role. A very high number of water molecules are present relative to the number of atoms in the molecule of interest. This results in a high computational burden since many calculations involve only solvent molecules. For this reason, it is necessary to use a simple yet accurate model to describe water. It is crucial to choose the right box size based on the size of the studied system and the proposed behaviour. It is more computationally efficient to use rigid water models like TIP3P because bond lengths and angles cannot vary.

4.5. Temperature and Pressure Control

4.5.1. Thermostat

It is well known that the natural time evolution of a classical system that has N particles within a distinct volume has a constant of motion, which is the energy E . As time averages are equivalent to ensemble averages (ergodic theorem), it is possible to describe the system as a microcanonical ensemble NVE. In a microcanonical ensemble or NVE ensemble, the total number of atoms N is constant as well as the temperature T and the total energy E of the system. Compared to experimental data, simulation results are more easily compared to other ensembles such as NPT or NVT. It is common for experiments to be conducted at constant pressures or temperatures. Therefore, the isothermal-isobaric ensemble NPT or canonical ensemble NVT is the goal. The thermostat algorithm can be used to generate a canonical NVT ensemble that conserves temperature.

The Berendsen Thermostat introduces an external heat bath with a particular reference temperature T_0 , which is coupled to the molecular system. In the Berendsen approach, the atomic velocities are rescaled, and the system is weakly coupled to the heat bath with the desired temperature T_0 . In this way, the system adapts slowly to its surroundings, thus forming an NVT ensemble (constant number of particles, constant volume, and constant temperature). The scaling factor λ is introduced:

$$\lambda = \sqrt{1 + \frac{\delta t}{\tau_T} \cdot \left(\frac{T_0}{T} - 1 \right)} \quad (4.22)$$

where δt is the integration time step and the coupling strength is described by τ_T . The large temperature differences are removed, but there is a reduction in kinetic energy fluctuations. Most simulations require a canonical ensemble, which is not generated by this thermostat. Therefore, in order to obtain a correct Maxwell-Boltzmann distribution of velocities using the Velocity-rescaling temperature method or the Nosé-Hoover thermostat, it is usually necessary to add a stochastic term to the Nosé-Hoover thermostat in order to solve the kinetic energy distribution problem.

4.5.2. Barostat

Similarly, an external pressure bath can be coupled to the system for temperature. As the simulation progresses, each time step adjusts the size of the simulation box. Similar to a thermostat, an additional term is added to the equation of motion to control the pressure:

$$\frac{dP}{dt} = \frac{P_d - P}{t_p} \quad (4.23)$$

Where P_d represents the desired pressure and t_p represents the exchange time. The volume (coordinates and box vector) of the box is rescaled every time step by the factor:

$$\chi = 1 - \beta_t \cdot \frac{\delta t}{t_p} (P_d - P) \quad (4.24)$$

A simulation timestep is represented by δ_t , and β_t represents the isothermal compressibility. In this way, we can generate an isothermal-isobaric NPT ensemble. There may be a problem, particularly for small systems, where the fluctuations in pressure or volume are significant (for example in order to calculate thermodynamic properties), that the exact ensemble for the weak-coupling scheme is not well defined and it does not accurately simulate the true NPT ensemble, especially when the fluctuations in pressure or volume are important. In such situations, the Parrinello-Rahman pressure coupling can be used to get the true NPT canonical ensembles.

4.6. Intermolecular Interactions

For a system with N atoms, evaluating non-bonded interactions of N^2 atoms pair is the most computationally expensive part. As with biological systems with a system size of 10^6 atoms, comparing all of the pair atoms should be avoided in biological systems. Cut-off distances can be used to restrict long-range interactions. Lennard-Jones potentials with fast decaying r^{-6} functions are well approximated by this method, but Coulomb potentials with slow decaying r^{-1} functions are approximated with a certain amount of error.

The Ewald summation method is one of the most popular approaches to overcome the problems associated with cutting off the coulomb potential by PME only. A periodic boundary condition is applied to Coulomb interactions. All charged particles in the main cell are added together to form the energy for the main cell.

$$U = \frac{f}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z^*} \sum_i^N \sum_j^N \frac{q_i q_j}{\mathbf{r}_{ij,n}}. \quad (4.25)$$

where N is the number of charged particles, (n_x, n_y, n_z) are box index vectors, and the star indicates that terms with $i = j$ should be omitted when $(n_x, n_y, n_z) = (0, 0, 0)$. In other words, $\mathbf{r}_{ij,n}$ describes the actual distance between the charges, not the minimum image. This sum is conditionally convergent but very slow. Originally, the Ewald summation method was introduced as a way of calculating the long-range interactions between periodic patterns located in crystals. In this method, the slow-converging sum is converted into two rapidly-converging terms and one constant term.

So the Ewald formula is summarised as

$$U = U_{\text{direct}} + U_{\text{reciprocal}} + U_{\text{correction}} \quad (4.26)$$

$$U_{\text{direct}} = \frac{f}{2} \sum_{i,j}^N \sum_{n_x} \sum_{n_y} \sum_{n_z^*} q_i q_j \frac{\text{erfc}(\alpha \mathbf{r}_{ij,n})}{\mathbf{r}_{ij,n}} \quad (4.27)$$

$$U_{\text{reciprocal}} = \frac{f}{2\pi V_{\text{box}}} \sum_{i,j}^N q_i q_j \sum_{m_x} \sum_{m_y} \sum_{m_z^*} \frac{\exp(-(\pi m/\alpha)^2 + 2\pi i m \cdot (\mathbf{r}_i - \mathbf{r}_j))}{m^2} \quad (4.28)$$

$$U_{\text{correction}} = -\frac{f\alpha}{\sqrt{\pi}} \sum_i^N q_i^2 \quad (4.29)$$

Where V_{box} is the volume of a unit cell and α is the Ewald splitting parameter that determines the relative contributions to the direct and reciprocal sums, m_x, m_y, m_z . Please refer to the source for a complete explanation of complex mathematics [39]. Unfortunately, the computational cost of the reciprocal part of the sum increases as $N^{\frac{3}{2}}$ and it is therefore not realistic for use in large systems. PME interpolates charges rather than directly summing wave vectors. The PME algorithm scales as $N \cdot \log(N)$ and is substantially faster than ordinary Ewald summation on medium to large systems. In order to perform the Ewald summation, a neutral simulation box will be necessary (assuming that N equals 0). In the Verlet cut-off scheme, a constant is shifted from the direct space potential of the PME to zero so that the cut-off potential of the PME is zero. Due to the small amount of shift and the close proximity of the net system charge to zero, the total shift is very small in comparison to the Lennard-Jones potential where all shifts contribute to the total shift. It does not matter how the shift is applied, as long as the potential equals the force's integral.

Net charge leads to artefacts where a background charge occurs. This is mainly important for simulations with different dielectric constants. For example, proteins with lipid membrane simulation will result in strong artefacts. Therefore, it is recommended to use counter-ions to neutralise the system.

4.7. Hybrid QM/MM Simulations

There are many complicated calculations that need to be performed in order to be able to describe chemical reactions such as the breaking of bonds, the formation of bonds, or an electron transfer reaction. Whenever there is a large system whose atoms are in an excessive number and cannot be explained by quantum mechanics directly due to its size, a hybrid approach is the method of choice. As a result, a molecular mechanical force field may be combined with quantum mechanics methods of electronic structure in one simulation in order to achieve the following result. In hybrid QM/MM schemes, the reaction occurs in a part of the system that is described with QM (for example the substrate and active site in an enzyme) and the rest of the system is described using MM (for example, the surrounding tissue). Molecular active sites are calculated based on semiempirical, *ab initio*, or density functional methods, while the remaining larger part of the molecule can be interacted with through non-covalent interactions as well.

In order to calculate the Hamiltonian of a system, the Hamiltonians of both MM and QM regions as well as their interactions are added up together (additive scheme):

$$\hat{H}_{total} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (4.30)$$

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \quad (4.31)$$

In my work, the active site for the proton transfer (chapter 7) was described with QM methods, the rest of the protein, the POPC membrane and the solvent with MM methods. The calculation of the QM region and the MM rest is performed in a standard manner, but to describe the coupling between the QM and MM regions, further description is required.

A link atom scheme will be used when the border between the QM and MM regions intersects covalent bonds. Generally, only sidechains are described with QM in proteins, and a link atom is placed at a fixed distance along the C_α and C_β bonds. C_β is saturated with a link atom that is often a hydrogen-like atom and the bond between C_β . It is described as part of the QM region, rather than the MM region because the linker hydrogen atom is described as part of the QM region. To prevent over-polarisation of the QM region, it is important that the charges associated with C_α and its hydrogen bond are distributed to the remaining MM atoms (N, NH, C and O) so the polarisation of the QM region is not too great. If two out of three atoms belong to the QM region, or if three out of four atoms belong to the QM region, the force field does not calculate bending and torsional energies. A force field is used to calculate van der Waals interactions between QM and MM atoms by using a Lennard-Jones potential under standard force field conditions. Any important residue must be included in a QM region, but its number cannot exceed a certain value.

In order to calculate the electrostatic interaction between MM and QM atoms, several schemes have been proposed such as mechanical, electrostatic and polarisation embedding.

4.7.1. Mechanical embedding

Mechanical embedding involves calculating the electrostatic energy at the force field level. A QM region is evaluated based on point charges associated with the atoms. Consequently, the QM atoms are not polarised by the MM system.

$$E_{QM/MM} = \sum_i^{QM} \sum_m^{MM} \left(\frac{q_i \cdot q_m}{r_{im}} + 4\epsilon_{im} \left(\left(\frac{\sigma_{im}}{r_{im}} \right)^{12} - \left(\frac{\sigma_{im}}{r_{im}} \right)^6 \right) \right) \quad (4.32)$$

For the QM-atoms, the Lennard-Jones parameters are assigned and the charges associated with them are taken into account as well. The electron density in the QM region is unaffected by the MM charges, which can lead to problems describing QM regions with polar surroundings.

4.7.2. Electrostatic embedding

The electronic embedding approach is an alternative to the mechanical embedding approach where the QM region interacts with the atomic charges of the MM in a precise way.

MM point charges are included in the Hamiltonian of the QM region, while interactions between the QM nuclei remain in the QM/MM term of the Hamiltonian.

$$\hat{H}'_{QM} = \hat{H}_{QM} - \sum_j^{QM \text{ elec}} \sum_m^{MM \text{ atoms}} \left(\frac{q_m}{\mathbf{r}_{jm}} \right) \quad (4.33)$$

$$E_{QM/MM} = \sum_i^{QM} \sum_m^{MM} \left(\frac{Z_i \cdot q_m}{\mathbf{r}_{im}} + 4\epsilon_{im} \left(\left(\frac{\sigma_{im}}{\mathbf{r}_{im}} \right)^{12} - \left(\frac{\sigma_{im}}{\mathbf{r}_{im}} \right)^6 \right) \right) \quad (4.34)$$

As part of this study, we have also used an electrostatic embedding scheme which includes the effects of polarisation. With DFTB as the QM method, all of the collective electrostatic potentials induced by all MM atoms enter the Hamiltonian matrix elements of the DFTB3 and affect the QM distribution of charges.

4.7.3. Polarisation embedding

Using polarisation embedding, both subsystems of the system are polarised at the same time. In order to do this, however, a polarisable force field has to be created. This is a computationally expensive process and, therefore, is not used very often.

5. Enhanced Sampling Techniques in Molecular Dynamics

5.1. Introduction

There has been a rapid growth in the use of computer simulations of biomolecular systems over the past few decades, having progressed from simulating very small proteins in a vacuum all the way to simulating large protein complexes in solvated environments. In the past decade, all-atom molecular dynamics (MD) simulations, utilising classical mechanics, successfully assisted in the study of a broad variety of biological systems, ranging from small molecules such as small peptides to very large proteins [40, 41] and to proteins embedded within membranes [4, 42]. Molecular dynamics simulations using a hybrid QM/MM approach allowed enzymatic activation [43] or polarisable molecules to be studied in biological membranes. While MD simulations have been successful, they are still limited in two critical respects, namely the lack of accuracy of the force fields as well as their high computational cost.

For the past few years, molecular dynamics simulations have been one of the standard methods to study the conformational changes of biomolecules using general sampling methods. There are many local minima in biological molecules, separated by high energy barriers, indicating rough energy landscapes. It is possible, in most traditional simulations, to get stuck in a non-functional state if the energy height impedes a barrier crossing. A number of free energy minima are capable of trapping molecules for a long period of time, which in a sense slows down the process of sampling and results in a poor characterisation of a protein's dynamic behaviour [44, 45]. For proteins to be active in biological systems, large conformational changes are often critical to their activity. Catalysis, for instance, involves large movements of high amplitudes that are extremely important for substrate binding. Gating substrates require motions and changing mechanisms in channels and transporters when transported through membranes. Due to their complexity and time-consuming nature, these processes far exceed the capabilities of simple MD simulations, and improved sampling algorithms are often required to obtain reliable results.

In accordance with the ergodic hypothesis, a trajectory will ultimately exploit all the microstates of a system that are accessible, such as all those states that are compatible with the conservation of energy of the system. In equilibrium, the average time that a trajectory spends in a given region of the state space is proportional to the number of states that are accessible in that region. Time averages equal ensemble averages if the ergodic hypothesis is true and equipartition appears to be a valid assumption when time

averages equal ensemble averages.

Ergodicity refers to the fact that in a system consisting of multiple particles (atoms, proteins), over time each particle will sample the same states as one would see if one were to sample multiple particles at the same time. Accordingly, if one follows a single particle long enough, one should end up with the same result as by examining the data of a group of particles at a single point; i.e. the time average equals the ensemble average.

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} \quad (5.1)$$

With a simulation, all the possible states are usually not determined at a single point in time, but a single particle is often tracked over an extended period of time. It is possible to derive meaningful information from a simulation if it is ergodic and performed long enough to visit all of the system's energetically relevant states. The Boltzmann probability function can be calculated by sampling the full $6N$ -dimensional phase space:

$$P(\mathbf{r}, \mathbf{p}) \propto \exp\left(\frac{-E(\mathbf{r}, \mathbf{p})}{k_B T}\right) \quad (5.2)$$

According to Boltzmann probability theory, the Boltzmann probability function exponentially varies with the system energy $E(\mathbf{r}, \mathbf{p})$, k_B is Boltzmann's constant and T is the absolute temperature of the system. Thermodynamic properties can also be calculated from the Boltzmann distribution function such as the Helmholtz free energy F , enthalpy, entropy, etc.

$$F = -\beta^{-1} \ln \int_{\Sigma} e^{-\beta U(x)} dx \quad (5.3)$$

where integration is done over a subset of Σ of the configurational space corresponding to the macrostate. F is thus a function of $\beta = \frac{1}{k_B T}$ which is the absolute temperature, $U(x)$ which represents the potential energy function and Σ which is configurational space.

There is however a problem with standard MD simulations, since the system is limited to configurations that are available at the given temperature and, as a result, it fluctuates around its equilibrium state. If one supposes two configurations that are separated by a barrier height of more than a $k_B T$ at room temperature, the time required for the transition can be as much as milliseconds, which is beyond the scope of standard MD simulations.

In the past few decades, several enhanced sampling techniques have been developed in order to deal with this condition. For example, replica exchange molecular dynamics (REMD), metadynamics, umbrella sampling (US), and simulated annealing have been developed for this purpose. The methods can be classified into two categories: (i) collective variables-based methods and (ii) collective variables-free methods. I present a short summary of some of the techniques which are frequently used and that I have used in my work.

5.2. Collective Variable

The concept of collective variables (CVs) describes the reaction pathway by predefined reaction coordinates. It is possible to define them as a function of atomic coordinates $\xi(\mathbf{x})$, such as distance between two atoms, angles between three atoms, or dihedral angles between four atoms, depending on the function. A function ξ is mapping the full n -dimensional configurations \mathbf{x} to a lower-dimensional representation \mathbf{z} .

$$\mathbf{z} = \xi(\mathbf{x}) \quad (5.4)$$

5.3. Potential of Mean Force

An analysis of the free energy profile along a preferred reaction coordinate is referred to as the potential of mean force (PMF) as seen in fig. 5.1. A free energy profile provides a description of the force that is applied to particles of interest as a result of averaging the force for all possible configurations of a given system (the ensemble average of the force). The free energy difference ΔF between two states ξ_A and ξ_B can be obtained from the given equation,

$$\Delta F = -k_B T \ln \left(\frac{P(\xi_B)}{P(\xi_A)} \right) \quad (5.5)$$

where $P(\xi_A)$ is the probability of finding the system in state ξ_A and $P(\xi_B)$ in the state ξ_B , respectively. A trajectory from an MD simulation can be analysed using histogram analysis to calculate the potential of mean force. PMF force along the reaction coordinate is calculated if the correct probabilities for finding the system in a given state are known.

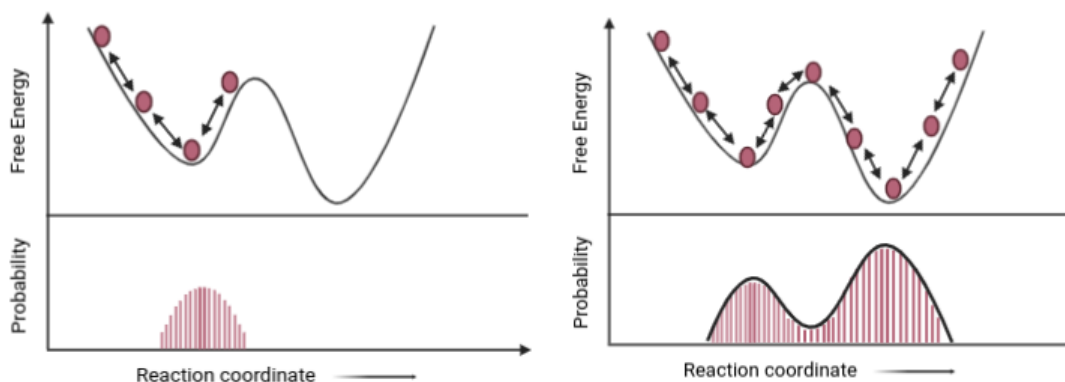


Figure 5.1.: The potential of mean force is shown as a graphical illustration with two minima separated by an energy barrier. It is possible for the system to overcome the energy barrier when sampled for a sufficient period of time. PMF can be calculated using the obtained probabilities of finding the system at a certain reaction coordinate value.

For instance, let us suppose there are two minima separated by a low energy barrier in a system under investigation. There is a fluctuation around the equilibrium in the system when it starts in the left minimum. The system can possibly cross the energy barrier if the MD simulation is performed long enough and/or if the energy barrier is not too high. By sampling the second minimum as well, the PMF can be determined based on the obtained probabilities. A simulation that is too short or the barrier too high will result in an inaccurate PMF.

5.4. Umbrella Sampling

The umbrella sampling (US) method as represented in fig. 5.2 was developed by Torrie and Valleau in 1977 [46]. In umbrella sampling, conformational transitions are modelled by dividing a CV into several independent windows. For each window, a simple harmonic potential $V(\xi)$ is added to the Hamiltonian, whose strength is determined by the force constant κ , which depends on the potential energy at the relevant reference points S_i along the CVs.

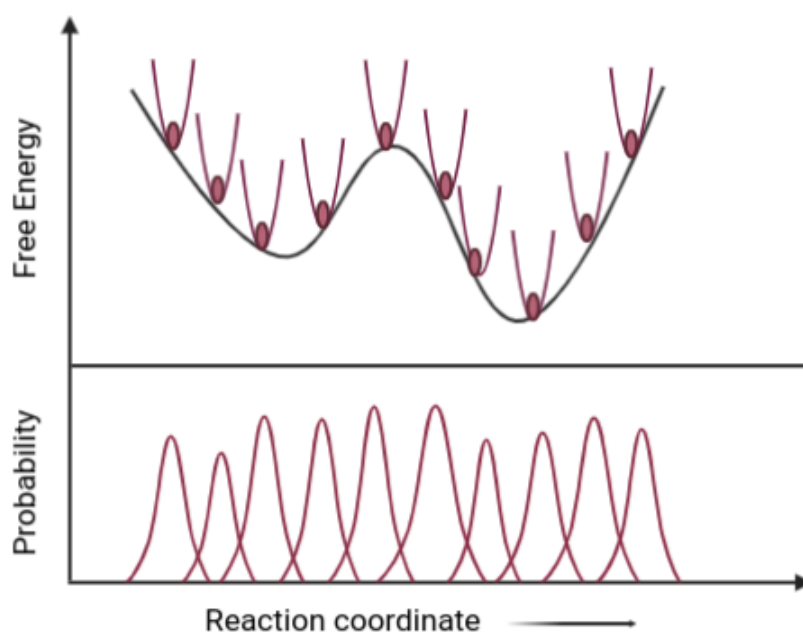


Figure 5.2.: Graphical illustration of umbrella sampling along the reaction coordinate ξ . The energy surface is divided into several windows, which is further restrained to a harmonic potential ξ_0 . Free energy can be obtained using biased probability if the windows are sufficiently overlapped.

The biasing potential depends only on the reaction coordinate and usually takes the form of a harmonic potential:

$$V(\xi) = \frac{1}{2}\kappa(\xi - \xi_0)^2 \quad (5.6)$$

where κ is the applied force constant and ξ_0 is the value of CV at which the potential is centred. MD simulations are performed for each window to obtain the biased probability $\mathcal{P}_i(\xi)$. For each window, the unbiased free energy $F_i(\xi)$ is obtained as follows:

$$F_i(\xi) = -k_B T \ln \mathcal{P}_i(\xi) - V_i(\xi) + C_i \quad (5.7)$$

where i is the window number, C_i is an unknown constant. Hence, $F_i(\xi)$ and $F_{(i+1)}(\xi)$ are shifted by $C_{i+1} - C_i$. Through an umbrella integration, unbiased free energy could be recovered following simulations using the weighted histogram analysis method (WHAM). This can, however, only lead to meaningful results if the biased probability distributions for each of the windows overlap sufficiently with those of the neighbouring windows.

5.5. Metadynamics

The Parrinello group proposed a new sampling algorithm referred to as metadynamics, which inserts memory into the sampling process [47]. The first methods to insert the memory into an enhanced sampling for biomolecules were local elevation [48] and conformational flooding [49]. Metadynamics discourages the resampling of previously visited states, which allows computational resources to be directed towards a broader exploration of the free energy landscape. In metadynamics, the Hamiltonian of a system is enhanced with a history-dependent bias potential based on CVs. It can be modelled as a sum of Gaussians deposited along the trajectory of the system within CVs space so that the system does not revisit configurations that have already been sampled.

5.5.1. Standard Metadynamics

In the standard metadynamics (METAD), a biased potential is added along the trajectory to the collective variable (CVs) [47] during the simulation as seen in fig. 5.3a. The added biasing potential forces the system to leave the local minimum and hence sampling of rarely visited states of chemical importance is possible. The evolution of the system is biased by the history-dependent potential, which is constructed as Gaussian functions. The collective variable as a function of atom coordinates is $s = s(q)$, and since the atom coordinates vary in time, it is $s = s(q(t))$. The biasing potential V is added in regular intervals τ .

$$V(s, t) = \sum_{k\tau} W(k\tau) \cdot \exp \left[-\frac{(s - s(q(k\tau)))^2}{2\sigma^2} \right] \quad (5.8)$$

where τ is the Gaussian deposition stride, σ is the width of the Gaussian for the i^{th} CV, and $W(k\tau)$ is the height of the Gaussian. Standard metadynamics has a constant height of W . In the long term, the biasing potential converges as the negative free energy as a function of the CV.

$$V(s, t \rightarrow \infty) = -F(\vec{s}) + C. \quad (5.9)$$

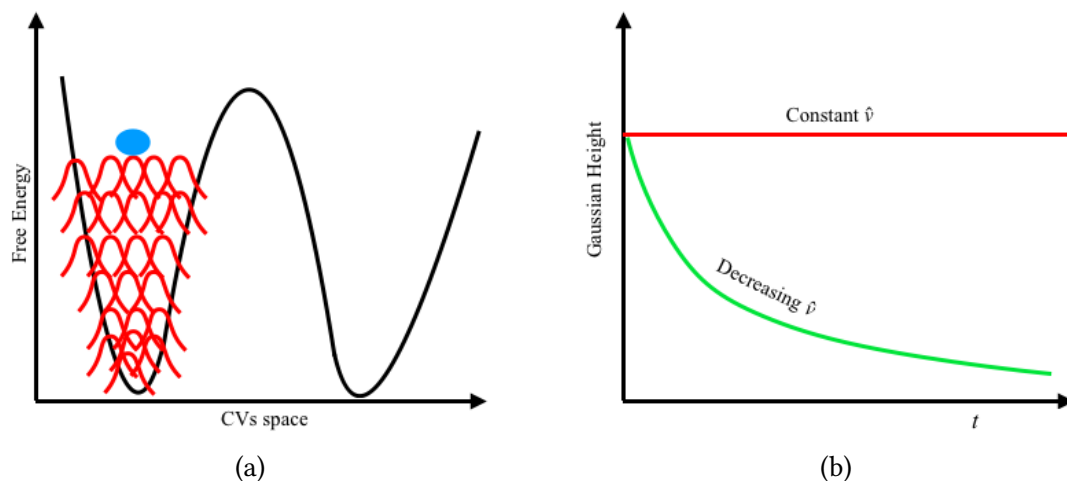


Figure 5.3.: Fig (a) shows the representation of the deposition of Gaussian potentials along the collective variable (CVs). Fig (b) shows the comparison of the shape of the Gaussian deposition rate between standard METAD (red line) and WT-METAD (green line).

5.5.2. Well-Tempered Metadynamics

In metadynamics, one of the challenges is being able to determine when to end a simulation. The FES does not converge in one run, making it difficult to determine when to stop the simulation. A rather interesting phenomenon is to observe that the result oscillates around the correct result, resulting in an average error proportional to the square root of the deposition rate. Metadynamics simulations can also cause the system to assume physical conformations that are irrelevant to its function if performed for a long period of time. Unlike standard metadynamics, Well-Tempered metadynamics (WT-METAD) [50] is inspired by the self-healing umbrella sampling method, which estimates the free energy surface (FES) that converges to the exact result in a long time limit while ensuring a relatively small error. It offers the possibility of controlling the regions of the FES which are relevant to chemical explorations. Well-Tempered metadynamics can resolve these problems due to the fact that it does not use constant-height Gaussian functions, but instead progressively decreases the bias heights over time.

The height of the Gaussian is decreased with the simulation time according to:

$$W(k\tau) = W_0 \cdot \exp \left[-\frac{V(s(q(k\tau)), k\tau)}{k_B \Delta T} \right] \quad (5.10)$$

where W_0 is an initial Gaussian height, ΔT is an input parameter with the dimension of a temperature, and k_B is the Boltzmann constant.

In this way, the FES converges to

$$V(S, t \rightarrow \infty) = \frac{\Delta T}{T + \Delta T} F(S) + \text{const.} \quad (5.11)$$

The bias factor γ must be chosen carefully in order to permit efficient crossing of the relevant free-energy barriers in the time scale of the simulation.

$$\gamma = \frac{T + \Delta T}{T} \quad (5.12)$$

where, T is room temperature, $T + \Delta T$ is the temperature of the CVs, and an additive constant C . The parameter ΔT determines how fast the bias decreases over time and is regulated by the bias factor γ as seen in fig. 5.3b. A bias factor $\gamma = 1$ ($\Delta T = 0$) corresponds to an unbiased MD simulation whereas a bias factor $\gamma \rightarrow \infty$ is equal to standard metadynamics.

5.5.3. Multiple-Walker Metadynamics

In order to reduce wall-clock time, one method that is often used is parallelising several metadynamics simulations (so-called walkers) in order to reconstruct the same FES at the same time. The multiple-walker metadynamics [50] approach is characterised by each walker sampling the FES independently, whereas the interactions between them are mediated only by their biasing potentials, which are the sum of all the Gaussians of all walkers. As a result, the metadynamics algorithm remains accurate and simulation time can be decreased efficiently.

Part III.
Contributions

6. Water and Proton Transport Across Tetrameric Charge Zipper Protein

6.1. Introduction

in order to protect itself from environmental stresses, *E. coli* produces an excessive amount of TisB. TisB is a twenty-nine residue amphiphilic α -helix that is located in the inner membrane. It is a toxic component of the bacterial Toxin-Antitoxin (TA) system [51, 52] where the TisB-IstR assembly plays a critical role in the stress response. The insertion of TisB in the lipid bilayer helps break the Proton Motive Force (PMF) by equilibrating the pH gradient across the membrane [53, 54]. This results in the temporary suppression of metabolic activities and other cellular functions such as ATP inhibition and transcription, which can lead to cell death. However, persister cells are in a state of dormancy, making them resistant to external environmental stresses, for example, antibiotics [53]. This promotes the formation of drug-tolerant persister cells due to the inactivity of the targets in the cell.

With polymer-exclusion experiments, it was demonstrated that TisB forms small pores in the lipid bilayer with a small aperture, which are anion-selective channels. They modulate the net charge of the protein and it has been proposed that the observed leakage of protons can be facilitated by the transfer of hydroxyl anions [55].

In the previous work from our group, Schneider et al. [42] performed Coarse-Grained (CG) followed by classical all-atom molecular dynamics (MD) simulations of a TisB dimer and a TisB tetramer. Both, the dimer as well as the tetramer, have a charge zipper motif [56] which is crucial for its stabilisation. Especially the tetrameric structure of TisB showed extensive patterns of salt bridges and hydrogen bonding interactions within the dimer as well as across the subdimer. Thus, the tetramer structure is very stable and it was observed that water can pass along the continuous polar interface of the TisB assembly. Hence, they proposed that a proton transfer might occur across the lipid bilayer.

6.1.1. Interaction between TisB Peptide

The tetrameric assembly of TisB peptide is formed by dimer-of-dimer, in which two tightly bound subdimers cross over each other in an X-shape. Two peptides form a subdimer by arranging themselves in anti-parallel fashion. As already stated before, each subdimers are charge-zipped with four salt bridges and one hydrogen bond.

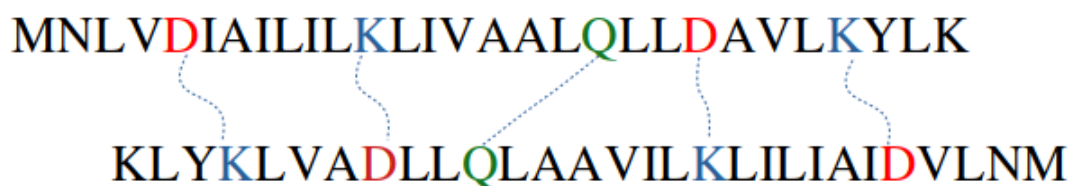


Figure 6.1.: Representation of the anti-parallel arrangements of the dimer formed by TisB. The four salt bridges are formed between lysine (shown in light blue colour) and aspartate (shown in red colour). The hydrogen bond is formed between the two glutamines (shown in green colour) which are present in the middle of the dimer assembly.

Water could be pulled into the hydrophobic bilayer core by the polar interface, which might explain how biofilms are formed. Positions 5 and 22 of the aspartic acid can form salt bridges with positions 12 and 26 of the lysine (D5–K26, K12–D22, D22–K12, and K26–D5). An additional hydrogen bond is formed by the glutamines on position 19 (Q19–Q19). TisB's charge zipper structure may be based on this motif, see fig. 6.1.

In the case of the tetrameric assembly of TisB, there is also a possibility of the formation of salt bridges across the subdimers, which provides extra stabilisation to the assembly. When one subdimer runs over the other subdimer, the lysines (K12, K26) of dimer one (d_1) can form a salt bridge with the aspartate (D5, D22) of dimer two (d_2). In a similar fashion, glutamine has now more possibilities of forming hydrogen bonds, due to the presence of four glutamines at the centre. Though, the cross-bridging is dynamic in nature and keeps on forming and breaking, indicating the flexibility of these salt bridges and hydrogen bonds. Salt bridges K12–D22 and D22–K12 can form cross bridging across the subdimers, which gives extra stabilisation to the tetramer. This stabilisation is also driven by the different modes of hydrogen bonding interaction of centrally located Q19 residue.

It has also been suggested by classical MD simulations that the water molecules are engaging in interactions with the charged and polar residue side chains along the polar "tunnel" at the centre of the TisB assembly. The water density along the TisB assembly decreases from both sides towards bilayer center. Even in the least hydrated part of the TisB assembly which is the centre of the lipid bilayer, there is always a water molecule penetrating into the centre of the TisB assembly. This indicates the hourglass-shaped body of the water molecules, compare fig. 6.6. The density of water can be symmetrical or asymmetrical, which depends upon the compactness of the tetramer assembly, from both sides toward zero of the vertical coordinate. Previous studies from our group showed that there is at least one water molecule per 0.3 nm from the centre of the tetramer TisB assembly [42]. As the water molecules are present near the centre of the peptide, they are capable of effectively forming water wires between the TisB peptide assemblies. Considering the fact that water penetrates the tetramer assembly to such a great extent, there are several questions related to the passage or the permeability of water molecules through

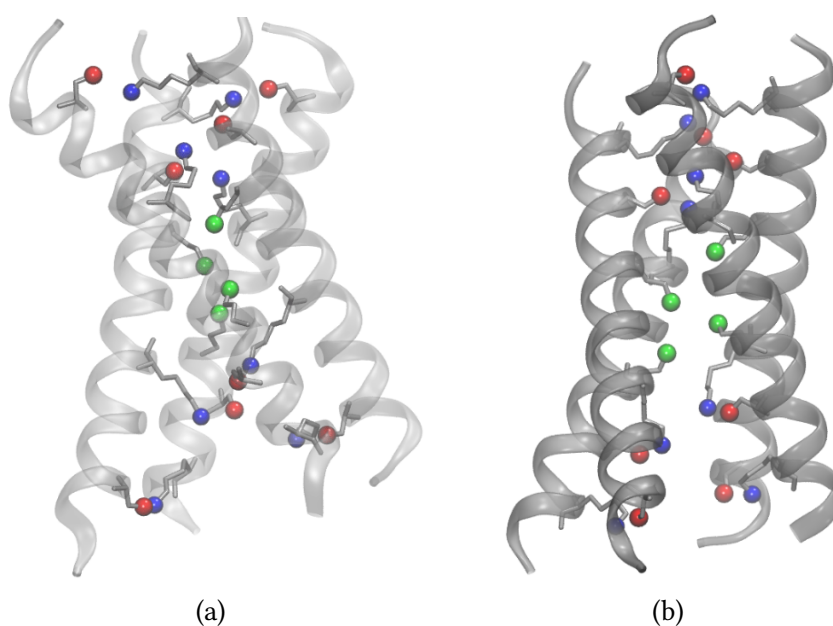


Figure 6.2.: Snapshots of TisB: (a) shows it viewed from the side and (b) is another representation of the TisB assembly, showing extended salt bridges and hydrogen bonds in a charge zipper fashion forming an anionic channel.

the assembly. Though there is no experimental evidence for the passage of water, the possibility of water transfer was never ruled out.

The main goal of this work is to study the energetics and mechanism of proton transfer through the assembly of TisB peptide. Apart from the main goal, we are also looking for the possibility of the passage of water through the assembly of TisB peptides and conditions under which water can pass through the TisB assembly. Energetics of proton transfer and passage of water are studied using QM/MM multiple walker metadynamics simulations.

6.2. Methodology

6.2.1. Classical MD Simulation

In the early stages of my research, I performed several classical Molecular Dynamics (MD) simulations to simulate the tetrameric assembly of the charge zipper TisB protein embedded in the POPC bilayer. Two final structures from classical all-atom simulations performed by Schneider et al. were taken as the starting structures for my simulations, but with modifications on some parameters. The two initial structures are referred to as structure 1 and structure 2, in this work. The entire system was enclosed in a rectangular periodic box, consisting of the protein, the lipid, water, and ions. The TisB peptide assembly was embedded in a lipid bilayer consisting of 278 POPC molecules, solvated with

11494 water molecules and neutralised with 4 chloride molecules. Gromacs [57, 58, 59, 60, 61] patched with Plumed plugin 2.5 [62] was used for the energy minimisation and equilibration.

Firstly, energy minimisation was performed using the steepest descent algorithm with a maximum of 1000 steps, until a gradient threshold (1000.0 kJ/(mol · nm)) was reached. Energy minimisation was followed by a constant-temperature (NVT) simulation at 300 K for 10 ns using velocity rescaling [63] with a coupling constant of 0.5 ps. This was followed by a constant-temperature-and-pressure (NPT) simulation of 300 K and 1 bar for 5 ns using the Nosé–Hoover thermostat [64, 65] with coupling constants of 1.0 ps. Further, production simulation of 0.5 μ s at 300 K and 1 bar was performed in an NPT ensemble using the Nosé–Hoover thermostat [64, 65], with coupling constants of 1.0 ps. Pressure during the simulation was kept constant using a semi-isotropic scheme such that the pressure in the membrane can be controlled separately. Parrinello–Rahman barostat [66] was applied with a reference pressure of 1 bar, a coupling constant of 10 ps, and compressibility of 4.5×10^{-5} bar⁻¹. The coordinates of atoms were saved every 10 ns.

Throughout all the above equilibration and production simulations, the backbone atoms have been subjected to additional torsional angle restraints that were applied to all residues except for the two residues located on the ends of the helices. The torsional restraint was restrained at an angle of -105° with a force constant of 1641 kJ/(mol·nm²). Along with this, eight distance restraints were applied on the salt bridge between the four Asp5–Lys26 and the remaining four between Asp22–Lys12, within the subdimer. Distance restraints were applied between Asp (C _{γ}) and Lys (N _{ζ}) with a force constant of 1000 kJ/(mol·nm²).

6.2.2. QM/MM Simulation

The final structures of the production simulations were taken as the initial structure for the QM/MM simulations. The QM/MM simulations and the equilibrations were set up for proton transfer in acidic as well as the basic medium. The QM region comprised the amino acid side chains of all the D5, K12, D22, K26, Q19 from all C _{β} atoms on, compare fig. 6.3. Furthermore, the QM region also included fourteen water molecules located in the proximity of these side chain residues, which are relevant for forming water wire along the assembly and facilitating proton transfer. In total, the QM zone comprised 187 atoms in an acidic medium which have one extra proton added to the QM region, which was covalently bonded with aspartate. In an acidic medium, there are a total of 187 atoms having an extra proton added to the system. This extra electron is covalently bonded with donor aspartate. In the basic medium, the QM zone consists of 185 atoms where one hydrogen is removed from one of the QM water. There is a charge of +1 in the acidic QM region, whereas in the case there is a charge of -1 in the case of the basic medium. QM atoms are treated using DFTB3 with the 3OB parameters and the remaining systems are described by a classical AMBER force field.

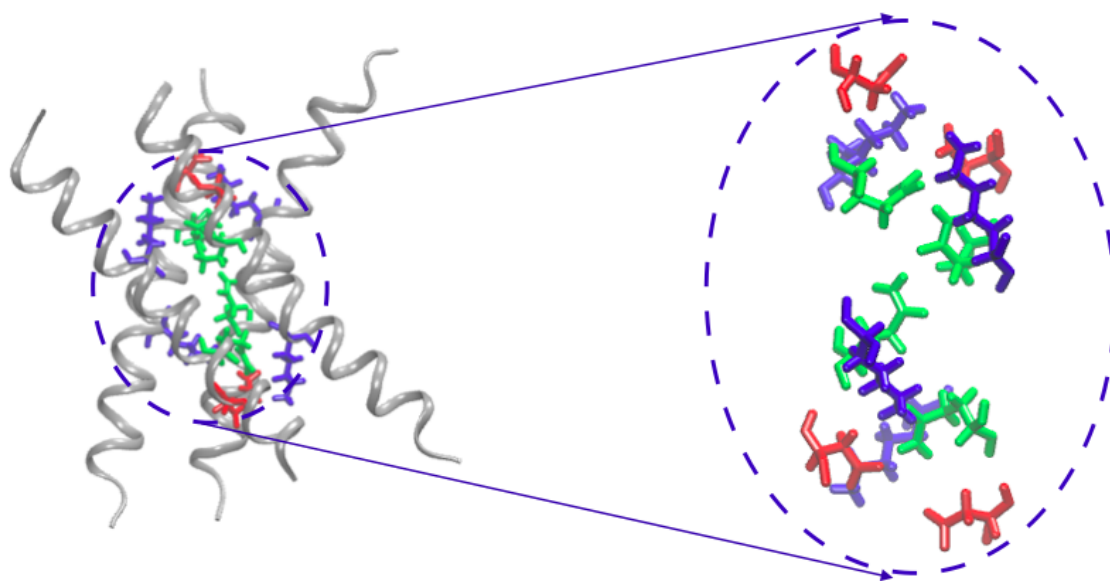


Figure 6.3.: Insight view of QM region, which consists of amino acid side chain residues of four aspartates (red), lysines (blue), and glutamines (green) each.

The transition from the QM to the MM region was done using the link atom approach, in which one hydrogen atom is placed between the C_α and C_β bond. A steepest descend minimisation algorithm was used for the QM/MM optimisation with a threshold of 1000.0 kJ/(mol · nm). The leap frog algorithm was used to solve Newton's equations of motion with a time step of 0.5 fs. The Nosé–Hoover thermostat and the Parrinello–Rahman barostat were used for temperature and pressure coupling, respectively. An NPT equilibration was conducted over 10 ps at 300 K and 1 bar followed by a production simulation over 500 ps for both acidic and basic systems.

Moreover, to prevent the leakage of the QM-treated water molecules out of the QM region harmonic position restraints were applied to the oxygen atoms of the fourteen water molecules. Thereby, the oxygen atom of each of the water molecules was allowed to move freely within 10 Å from the centre of mass of the alpha carbon of Q19. Whenever the distance from the initial position increased beyond 10 Å, a harmonic restraint with a spring constant of 1000 kJ/(mol·nm²) set in, pulling the water back within 10 Å. Additional restraints were applied along the water wires, i.e., harmonic restraints were applied to the distance between oxygens of neighbouring water molecules in order to keep the water wire intact. The spring constant was set to 1000 kJ/(mol·nm²) which was activated when the distance was greater than 3 Å. A very strong harmonic restraint with a spring constant of 50000 kJ/(mol·nm²) was applied between hydrogens and oxygens of the QM water molecules, which were not part of the QM proton transfer relay member. This is done to make sure that the proton transfers along the water wire.

In order to investigate the proton transfer in acidic and basic mediums, a long-range proton transfer coordinate was used based on the centre of excess charge (mCEC) introduced by König et al. [67]:

$$\xi = \sum_{i=1}^{N_H} \mathbf{r}^{H_i} - \sum_{j=1}^{N_X} w^{X_j} \mathbf{r}^{X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{sw} \left(d^{H_i, X_j} \right) \left(\mathbf{r}^{H_i} - \mathbf{r}^{X_j} \right) \quad (6.1)$$

The first term in equation 6.1 is a sum of all hydrogen coordinates \mathbf{r}^{H_i} , and the second term is a weighted sum \mathbf{r}^{X_j} of the positions of the heavy atoms participating in the PT process (with w^{X_j} being the number of hydrogens coordinated to X_j in the least protonated configuration). The last term corresponds to a correction that runs over all distances between hydrogens and heavy atoms and that decides which H_i and X_j are connected by bonds based on a switching function $f_{sw} \left(d^{H_i, X_j} \right)$.

$$f_{sw} \left(d^{H_i, X_j} \right) = \frac{1}{1 + \exp \left[\left(d^{H_i, X_j} - r_{sw} \right) / d_{sw} \right]} \quad (6.2)$$

The steepness and centring of the switching function are controlled by the parameters $r_{sw} = 0.125$ nm and $d_{sw} = 0.006$ nm, respectively. An excess proton being transferred in a given location can be represented by the vector variable ξ . In order to convert ξ into a scalar quantity ζ , the distances between mCEC and the initial proton donor D and between mCEC and the final proton acceptor A, are considered.

$$\zeta = \frac{d_{\xi, D} - d_{\xi, A}}{d_{\xi, D} + d_{\xi, A}} \quad (6.3)$$

For a proton that is located on the initial donor, $\zeta = -1$, while for a proton located on the final acceptor, $\zeta = 1$. There were fourteen molecules of water from the QM/MM region as well as oxygen atoms that were considered potential proton coordination sites during the PT reaction from the initial donor to the final acceptor. The QM waters which form waters wire are described by the mCEC coordinate.

6.2.3. Water Transfer – Classical Metadynamics

The final structures from the 500 ns production simulations (structure 1 and structure 2) were taken to study the water passage. For this, the water molecule which was the closest to the centre of the peptide assembly was selected and hills were deposited along the d_z reaction coordinate, i.e., the z -distance between the oxygen of the water molecule and the centre of mass (COM) of the four glutamine residues (Q19) in the centre of the pore, compare fig. 6.4. The free energy profile was obtained with well-tempered metadynamics using d_z as the reaction coordinate. The bias factor was set to 15, the width of the gaussian to 0.02 nm, and the initial height 0.5 kJ/mol. The Gaussians were deposited every 500 steps along the reaction coordinate over a total simulation time of 100 ns. The simulations were performed at 300 K and 1 bar.

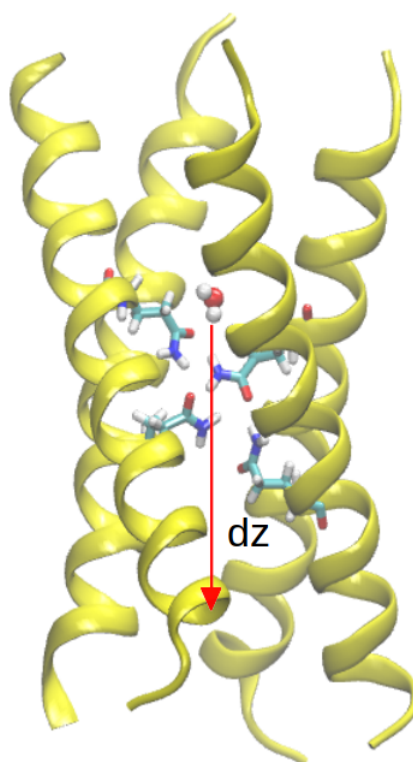


Figure 6.4.: Graphical representation of the distance (d_z) of the nearest water molecule along the bilayer normal in a collective variable through the tetramer assembly of TisB. Glutamines are present at the centre of the tetramer assembly.

6.2.4. Proton Transfer – QM/MM Metadynamics

The proton transfer was simulated with multiple walker well-tempered metadynamics, once in acidic and once in basic medium, compare fig. 6.5. The initial structures were taken from NPT equilibrated QM/MM simulations of the TisB assembly in the respective medium. Two batches of simulations were performed. For the first batch, the modified centre of excess charge (mCEC) coordinates transformed to the scalar quantity ζ were used as the reaction coordinate. In the acidic medium, $D22 - C_\gamma$ (ASP22) with the extra proton was considered as the initial donor and $D99 - C_\gamma$ (ASP99) on the other side of the assembly as the final acceptor. In the basic medium, $K109 - N_\zeta$ (LYS109) was defined as the initial donor and $K5 - N_\zeta$ (LYS5) is the final acceptor.

For the second batch, ζ was also used as a reaction coordinate but only the z-component of ζ was considered. In this case, no definition of the final donor and acceptor was required.

In our simulation, we often saw lysine residues in the QM region, get protonated and deprotonated. In order to avoid this, a very high harmonic restraint is applied on hydrogen attached to nitrogen in lysine, with a spring constant of $100000 \text{ kJ}/(\text{mol}\cdot\text{nm}^2)$, if it is not defined as the initial donor or final acceptor. In the case of an acidic medium, all four lysines in the QM region are restrained with coordination restraints between N_ζ and hy-

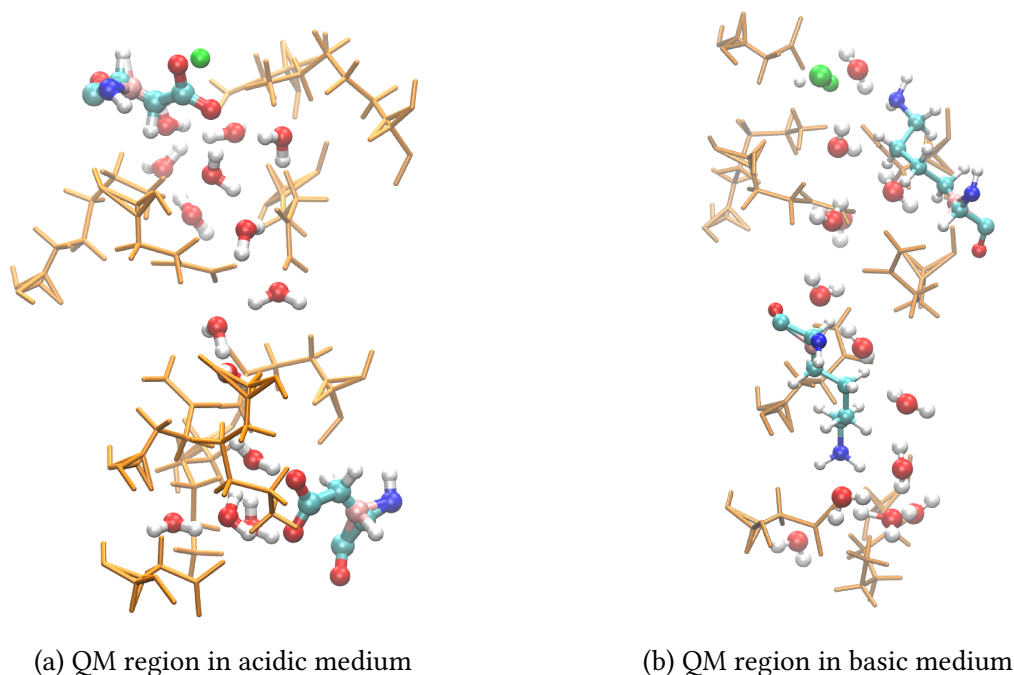


Figure 6.5.: Snapshots of the QM region in (a) acidic medium, where the extra proton is added manually into the QM region, which is shown in green colour, which is covalently bonded initial donor aspartate. In the case of an acidic medium, aspartate containing the extra proton acts as a donor, whereas the aspartate on the other side acts (lower part) as a final acceptor. (b) is the QM region in basic medium, a proton is manually removed from one of the water molecules, making it OH^{1-} , which is shown in green colour. In a basic medium, charged lysine (lower part) is considered the initial donor and neutral lysine (above) acts as the final acceptor.

drogen atom covalently bonded with it, in both ζ as well as z-component of ξ as a reaction coordinate. Whereas in the case of basic medium, two lysine residues are restrained with such a restraint when ζ and z-component of ξ reaction coordinate is used.

All simulations employed 24 walkers, with a bias factor of 60, a Gaussian height of 0.6 kJ/mol, a Gaussian width of 0.05 nm, and a deposition stride of 1 ps. The walkers exchanged their deposited Gaussians every 1 ps.

6.3. Results and Discussion

6.3.1. Water Transport

The initial structures of the tetramer TisB assembly, obtained from classical MD simulation, are well hydrated by water molecules. Thus, my simulations were able to reproduce the

earlier findings from our group [68]. It was reported that the salt bridging pattern between the residues aspartate and lysine plays a vital role in stabilising the tetrameric structure, which is consistent with our MD simulation. The salt bridges within the subdimer are more stable than the salt bridges across the subdimer. The distance of these salt bridges within the subdimer is smaller than the distance across the subdimer. It was reported that the tetrameric structure can allow the passage of water along the polar interface of the assembly, which is consistent with our simulation. The water passage is additionally influenced by the presence of glutamine at the centre of the TisB assembly. The different ways of interaction between glutamines at the central core region of the tetrameric assembly results in the opening and closing of the gate. The water density on both sides of the lipid bilayer decreases towards the centre of the TisB assembly and approaches but never reaches zero, compare fig. 6.6. Moreover, it is possible for extended salt bridges and hydrogen bonding interactions within and across the subdimers to result in an asymmetric or symmetric distribution of water along the vertical coordinate of the lipid bilayer.

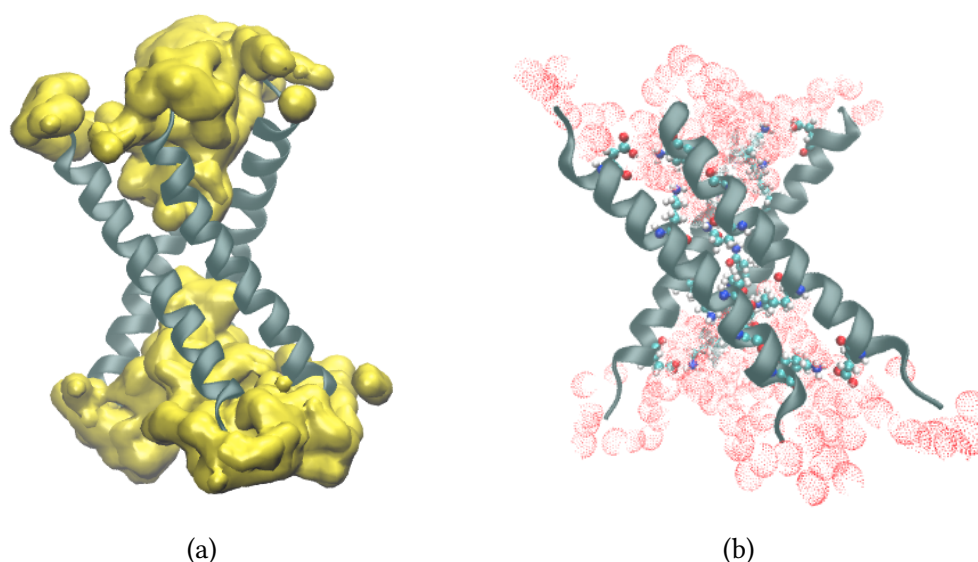
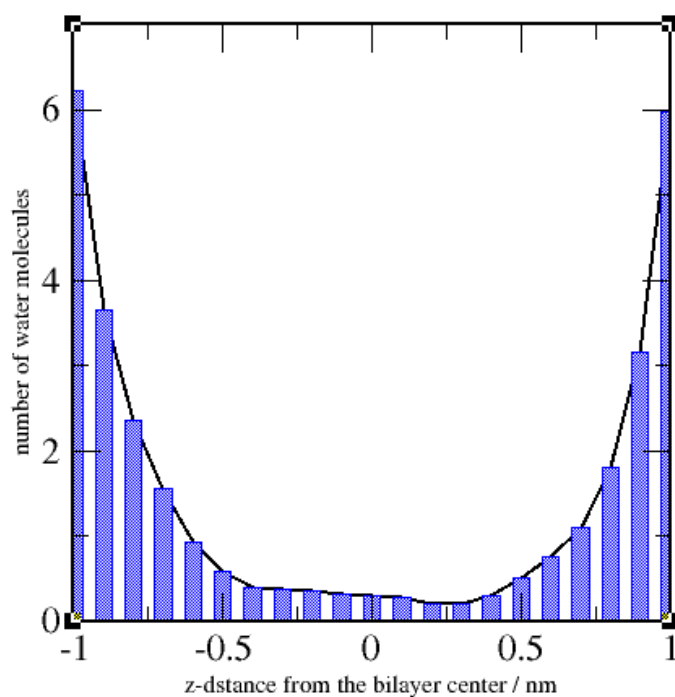
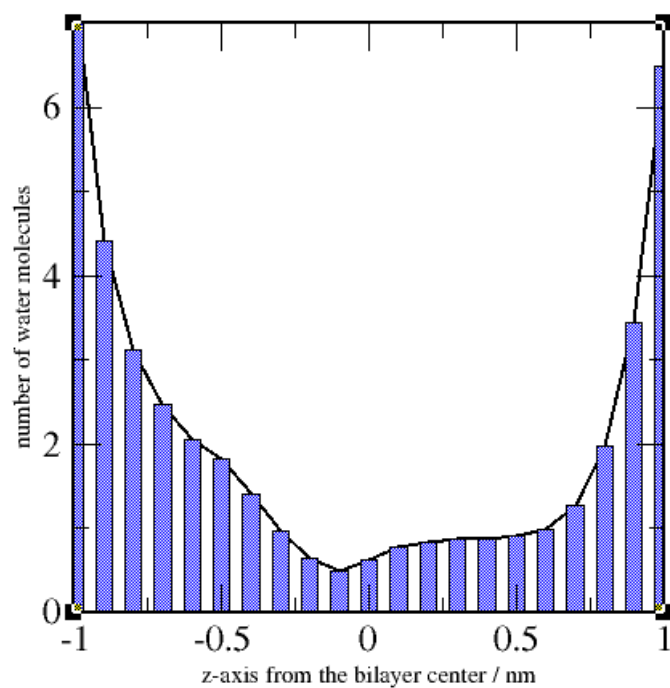


Figure 6.6.: (a) shows the hourglass-shaped of water molecule into the TisB assembly of a peptide. (b) shows the backbone of the peptide along with the charged and polar side chain solvated in the water molecule.

Previous studies suggested that it is impossible for water to cross the TisB assembly as shown in fig. 6.8a. We also find that interactions between charged residues (K12 and D22) and interactions between the glutamine residues (Q19) in the centre of the assembly can modulate the hydration level in the assembly. Through the dynamic exchange of the salt bridges and hydrogen bonds between the residues, within and across the TisB assembly, the two subdimers interact with each other in order to maintain a stable interaction and in some configurations water may pass through the centre.



(a)



(b)

Figure 6.7.: The graphical representations show the number of water molecules within the TisB assembly counted in the bin width of 0.1 nm along the lipid bilayer normal of (a) structure 1 and (b) structure 2.

In this regard, the hydrogen bonding that occurs between the glutamine (Q19) residue in the middle and the water molecules seems to regulate the water passage. In the present study, it has been observed that if there are hydrogen bonding interactions between glutamine (Q19) across subdimers, then there will not be the passage of water molecules, compare fig. 6.8b. Whenever there is no hydrogen bonding interaction between glutamine across (Q19) residues across the helix, one can observe a higher number of water molecules at the centre of the TisB assembly. There is a greater probability that water can pass through the tetrameric assembly of the peptide.

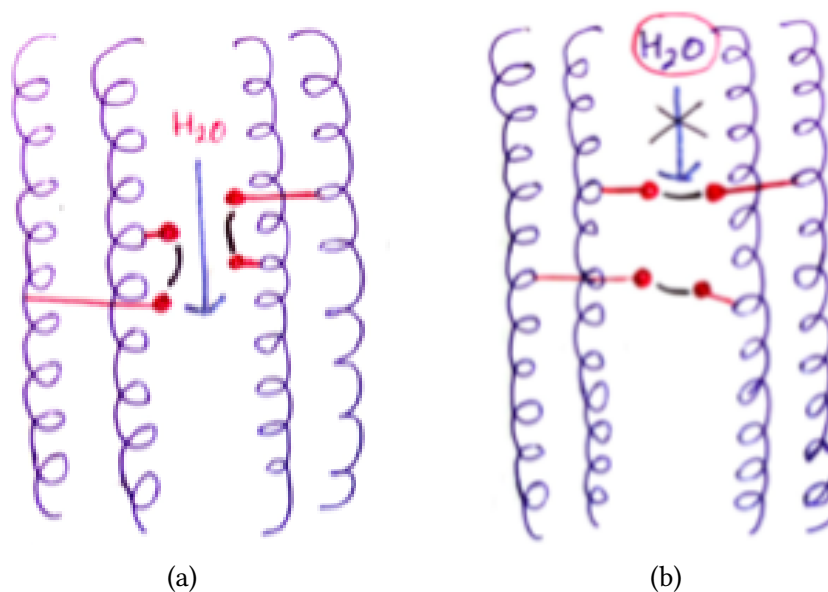
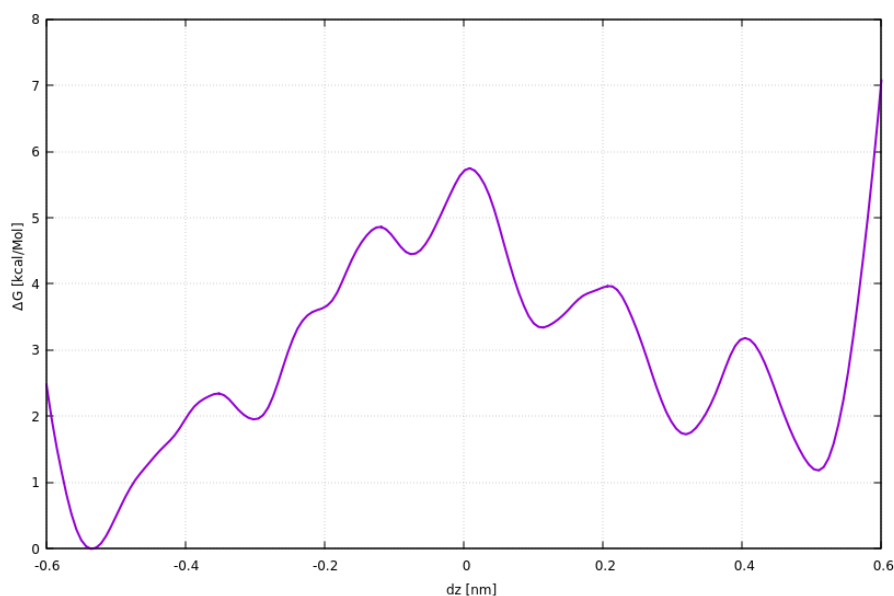


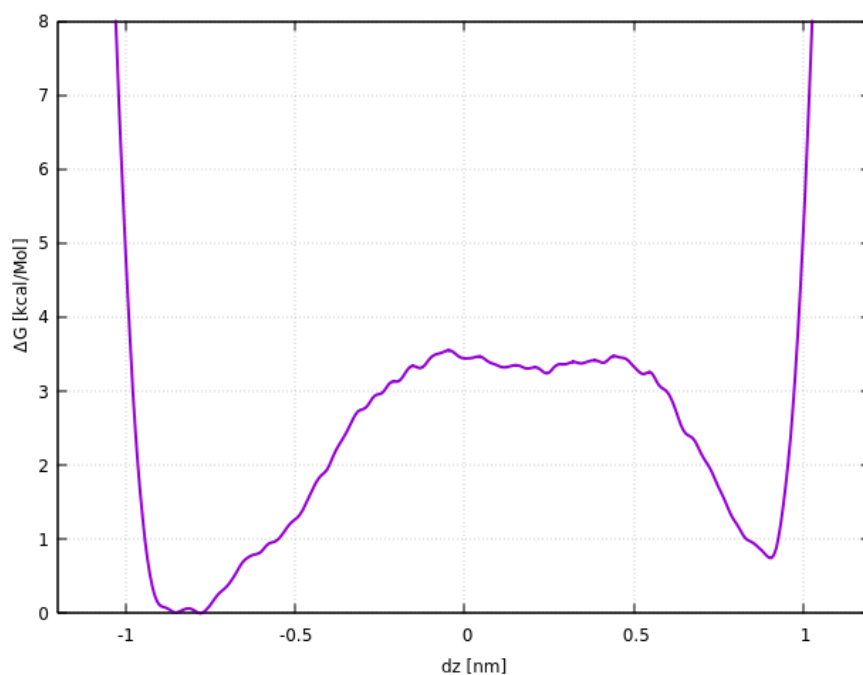
Figure 6.8.: (a) shows the pictorial representation of hydrogen bonds within the subdimers in the TisB assembly. The involvement of the glutamines (Q19) in hydrogen bonds within the subdimers results in creating more space for the water to enter. (b) shows the pictorial representation of the hydrogen bond between glutamine (Q19) across the subdimers. It makes the space narrower and results in blocking the water passage.

Symmetric or asymmetric distribution of water molecules along the lipid bilayer normal can be attributed to the salt bridges between K12–D22, D5–K26, D22–K12, K26–D5 across the dimer and the hydrogen bonds interaction Q19–Q19 across the subdimers. If the uniform salt bridges are formed across the subdimers along both sides of the TisB assembly, the water density tends to be symmetrical along the bilayer normal from both sides of the TisB assembly. Whereas, if the salt bridging is non-uniform, meaning unstable salt bridges across the subdimer, there are more chances of getting unsymmetrical distribution of water along TisB assembly.

It is even more clear from the graphical representation of water density distribution along structure 1, where the salt bridges and hydrogen bonding interactions across the subdimers are strong and uniform on both sides of the TisB assembly. As a result, the TisB assembly develops a symmetrical density distribution of water along the vertical direction



(a) ΔG free energy of the structure 1



(b) ΔG free energy of the structure 2

Figure 6.9.: (a) and (b) show the Gibbs energy ΔG of water across the bilayer normal, dependent on the collective variable d_z for structure 1 and structure 2, respectively.

of the assembly, as shown in fig. 6.7a. Whereas in the case of structure 2 (fig. 6.7b), the salt bridges and hydrogen bonding are not uniform from both sides of the TisB assembly. On one side of the TisB assembly, the salt bridge across the lysine and aspartate has been found to be stronger in comparison to the salt bridge across the lysine and aspartate on the

other side. As a result, the water density distribution along the TisB assembly in structure 2 results in an asymmetric distribution of density along the TisB assembly.

My simulations and previous studies by Schneider et al. [68] showed that the water density is never zero at the centre of the peptide assembly. Schneider et al. also suggested that the charged interface of the tetrameric charge zipper protein can facilitate the water passage across the assembly.

In order to get the underlying free energy landscape of the water passage, well-tempered metadynamics was performed, comparing fig. 6.9. Though there is no experimental evidence of the passage of water across the TisB assembly, our well-tempered metadynamics simulation reveals a low energy barrier for the passage of water. In the case of structure 1 (fig. 6.9a), the minima are at -0.5 and 0.5 nm and the barrier height at 0 nm is ~ 6.5 kcal/mol, which is lower than expected. The high barrier at the centre of the peptide is caused by the formation of hydrogen bonds between Q19–Q19 across the subdimer. The formation of hydrogen bonds obstructs the pathways for the passage of water and more energy is required for the passage of water across the assembly of TisB peptide. In the case of structure 2 (fig. 6.9b), the minima are located at -0.9 and 0.9 nm. The energy curve shows a low energy barrier of ~ 3.5 kcal/mol, which is less than for structure 1. The smaller barrier can be accounted to the fact that there is less or hardly any obstruction offered by the Q19 residues at the centre of the peptide. In both cases, the energy barrier for the passage of water is less than what one would expect considering the compact structure of the TisB assembly.

6.3.2. Proton Transfer or Proton Hole Transfer

The proton transfer is fundamental to the chemical processes that occur in solution and in biological systems. The PT mechanism by which an 'excess' proton diffuses through the hydrogen bond network of water molecules is known as the Grotthuss mechanism. Here, the proton from the initial donor hops in direction of the acceptor, via intermediate proton acceptors such as water. The presence of an extra proton on the donor makes it acidic in nature. An alternative to the Grotthuss mechanism involves protonating the acceptor first, for example by a water molecule. This mechanism involves the transfer of "proton holes" from the acceptor to the donor, for example along a chain of water molecules. In this work, such a hydroxyl ion transfer is referred to as a PT occurring in a basic medium.

6.3.2.1. Proton Transfer in Acidic Medium

The free energy profile of the PT in an acidic medium, i.e., an excess proton added to the system, is shown in fig. 6.10. The ζ values of -0.6 and 0.6 correspond to the protonated carboxyl group of D22 (donor) and D109 (acceptor), respectively. The barrier to PT from donor to acceptor is ~ 16.0 kcal/mol, whereas the energy barrier for the backreaction (acceptor to the donor) is ~ 6.0 kcal/mol. Using the zeta coordinates, the minima obtained by the free energy surface are not at equal depth, as one would expect. The unequal depth

in the minima of two acceptors and donors can be explained by the presence of extensive salt bridges between aspartate and lysine as well as a hydrogen bond between glutamine at the centre of the QM region. These salt bridges and hydrogen bonds disrupt the formation of water wire. Due to this disruption of the water wire, protons can no longer migrate on one side system. Thus, the minima are so deep on the side which has an extra proton.

In the QM region, there are 14 water molecules but only 6–7 water molecules usually participate in the formation of the water wire. Thus, there is a possibility of the proton being taken up by which water molecules, which we have seen on multiple occasions in our simulation. In this situation, the proton gets stuck with water which is not the part of the water wire members. This can create deep minima on that side, preferably the proton donor side. In our simulation, we have also seen the extra proton sometime migrate to the glutamine and stuck there for the whole time. Which can also create deep minima in some cases.

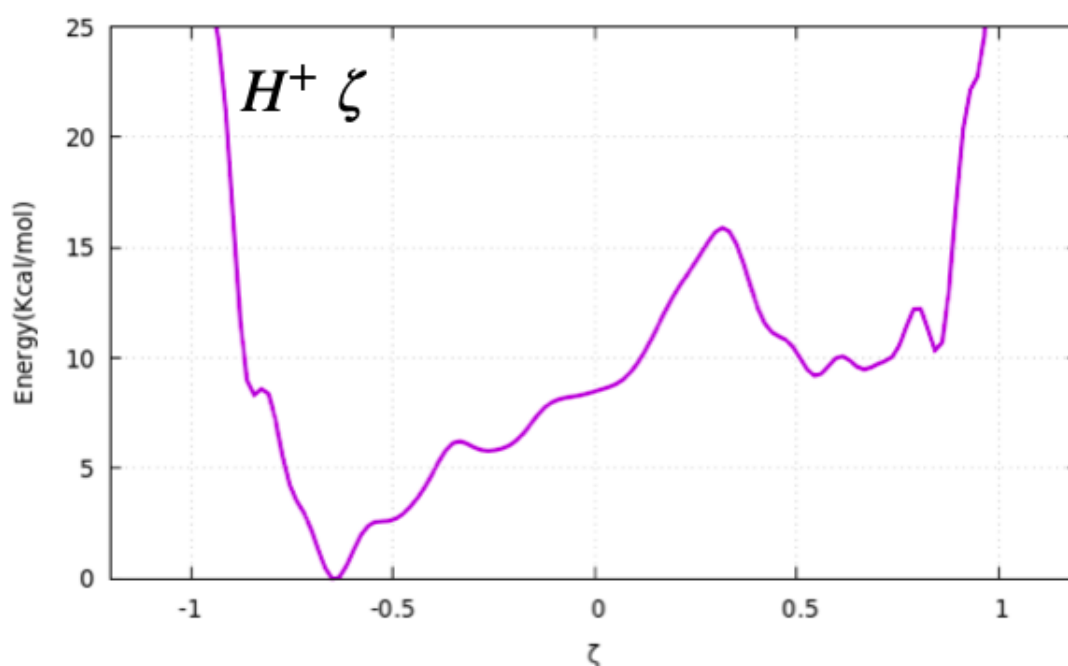


Figure 6.10.: Free energy diagram of proton transfer in the system where an extra proton is present (acidic medium) using ζ as the reaction coordinate.

Using the second batch of reaction coordinates (z -component of ξ), we observe two minima more or less the same position for the donor and the acceptor as seen in fig. 6.11. For the donor, the ΔG are observed around ~ 0 kcal/mol and for the acceptor, it is observed at ~ 12.0 kcal/mol. The barrier for the PT using this reaction coordinate is ~ 9.0 kcal/mol, which is almost half the energy barrier obtained from ζ reaction coordinates. Nevertheless, these energy curves had not reached convergence, and simulations are still running to get the converged energy curves, with changed settings in PT coordinates. The oxygen of the

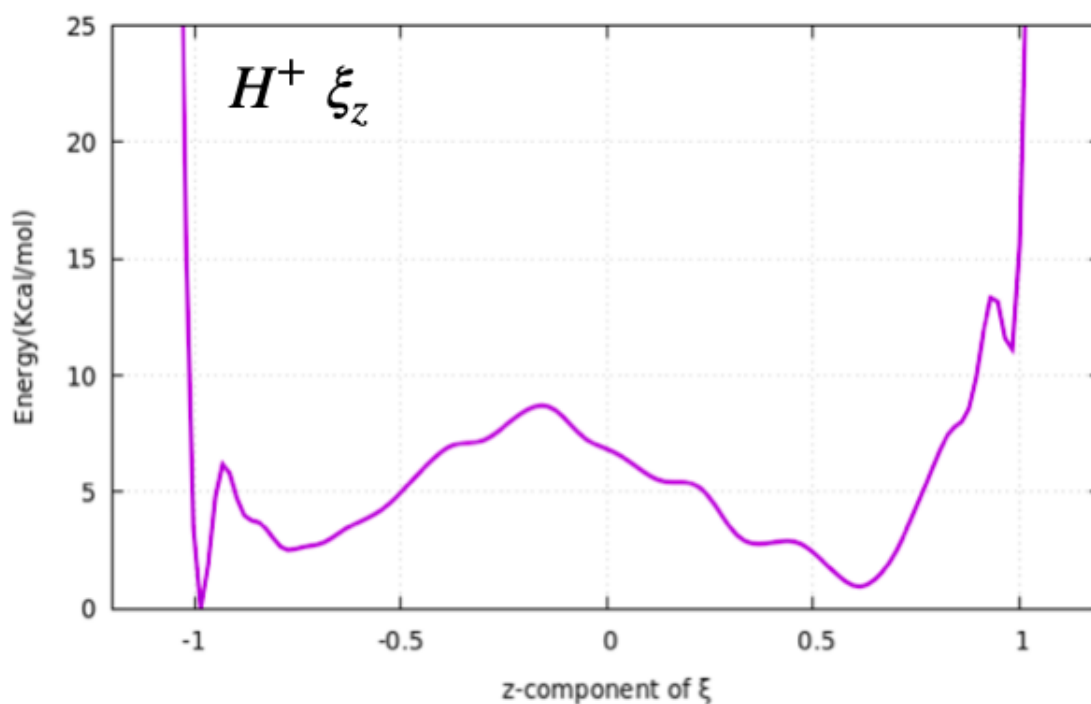


Figure 6.11.: Free energy diagram of proton transfer in the system where an extra proton is present (acidic medium) using z-component of ξ as the reaction coordinate.

water involved in the water wire is restrained using bias potential, making sure that an extra portion goes into the water wire.

6.3.2.2. Proton Transfer in Basic Medium

The free energy profile of the PT in a basic medium, i.e., a proton is removed from the system is shown in fig. 6.12. The ζ value of -1 corresponds to the charged lysine which has three hydrogens attached to N_ζ of K12. The ζ value of +1 corresponds to the neutral lysine (K107) which has two hydrogens attached to the N_ζ . PT in the basic medium using ζ reaction did not yield a well-defined energy barrier height and the barrier height keeps on increasing from donor to acceptor. Though a deep minimum is observed for the proton donor lysine (K12) in this case.

Using the second batch of reaction coordinates using the z-component of ξ , the free energy profile of the PT is shown in fig. 6.13. In this case, there is no need to assign an acceptor or donor atom. Two minima of unequal depth can be seen, which is not expected. The ΔG for the donor is observed at 0 kcal/mol, whereas the ΔG for the acceptor lysine is observed at ~ 6.0 kcal/mol. The energy barrier for PT from donor to acceptor is ~ 13.0 kcal/mol and the energy barrier for the backreaction is ~ 7.0 kcal/mol.

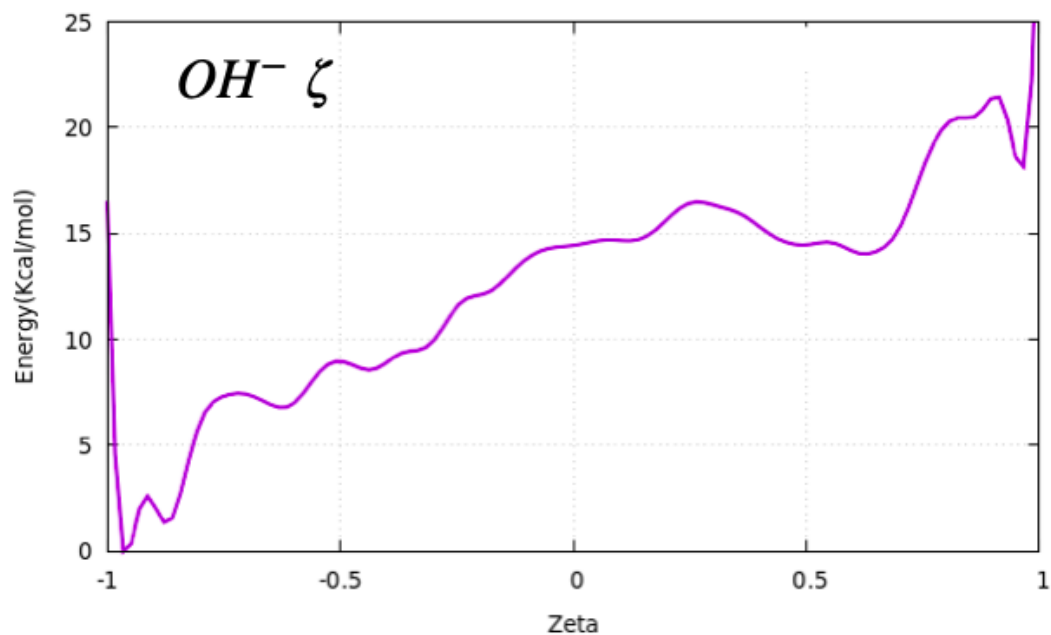


Figure 6.12.: Free energy diagram of proton transfer in the system where a proton is removed (basic medium) using ζ as the reaction coordinate.

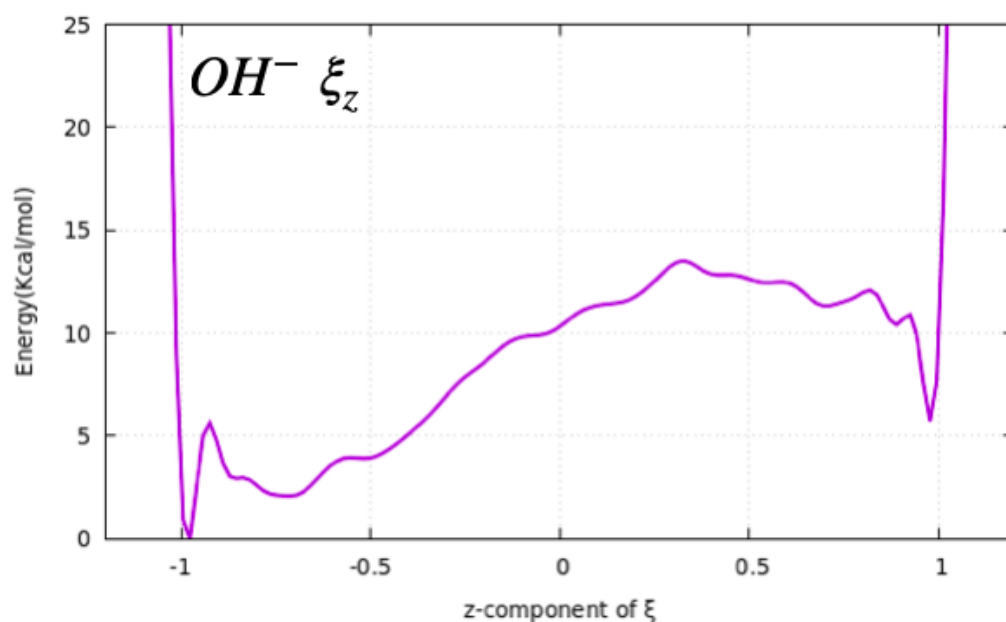


Figure 6.13.: Free energy diagram of proton transfer in the system where a proton is removed (basic medium) using z-component of ξ as the reaction coordinate.

As discussed, there were many problems encountered in PT in both mediums (acidic and basic). These problems include the interference of MM water entering the QM region, the uptake of a proton by the non-wire water molecules in the QM region, protonation of the GLN19 residues as well as the disruption of water wire. In order to encounter these difficult situations we have employed different strategies. We have restrained QM water such that it remains in the QM region and thus the MM water entering into the WM region is hindered. We have also restrained the water wire in the QM region such that the water wire remains enacted through the PT. In a similar way, a harmonic restrain was also applied to GLN such that it never gets protonated.

To conclude the proton transfer in acidic and basic medium, the energy barrier is found to be an upper bound. The work is still in progress and looks good with the new settings.

7. Structural Analysis of E5-PDGFR β Transmembrane Proteins

7.1. Introduction

A central role played by transmembrane receptor proteins in signal transduction is their ability to transmit signals across cell membranes. Biological functions have led to the classification of membrane proteins into many different families. Adhesion molecules, such as integrins and selectins, are crucial to cell-to-cell communication [69]. Proteins that transport molecules across membranes, including channel and carrier proteins, play a critical role in the transfer of molecules. Another major class of transmembrane proteins are transduction receptors, which transmit signals from the external environment to the inside of the cell, thus resulting in specific cellular responses [70]. Examples are receptor tyrosine kinases (RTK) and G-protein coupled receptors [71].

7.1.1. Tyrosine Kinases Receptor

A receptor tyrosine kinase is an important component of the signal transduction pathway, which is involved in regulating cell growth, differentiation, migration, and apoptosis [72]. Platelet-derived growth factor receptors (PDGFR) β are typical members of the RTK family. The protein consists of three domains: an extracellular domain that binds PDGF, a single transmembrane domain (TMD), and a cytoplasmic domain that operates as a tyrosine kinase. Due to its pathological functions, particularly in tumour and cancer cells [73, 74, 75, 76], it has been the subject of intense research for several decades.

There are many growth factors that are associated with oncological diseases within the RTK family, including insulin, skin growth factors, platelet-derived growth factors (PDGF), fibroblast growth factors (FGF), vascular endothelial growth factors (VEGF), and hepatocyte growth factors (HGF). There is an abundance of them in all metazoans, including sponges and humans [77]. A monomeric PDGFR β is present when it is inactive and can be activated by the binding of PDGF, which results in the dimerisation of the receptor.

From *Caenorhabditis elegans* to humans [78], all RTKs have a similar molecular architecture, as seen in fig. 7.1, that consists of a ligand-binding extracellular domain, a transmembrane domain (TMD) and an intracellular domain. Extracellular regions of RTKs can vary significantly in length and subdomain composition, which defines RTK subfamilies. An RTK from the extracellular region of the membrane is composed of a linear array of domains such as immunoglobulin receptors (Ig)-like domains, cysteine-rich domains,

fibronectin-like domains, and epidermal growth factor (EGF)-like domains. In contrast, intracellular domains are more uniformly organised. The intracellular domain consists of a juxtamembrane domain (JMD), followed by a tyrosine kinase domain (TKD) and a C-terminus [79, 80]. TKD lengths are relatively constant across RTK families, whereas JMD and C-terminal lengths differ strongly among RTK families, depending on the specific nature of intracellular signals [81]. Furthermore, the total amount of tyrosine found within the intracellular region as well as its distribution is quite different. [82] It is important to note that TMD are inserted into the cell membrane in order to connect the extracellular and intracellular domains of all RTKs.

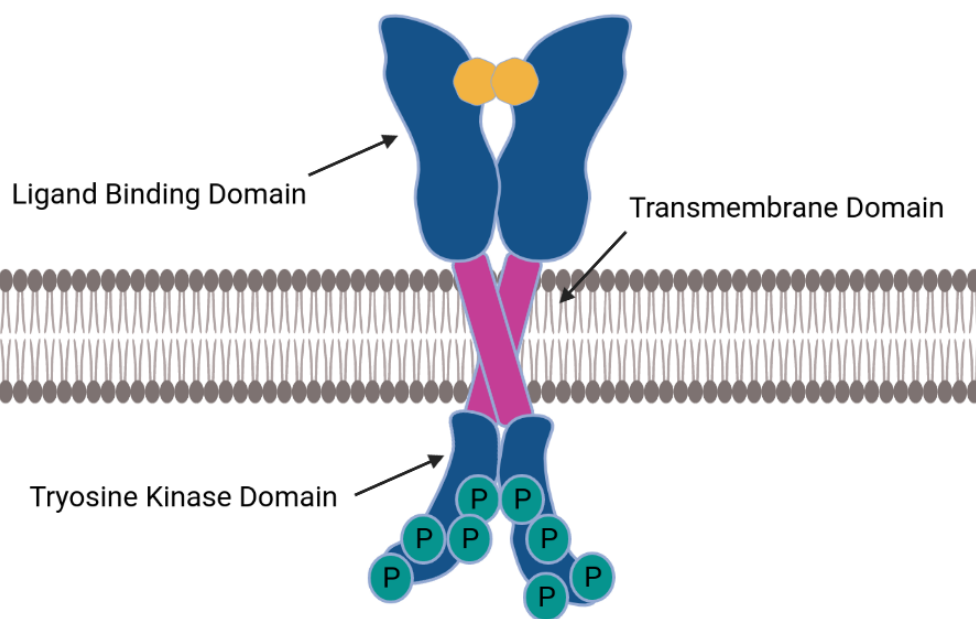


Figure 7.1.: Schematic representation of receptor tyrosine kinases (RTK). It consists of an extracellular ligand-binding domain, a transmembrane domain (TMD) and an intracellular region. The TMD connects the extracellular and intracellular domains through a helical structure in the cell membrane.

Crystal structures of receptor tyrosine kinases have revealed a lot of information regarding the mechanism of receptor activation [80, 79]. It is generally believed that receptor oligomerisation, especially dimerisation, occurs in response to ligand binding to the extracellular domain of RTK [78, 74]. There are however cases where a receptor can form a dimer without any ligand, such as insulin and the EGF receptor [83]. The binding of the ligand is necessary for the receptor to be activated regardless of whether the inactive state is monomeric or oligomeric.

Several studies have shown the ligand dimerisation technique, where the monomeric receptors form a disulfide bond with the dimeric ligand, such as PDGF or VEGF [84, 85]. Additionally, there is growing evidence that many receptors exist in the dimeric form before ligand binding, and activation of these receptors is not explained by the ligand-induced

dimerisation hypothesis. An alternative mechanism referred to as the "rotation model" has been proposed for this process, which has been demonstrated for a variety of receptors, including Neu receptors and EGF receptors [86, 87, 88, 89, 90]. This dimeric receptor has a rotationally flexible extracellular domain, and the binding of ligands results in conformational changes that allow rotation of TMD and rearrangement of the intracellular domain. In addition to TMD rotation, the angles and distances between interhelical crossings are altered. As a result, the activation of TKD is favoured. Monomeric receptors are dimerised by noncovalently binding ligands to their extracellular portions. These dimeric receptors undergo structural changes that facilitate the process of tyrosine autophosphorylation between cytoplasmic domains.

In light of the fact that RTKs play a critical role in cell proliferation, it is not surprising that a wide range of diseases are associated with RTKs. A number of cancers, diabetes, inflammations, severe bone disorders, arteriosclerosis, and angiogenesis are associated with the activation of RTK mutations. In addition, overexpression of RTKs is frequently detected in animal and human tumours. The EGF receptor was the first receptor to confirm that overexpression of receptors is associated with cancer [91, 92]. Further structural and functional studies are required to gain a more comprehensive mechanistic understanding of these receptor classes and open new avenues for diagnosis and treatment.

7.1.2. Oncoprotein E5

There are 44 amino acids in the protein E5 from the bovine papillomavirus, most of which have a hydrophobic nature. This is the smallest oncoprotein known. In transformed cells, it is primarily present in the Golgi and endoplasmic reticulum membranes [93, 94]. A highly specific interaction between E5 and platelet-derived growth factor receptor (PDGFR β) leads to the activation of the signal cascade independent of the ligand, resulting in cancer [94, 95, 96].

The structure of E5 itself is believed to form stable homodimers due to possible hydrophobic contacts in the transmembrane region as well as two disulfide bridges at the C-terminus [97]. As a result, it has been shown that truncated transmembrane segments of the E5 protein do not require the Cys37-Ser38-Cys39 motif for dimerisation, and the hydrophobic helical portion of the protein can self-dimerise [96, 98, 99]. E5 homodimer stability is predicted to be dependent on hydrogen bonds between Gln17 of the two helices [96, 100]. Additionally, MD simulations also predicted two dimerisation interfaces in addition to the Gln17-Gln17 interaction [97]. Two interfaces are referred to as cluster 1 and cluster 2. Cluster 1 has interaction between Ala14, Leu18, Leu21, and Leu25. Cluster 2 has interaction between Leu10, Ala14, Gln17, Leu21, and Leu24. As Gln17 lies within cluster 2, it is preferred by membranes to the conformation mediated by cluster 2. In recent years, Ulrich's group has elucidated the membrane alignment of E5 alone, providing evidence for the speculation above [101]. E5 is also found to self-associate into dimers even without the interaction of these motifs. Thus, the truncated sequence from Met1 to Glu36 was used in

Ulrich's experiments as well as these computational studies.

7.1.3. PDGF-Receptor β

It is also well known that the PDGF receptor has two subunits, PDGFR α and PDGFR β , both belonging to the RTK family [102]. It is believed that both are induced by the growth factor, PDGF, which plays an important role in regulating cell growth, mobility, and development [102, 74]. It has been known for several decades that these receptors play an important role in a variety of cellular processes through different signal transductions. Their pathological functions, especially in tumours and cancer cells [76, 74, 73, 75], have made them the subject of intensive research. Oncoprotein E5 activates PDGFR β via its transmembrane domain (TMD) [103, 104]. Despite the close relationship between PDGFR α and PDGFR β [105], E5 does not bind or activate PDGFR α , implying that E5 is highly specific in its recognition of PDGFR β .

PDGFR β is primarily monomeric when it is inactive and could be activated by binding to PDGF, which induces receptor dimerisation. A liquid-state NMR study of detergent micelles has found a stable left-handed dimer with a crossing angle of approximately 20° [106]. In the same manner as E5, hydrophobic contacts across the interface play an integral role in causing the interface to self-assemble.

It has therefore been proposed that dimerisation is mediated by a leucine zipper-like motif consisting of Ala505, Ala508, Leu512, Ile515, Ile519, and Met522. As has been demonstrated by solid-state ¹⁵N NMR, this motif is incompatible with the membrane orientation of the protein PDGFR β . It appears more likely, that dimerisation occurs through another motif, which has been proposed to take part in PDGFR β oligomerisation, namely Val502, Ala505, Leu509, Leu512, Ser516, and Ile519, as well as Leu523 [96].

7.1.4. E5-PDGFR Complex

A stable complex may form between E5 and PDGFR β in order to activate this receptor [107, 108, 104, 109]. The E5-PDGFR β complex is stabilised by two interactions, namely hydrogen bonds between Gln17 from E5 and Thr513 from PDGFR β and a salt bridge between Asp33 from E5 and Lys499 from PDGFR β [110, 111, 112, 110]. Compared to other amino acids, Gln17 and Asp33 are highly conserved in E5. Further, earlier mutational analyses of E5 and PDGFR β have identified several other residues that are involved with the interaction, such as Leu10 and Leu24 in E5 and Ile506 and Leu509 in PDGFR β . There is, however, still some uncertainty regarding the mechanism of this process.

Various models based on mutational experiments and molecular modelling have been proposed as a result of the fact that the structure of the E5-PDGFR β complex has not yet been determined. In accordance with the model most commonly used, E5 lies in an anti-parallel orientation with respect to the PDGFR β -TMD. This proposed model agrees with

previous findings that PDGFR β must be activated by dimerised E5 to function effectively, and that binding to E5 will result in dimerisation of the receptor [94, 95].

In a recent study, MD simulations were used to verify a new hexameric model consisting of two E5 dimers and a dimer of PDGFR β TMD in the middle. As a result of the development of a four-helix model, in which both the E5 dimer and PDGFR β dimer adopt left-handed crossing angles, a six-helix model was developed. The backbone of the PDGFR β dimer shows fewer fluctuations when the E5 dimer is added to the simulation, and the overall complex shows better stability than the four-helix complex. Because of the hydrogen bond between Gln17 and Thr513 and the salt bridge between Asp33 and Lys498, the six-helix complex could be favoured [113].

Further, the structural characterisation of E5-PDGFR β lacks evidence of interaction patterns. A tetrameric complex and hexameric complex models (as shown in fig. 7.2) were developed using solid-state 2D ^{15}N NMR, where these specific helix orientations and electrostatic interactions were considered. A dimer of E5 is surrounded by two single strands of PDGFR β proteins in tetrameric models. The change in orientation allows hydrogen bonds to form between Gln17 of E5 and Thr513 of PDGFR β as well as a salt bridge between Asp33 and Lys499 of PDGFR β . Interestingly, the proposed hexameric model that is used in this study differs from what was previously suggested using MD simulations [113].

Our experimental collaborator has proposed a new model in which the PDGFR β dimer is composed of two right-handed dimers. In contrast, in the model proposed earlier, this dimer is composed of two left-handed dimers. However, two dimers of E5 still interact with one dimer of PDGFR β in this new model 7.2. The E5 dimers stabilise a PDGFR β molecule by interacting with Gln17 of one E5 and Asp33 of the other because Thr513 and Lys499 in PDGFR β are simultaneously interacting with Gln17 and Asp33 of the two E5 dimers. Moreover, these models provide a better understanding of the mechanism through which E5 activates PDGFR β .

In this work, I have used these proposed tetrameric and hexameric models for all-atom simulations using state-of-art computer simulation. In the next section, I have explained in detail the steps followed for setting up the free and restrained simulations.

7.2. Methodology

Four idealised α -helical peptide chains were assembled manually to roughly form a tetrameric complex corresponding to the experimentally observed contacts. This structural model of the tetrameric E5-PDGFR complex was embedded in a C20:1(11c) di-gondoic acid (DGPC) lipid bilayer in a rectangular box sized $7.53 \times 7.53 \times 10.66 \text{ nm}^3$ using the CHARMM-GUI membrane builder [114]. The CHARMM36 force field was used to describe the entire system constituting protein, lipid, water and ions. The peptide-membrane complex was solvated with 10582 CHARMM-TIP3P water molecules and was neutralised

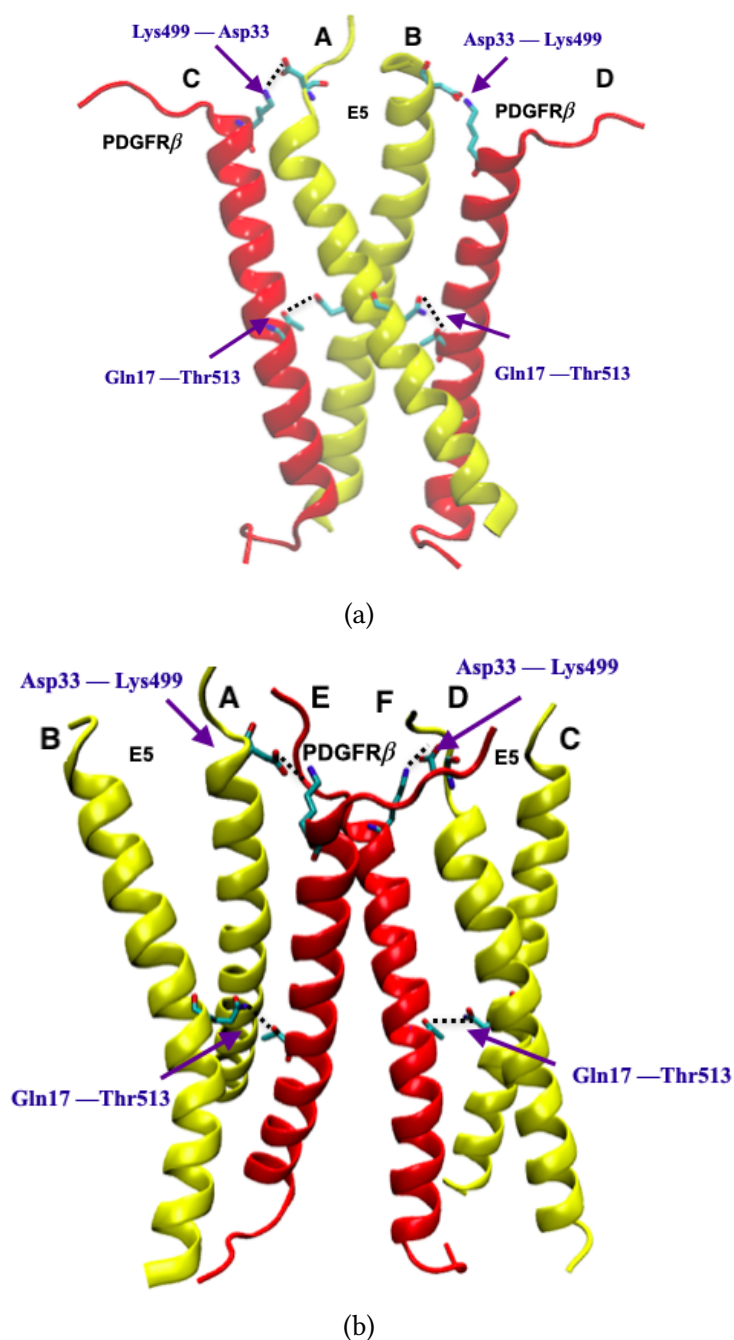


Figure 7.2.: Models for E5-PDGFR β complex. Based on orientation measurements, a model for the E5-PDGFR β complex has been proposed as (a) a tetrameric and (b) a hexameric form. The residues participating in intermolecular interactions are highlighted in both images.

with 2 chloride ions. The entire molecular system consists of 150 amino acid residues, 149 DGPC lipids, 10582 TIP3P waters, and two chloride counter ions. Gromacs version 5.0 [57, 58, 59, 60, 61] patched with Plumed plugin 2.1.1 [62] was used for the energy minimisation

and equilibration.

Energy minimisation was performed using the steepest descent minimisation algorithm with a maximum force threshold of 1000 kJ/(mol·nm). Various restraints were applied during minimisation and NVT equilibration: position restraints of 4000 kJ/(mol·nm²) on the backbone atoms and of 2000 kJ/(mol·nm²) on the side chain heavy atoms of the peptide. Position restraints were also applied between nitrogen (N) and phosphorus (P) of the lipid head groups, with a force constant of 1000 kJ/(mol·nm²) and three dihedral restraints in each of the lipid molecules (C1–C3–C2–O21, C210–C211–C212–C213, and C310–C311–C312–C313) with a force constant of 1000 kJ/(mol·rad²). The minimised structure was equilibrated employing a series of six NVT MD simulations employing the Berendsen thermostat, during which the restraints were relaxed gradually. Different force constants for position and dihedral restraints were applied throughout the equilibration, which are shown in table 7.1.

Equil. step	Time	Backbone	Side Chain	Lipid	Dihedral
1	125	4000	2000	1000	1000
2	125	2000	1000	400	400
3	125	1000	500	400	200
4	500	500	200	200	200
5	500	200	50	40	100
6	500	50	0	0	0

Table 7.1.: Lengths of simulations and force constants of position and dihedral restraints applied during each of the NVT equilibration step; lengths in ps, values for position restraints in kJ/(mol·nm²) and those for dihedral restraints in kJ/(mol·rad²).

For the equilibrations and production simulation, the leap frog integrator was applied with a time step of 2 fs. All covalent bonds were constrained to their respective equilibrium lengths using the LINCS algorithm [33]. All equilibrations were performed in the NPT ensemble, and the temperature was kept constant at 310 K using the Berendsen thermostat, with a coupling constant of 1.0 ps. The pressure during equilibrations was maintained constant with a semi-isotropic scheme. The Berendsen barostat was applied with a reference pressure of 1.0 bar and compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. The productive simulation of 2 μs was performed in the NPT ensemble, and the temperature of the entire system was kept constant at 310 K using the Nosé–Hoover thermostat [64, 65], with coupling constants of 1.0 ps. The pressure was maintained constant with a semi-isotropic scheme so that the pressure in the membrane plane was controlled separately from the pressure in the membrane’s normal direction. The Parrinello–Rahman barostat [66] was applied with a reference pressure of 1 bar, a coupling constant of 5 ps, and compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. The coordinates of atoms were saved every 10 ns.

All of the equilibrations and production simulations involved four additional distance restraints (“upper walls” of Plumed). These restraints were applied on the two salt bridges

between Asp33 & Lys499 and two hydrogen bridges between Gln17 & Thr513. In the case of the tetramer, distance restraints were applied between Asp33(C γ) of E5 (chain A) and Lys499(N ζ) of PDGFR (chain C) with a force constant of 500 kJ/(mol \cdot nm²). The same pair of distance restraints were applied between chain B and chain D. Also, distance restraints were applied between Gln17(N ϵ 2) of E5 (chain A) and Thr513(O γ) of PDGFR (chain D) with a force constant of 1000 kJ/(mol \cdot nm²). The same pair of distance restraints were applied between chain B and chain C. All of these distance restraints set in whenever the distance exceeds 3.5 Å.

In the case of the hexamer, the distance restraints were also applied between Asp33(C γ) of E5 (chain A) and Lys499(N ζ) of PDGFR (chain E) with a force constant of 500 kJ/(mol \cdot nm²). The same pair of distance restraints were applied between chain F and chain D. Also, distance restraints were applied between Gln17(N ϵ 2) of E5 (chain B) and Thr513(O γ) of PDGFR (chain E) with a force constant of 1000 kJ/(mol \cdot nm²). The same pair of distance restraints were applied between chain F and chain C. All of these distance restraints set in whenever the distance exceeds 3.5 Å.

VMD was used to visualise the trajectories and for rendering the snapshots. Analysis and plotting of data was performed with the GROMACS tools and Matplotlib libraries.

7.3. Results and Discussion

The proposed tetrameric (two E5 chains and two PDGFR β chains) and hexameric (four E5 chains and two PDGFR β chains) structure models of oligomers of E5 and PDGFR β were embedded in a solvated DGPC bilayer. Both of these structures were subjected to two microseconds of partially restrained molecular dynamics simulation. The proteins were restrained to varying degrees by three different simulation protocols for each of the two molecular complexes. The three simulation protocols are as below:

1. All the equilibration and production simulations were performed freely (SP1).
2. The distance restrains between Asp33 & Lys499 and two hydrogen bridges between Gln17 & Thr513 were applied during equilibration and released during productive simulation (SP2).
3. Distance restrains between Asp33 & Lys499 and two hydrogen bridges between Gln17 & Thr513 were applied throughout the equilibration and productive simulation. (SP3)

All the MD trajectories of both of these protein complexes were analysed, which we discuss in more detail below.

7.3.1. RMSD and RMSF

For quantifying differences in macromolecular structure and dynamics, root-mean-square deviation (RMSD) is undoubtedly the most popular method. The method is commonly

used to establish the equilibration period and the quality of biomolecular simulations and to cluster similar conformations in the analysis of molecular dynamics (MD) trajectory data [115, 116]. The RMSD is computed by performing a least-squares fit between two superimposed structures to minimise their differences [117]. As illustrated in the figure given below 7.3, a reference was made to the initial structure as well as the average structure of the trajectory. In comparison with the hexamer complex 7.4, the tetrameric complex equilibrates faster with all three simulation protocols. With all three simulation protocols, the tetrameric complex attains its equilibrium structure roughly within 400 ns. For the tetrameric complex, the RMSD accounting for deviations from the initial structure was ~ 4 Å with simulation protocol 1, ~ 3 Å with simulation protocol 2, ~ 4 Å with simulation protocol 3 (see figures 7.3a, 7.3c and 7.3e, respectively).

On the other hand, the hexameric complexes keep on changing from the initial structure with all three simulation protocols. The RMSD accounting for deviations from the initial structure was ~ 6 Å with SP1, ~ 7 Å with SP2, ~ 5 Å with SP3 (fig. 7.3b, 7.3d and 7.3f, respectively). In all the simulation protocols, only the all restrained simulation (SP3, fig. 7.3f) seems to be equilibrated after 500 ns. When compared to the tetrameric complex, the hexameric complex has undergone more structural changes, which suggests that the tetrameric complex is more stable.

In the case of hexamer simulations, the application of restraints stabilises the hexameric complex. Simulating without any restraints (Fig. 7.3b) leads to an increasing RMSD for hexameric complexes and failure to reach equilibrium conformation. The application of restraints during equilibration and the performance of a free productive simulation also did not converge to an equilibrium conformation: the RMSD keeps on changing and the final structure has a RMSD value greater than ~ 7.5 Å as seen in fig. 7.3d. By applying restraints throughout the equilibration and productive simulation, the RMSD value (Fig. 7.3f) is lower than in the other two cases. Despite the fact that the all-restrained simulation did not reach an equilibrium conformation, the applied restraints have a significant impact on the RMSD. There were no notable changes in the structure from the initial structure in comparison to the two previous simulations. This indicates that the interaction between Lys499-Asp33 and Thr513-Gln17 is crucial for the stability of the hexameric complex. It is true that all the hexameric simulations have not reached equilibrium, but the final structure is similar in all simulations.

In contrast to RMSD, Root Mean Square Fluctuation (RMSF) is a measure of residue flexibility as it indicates how much a single residue moves (fluctuates) over time [118, 119]. The results of this study provide an improved understanding of protein flexibility and structural fluctuations. It is common to plot the RMSF per residue against the residue number and this information can be used to determine structurally which amino acids have the greatest impact on molecular motion.

The RMSF profile of each of the tetrameric and hexameric complexes was scanned. It indicates that the RMSF profiles of all the helices are similar in the case of the tetramer and hexamer. In both complexes (tetramer and hexamer) atoms that have the greatest flexibility

7. Structural Analysis of E5-PDGFR β Transmembrane Proteins

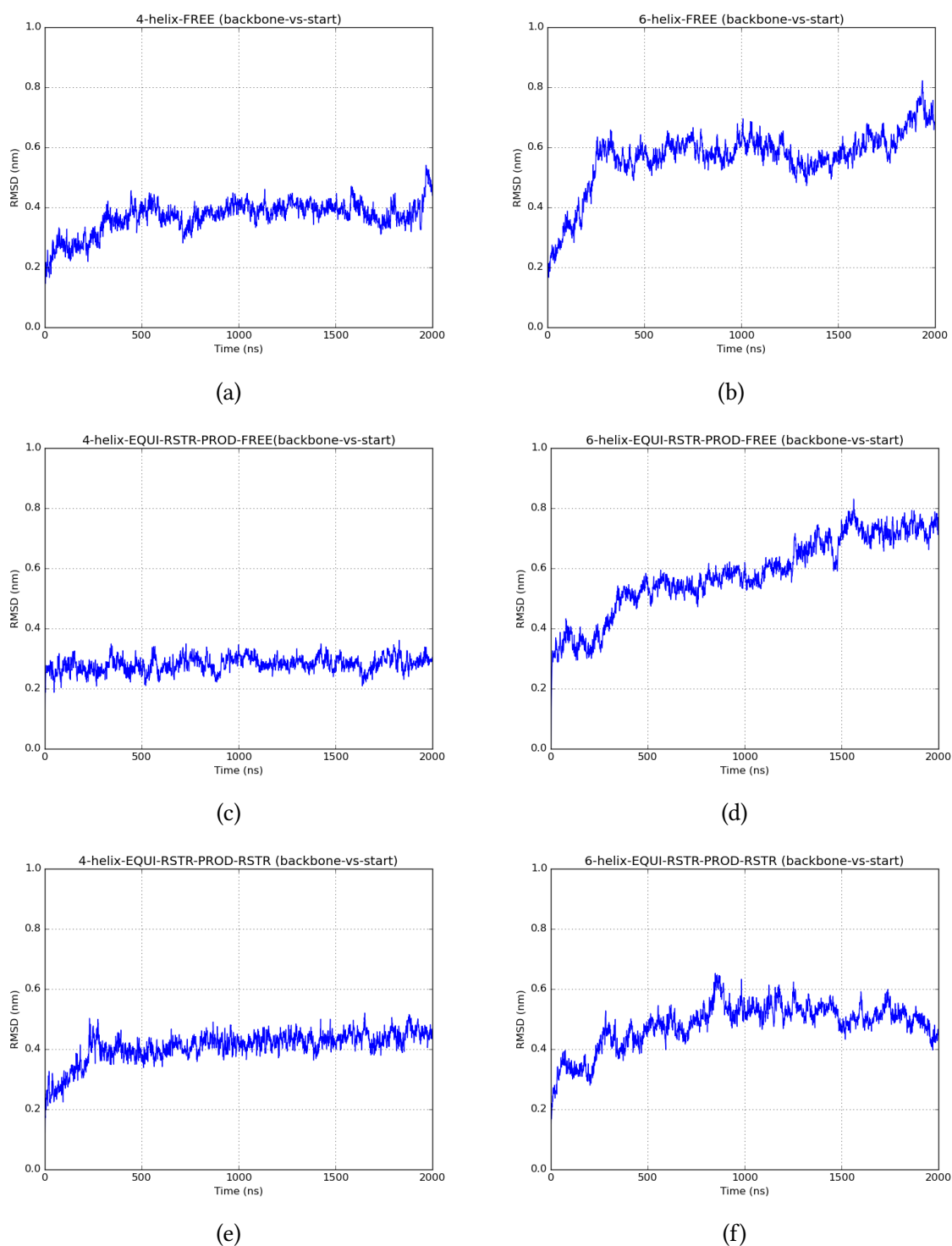


Figure 7.3.: Backbone RMSD analysis of 4-helix and 6-helix complexes with three simulation protocols. (a), (c) and (e) show the backbone RMSD plots of the 4-helix complex among all simulation times using the starting structure as a reference. (b), (d) and (f) show the backbone RMSD plots of the 6-helix complex among all simulation times using the starting structure as a reference.

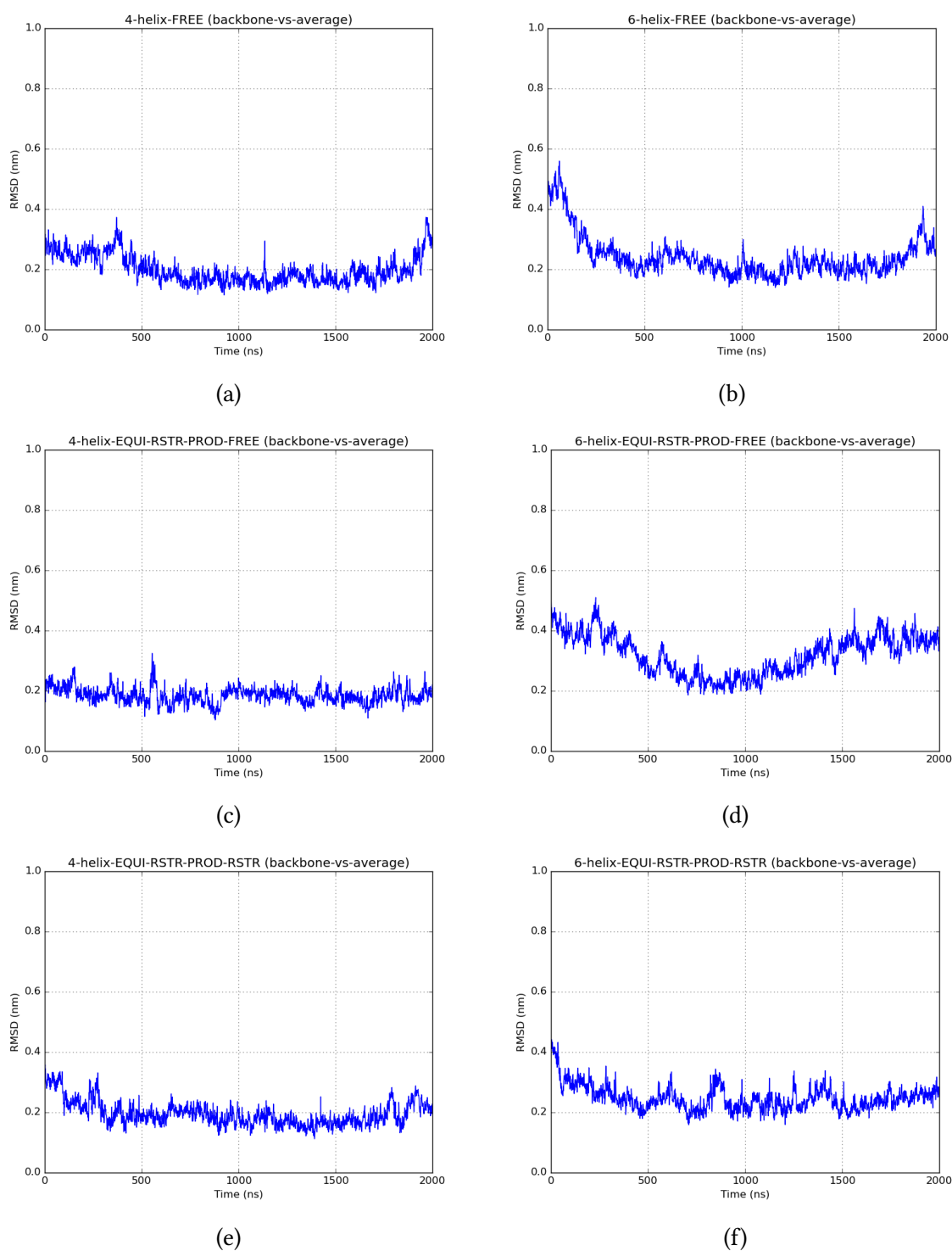


Figure 7.4.: Backbone RMSD analysis of 4-helix and 6-helix complexes with three simulation protocols. Left column figures (a,c,e), shows the backbone RMSD plots of the 4-helix complex among all simulation times using the average structure as a reference. Left column figures (b,d,f), shows the backbone RMSD plots of 6-helix complex among all simulation times using the average structure as a reference.

are located at the ends of the TMDs, and those with the least flexibility are located in the middle of the TMDs. As discussed earlier, three different simulation protocols have been employed for both complexes.

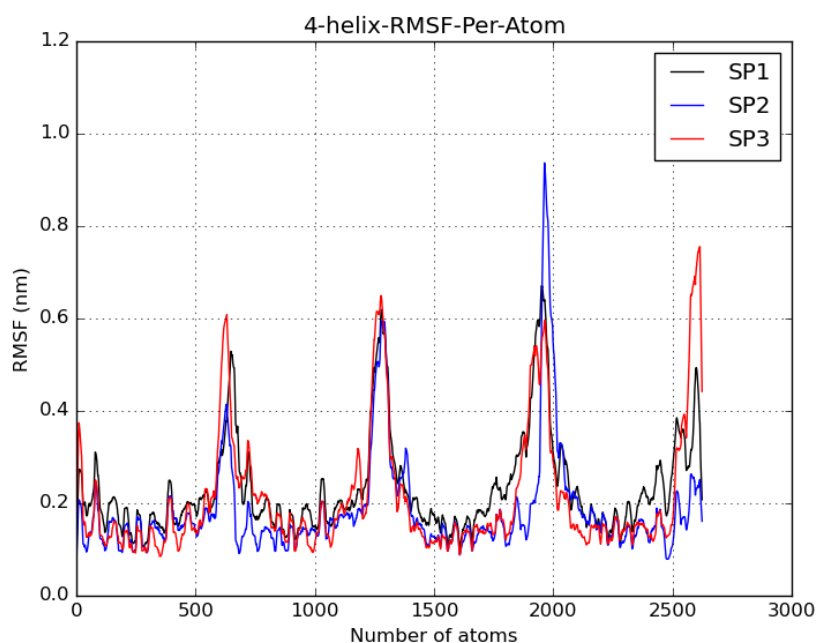


Figure 7.5.: RMSF scan along all atoms in tetrameric complexes. These data originate from the three different simulation protocols (SP1, SP2, and SP3). The tetrameric complex has two E5 (each E5 has 642 atoms) and two PDGFR β (each PDGFR β has 672 atoms).

RMSF profiles were scanned along all four helices from N- to C-terminus of a 4-helix complex, and it was found that all three simulation protocols produced similar results for all four helices. Despite the fact that there are small differences when restraints are applied during equilibration (SP2) and during all restrained simulations (SP3). There is similar behaviour of the E5 protein under all three simulation protocols, though the RMSF value will always be lower in restrained simulations (SP3) than in unrestrained simulations (SP1, SP2). In spite of the similarity between the RMSF for PDGFR β proteins, there is a clear difference between unrestrained simulation (SP1) and restrained simulation (SP2, SP3). In all free simulations (unrestrained, SP1), the black curve in fig. 7.5 always shows higher RMSF values than in all restrained simulations (SP3).

The RMSF of the 6-helix complex was also scanned using all three simulation protocols from N- to C-terminus. All three simulation protocols show similar RMSF profiles for all six helices. A clear difference can also be observed between the RMSFs of free (SP1) and restrained (SP2, SP3) simulations. The restrained simulation (SP3, red curve in fig. 7.6) always exhibit lower fluctuations than the free simulation (unrestrained, SP1). E5 in the

tetramer is sandwiched between the two PDGFR β and shows lower fluctuation, whereas in the case of the hexamer complex, the E5 protein shows a higher RMSF, where E5 is present in the outer region and the core is occupied by PDGFR β .

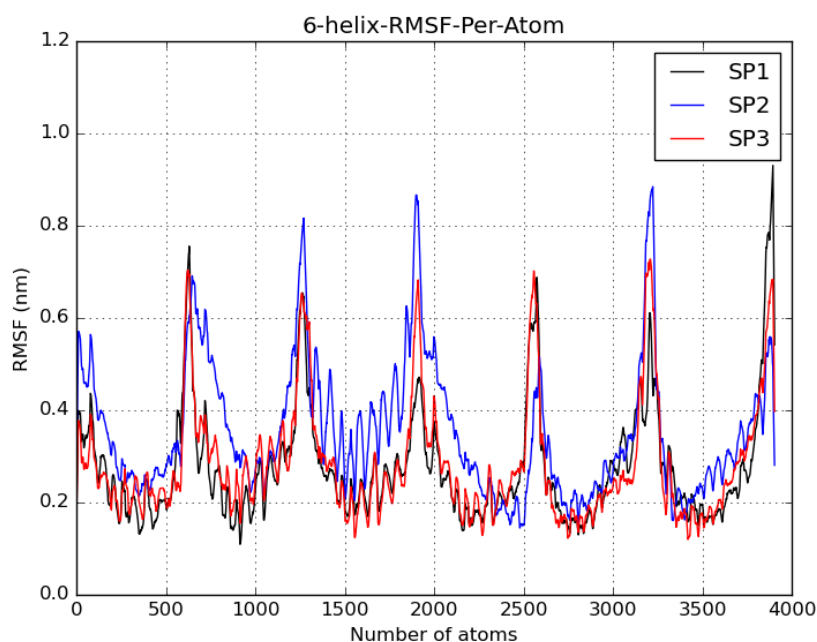


Figure 7.6.: RMSF scan along all atoms in hexameric complexes. These data originate from the three different simulation protocols (SP1, SP2, and SP3). Hexameric complexes consist of four E5 proteins (each E5 has 642 atoms) and two PDGFR β (each PDGFR β has 672 atoms).

The observations above suggest in all the simulations of both, the tetramer and the hexamer, that the restrained simulation (SP3) is more stable and more equilibrated than the unrestrained simulations (SP1, SP2). The results of this study further indicate that Gln17 is important for stabilising the 4-helix or 6-helix complex between E5 and PDGFR β by interacting with Thr513. Apart from the hydrogen bond between Gln15(E5) – Thr513(PDGFR β), the salt bridge between Asp(E5) – Lys499(PDGFR β) is important for the stability of the tetrameric and hexameric complex. Comparing the average structure of the tetramer and hexamer from their respective initial starting structure, it seems that assembled oligomers have restored the proposed structure despite the slightly high RMSD value and high root mean square fluctuation.

7.3.2. Helix Tilt and Rotational Angle

It is essential to have a thorough understanding of the structural properties of membrane proteins in order to understand their mode of action on a molecular level. This includes knowledge of the backbone structure and tilt angles of the transmembrane segments. Moreover, as membrane proteins are embedded in a lipid bilayer, it is important to determine

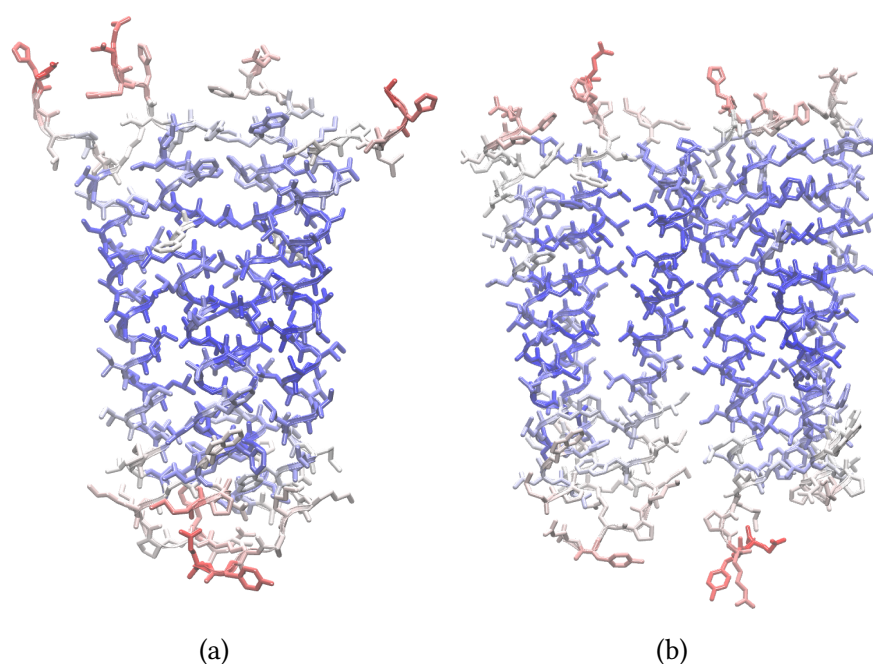


Figure 7.7.: Average structure of (a) 4-helix and (b) 6-helix complexes coloured by RMSF value from the simulation protocol 2, where the restraints were applied during equilibration and released during production simulation. The midpoint of the RMSF value is set at 0.1 and the offset value is set to 0. The colour changes from blue (small fluctuation) \rightarrow white (slightly larger fluctuation) \rightarrow red (large fluctuation).

how and to what extent the lipid environment affects these properties [120]. The protein may adopt a tilted orientation to achieve hydrophobic matching if the hydrophobic portion of a transmembrane segment is longer than the bilayer formed by the surrounding lipids. It is also possible that such a mismatch could result in (local) changes in the backbone structure of peptides, thereby decreasing the effective length of the protein. Mismatch responses of both types could clearly have functional implications as they could affect the proteins' structure and, consequently, their activity. A schematic representation of tilt and roll angle is shown in figure 7.8.

Based on the relative orientation of the individual peptide helices, the structure of the oligomeric complexes was characterised. The first step in this process was to determine the helical axis vectors. The Simulaid software was used to determine the angles between them as well as the angles with respect to the laboratory coordinate frame. Tables 7.2 and 7.3 provide the tilt and azimuthal angles of all helices in both complexes, tetramer and hexamer, respectively. All helices in the tetramer have tilt angles approximately equal to 11–17° (or the complement of 180°) using all three simulation protocols. This further indicates a symmetry in the structure throughout all simulation protocols and maintain a relatively upright orientation in the membrane. The standard deviation in tilt angle measurement throughout the simulation for E5 and PDGFR β using all three simulation

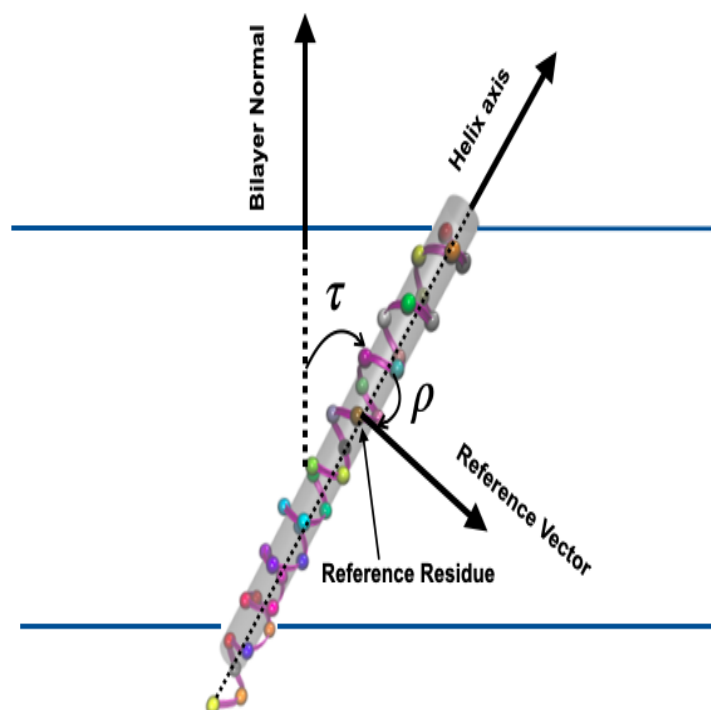


Figure 7.8.: Schematic representation of tilt and azimuthal angle.

protocols is small and in the range of ± 0.5 to ± 3.0 .

Variations in azimuthal angles are significant and do not follow a regular pattern. Azimuthal angles are difficult to define because the helical axes of E5 and PDGFR often deviate from end-to-end regular helical conformations. This effect is more prominent in PDGFR β helices than in E5. In the case of the E5 protein (helices A and B), the azimuthal angle is not uniform across all the simulation protocols, which is in the range of $5^\circ \sim 54^\circ$. Whereas the standard deviation is in the range of $5^\circ \sim 20^\circ$ across three different simulation protocols. The azimuthal angle of E5 proteins in case of all restrained simulation (SP3) behave in a similar fashion with angle $-18.3^\circ \pm 12.3^\circ$ for chain A and $-13.3^\circ \pm 19.2^\circ$ for chain B.

PDGFR β (helix C and D) also behaved non-uniformly with all three simulation protocols. Helix C shows a very high azimuthal angle with all three simulation protocols which range from $150^\circ \sim 230^\circ$ with a high standard deviation ranging from $40^\circ \sim 60^\circ$. Whereas helix D shows a relatively lower azimuthal angle in comparison to helix C. The azimuthal angle of helix D range from $4^\circ \sim 55^\circ$ with a very high standard deviation ranging from $46^\circ \sim 60^\circ$. Overall azimuthal angle distribution is fairly broad for E5 as well as PDGFR β protein.

Considering that there is no uniform azimuthal angle for the 4-helix complex across all three simulation protocols, it appears that all helices with a very high standard deviation are not perfectly helical, and the angle indicates there may be some preferred equilibrated values. As a result, the large standard deviation of more than 40° represents the fact that

Helix	Tilt	Azimuthal
A	11.2 ± 1.1	6.8 ± 10.3
B	17.0 ± 1.6	-53.7 ± 10.0
C	161.2 ± 1.2	227.3 ± 41.3
D	162.1 ± 0.7	4.7 ± 47.8

(a) SP1

Helix	Tilt	Azimuthal
A	13.2 ± 0.7	-26.3 ± 7.7
B	11.8 ± 1.1	-40.4 ± 11.9
C	167.4 ± 1.2	160.2 ± 59.4
D	168.2 ± 2.7	29.9 ± 60.2

(b) SP2

Helix	Tilt	Azimuthal
A	13.6 ± 1.5	-18.3 ± 12.3
B	12.2 ± 1.7	-13.3 ± 19.2
C	165.2 ± 1.4	150.0 ± 57.8
D	168.1 ± 1.3	54.4 ± 56.5

(c) SP3

Table 7.2.: Tilt and azimuthal angle (in degrees, mean \pm std. deviation) from MD simulation with three simulation protocols ((a) SP1, (b) SP2 and (c) SP3) for the tetramer. For the computation of the azimuthal angle, the following residues were utilised: L19 in E5 and L512 in PDGFR β . A and B of the tetramer are comprised of protein helices E5 while C and D are composed of PDGFR β . In this case, the tilt angle is approximately 180°, indicating the opposite orientation of the PDGFR β helix in the membrane (anti-parallel to E5).

rotation visits a variety of values. The high or low value of the azimuthal angle can result due to the bending or formation of a kink in the helix.

The hexamer consists of two pairs of the E5 helix (A, B, C, D) and a pair of PDGFR β (E, F) helices. There is also symmetry in the hexamer, with helices A and D at 9–12° with SP1, 13–15° with SP2, and 13–15° with SP3. A similar symmetry is present in the tilt angle of helices B and C and shows an angle of 9–11° with SP1, 8–11° with SP2, and 8–10° with SP3. The PDGFR β helices E and F also show symmetry in the tilt angle at a complement to 14–15°. The standard deviation in tilt angle measurement throughout the simulation for E5 and PDGFR β using all three simulation protocols is small which is (approximate ~ 1.0). Overall, the hexameric complex shows a similar trend with all three simulation protocols with very low standard deviation, which suggests that all the helices are placed upright in the membrane. There is a high probability that the tilt angles in the tetrameric and hexameric assemblies are too similar to allow comparison with the experiment to distinguish between them.

Unlike the tetrameric complex, the hexameric complex also does not follow a regular pattern when it comes to the azimuthal angle of each helix (four E5 and two PDGFR β). Though there is no similar trend in the 6-helix complex, the standard deviation is higher in the case of PDGFR β in comparison to the four E5 helices. It is also observed that the PDGFR β helix E always shows a similar trend in all three simulation protocols, which

Helix	Tilt	Azimuthal
A	9.5 ± 0.9	-4.9 ± 45.6
B	9.9 ± 0.4	-1.0 ± 18.0
C	10.4 ± 0.4	-70.4 ± 13.5
D	12.2 ± 0.8	16.4 ± 18.1
E	171.2 ± 0.8	-58.8 ± 49.7
F	165.2 ± 0.8	-137.2 ± 39.3

(a) SP1

Helix	Tilt	Azimuthal
A	13.6 ± 0.7	7.5 ± 22.6
B	10.6 ± 0.9	-22.0 ± 29.4
C	8.1 ± 0.7	-45.0 ± 35.7
D	14.9 ± 0.3	13.0 ± 18.7
E	165.9 ± 0.6	-64.2 ± 55.4
F	164.7 ± 0.7	194.2 ± 30.5

(b) SP2

Helix	Tilt	Azimuthal
A	13.9 ± 0.7	-5.2 ± 13.0
B	9.9 ± 0.4	-41.1 ± 13.3
C	8.9 ± 0.7	-19.6 ± 24.7
D	13.7 ± 0.7	23.6 ± 52.8
E	166.4 ± 0.6	-79.1 ± 33.2
F	166.2 ± 0.7	179.6 ± 49.3

(c) SP3

Table 7.3.: Tilt and azimuthal angle (in degrees, mean \pm std. deviation) from MD simulation with three simulation protocols ((a) SP1, (b) SP2 and (c) SP3) for the hexamer. For the computation of the azimuthal angle, the following residues are utilised: L19 in E5 and L512 in PDGFR β . A, B, C, and D of the hexamer are comprised of protein PDGFR β , while C and D are composed of protein E5. The tilt angle is approximately 180° , indicating the opposite orientation of the PDGFR β helix in the membrane (anti-parallel to E5).

is approximately $-58^\circ \sim -80^\circ$ with a standard deviation in the range $-33^\circ \sim -55^\circ$. The standard deviation in case E5 is approximately $13^\circ \sim 52^\circ$, whereas, in the case of PDGFR β , it is approximately $30^\circ \sim 56^\circ$.

There is a significant variation in the helices (four E5 and two PDGFR β) throughout the simulation, suggesting that the helices are not perfectly α -helical. This might happen due to the formation of the kink or bending of the α -helical structure. Which can result in a change in magnitude and direction in the rotation vector of a helix. This implies that the rotation vector is not equilibrated and thus explores more space.

From our three simulation protocols, I have also measured the angle between each pair of helices. The results for the tetramer and the hexamer are shown in tables 7.4 and 7.5, respectively. A similar trend can be seen between each pair of the helices, with all three simulation protocols with small standard deviations. In the tetramer (Table 7.4), helix A and B (E5) cross each other at an angle of approximately $\sim 25^\circ$. Whereas there is a slight difference between the angle between helix C and D (PDGFR β) in free simulation (SP1) and restrained simulation (SP2, SP3). In the free simulation, the angle between the helix C

SP1			
Helix	B	C	D
A	27.3 \pm 1.7	168.1 \pm 1.7	151.0 \pm 1.4
B		150.6 \pm 1.8	172.6 \pm 1.8
C			33.7 \pm 1.3
SP2			
A	24.1 \pm 1.2	176.8 \pm 1.2	157.1 \pm 1.9
B		157.3 \pm 1.7	175.6 \pm 2.0
C			21.2 \pm 1.9
SP3			
A	25.2 \pm 2.2	175.7 \pm 1.5	156.5 \pm 2.1
B		154.5 \pm 2.3	174.8 \pm 2.3
C			23.2 \pm 2.0

Table 7.4.: Angles (in degrees, mean \pm std. deviation) between the individual helices of the tetramer. Data are derived from three different simulation protocols (SP1, SP2, SP3) in the analysis simulation. There is an anti-parallel orientation of the involved helices when values are close to 180°. A and B of the tetramer are comprised of protein helices E5 while C and D are composed of PDGFR β .

and D is $\sim 33.7^\circ$, with a standard deviation of approximately ± 1.3 . However, in restrained simulations (SP2, SP3), the angle between helix C and D is approximately $\sim 23^\circ$, with a standard deviation of approximately ± 2.0 . It is important to note that the remaining two pairs of helix A and C and helix B and C are anti-parallel to one another. As a result, the angle between these helices is complementary to 180°. All simulation protocols, however, follow a similar trend for the angles between A–C and B–C.

The hexamer is composed of four E5 oncoproteins (A, B, C, D) and two PDGFR β (E, F) proteins, where E5 and PDGFR β are oriented anti-parallel to each other. The helix angle between A and B (E5) shows similar values of approximately 20° across all three simulation protocols (Table 7.5), with a standard deviation of $\sim 1.0^\circ$. Similarly, the helix angle between C and D (E5) also shows similar values of approximately 20° across all three simulation protocols, with a standard deviation of $\sim 1.0^\circ$. These two pairs of helices (A–B and C–D) are present on either side of the two PDGFR β helices (E–F). The angle between helix E and F (PDGFR β) shows symmetry in all three simulation protocols, with the lowest in the free simulation (SP1) of 20.8° and approximately 25.0° in the restrained simulations (SP2, SP3). Though, the difference is small, with a small standard deviation in all cases.

There is no significant effect on the helix angle caused by the distance restraints between Asp33-Lys499 in helices A–E and D–F, which were applied during the equilibration in SP2 and maintained throughout SP3. Whatever protocols were used to run the simulations, the angle between helices A–E and D–F remained the same. As a result of restrained SP2 and SP3 simulations, the helix angle between A and E is smaller than that experienced in the unrestrained simulation (SP1). It does not matter which simulation protocol is used,

SP 1					
Helix	B	C	D	E	F
A	19.2 ± 1.1	7.0 ± 1.0	21.5 ± 0.9	173.6 ± 1.1	156.0 ± 1.2
B		17.8 ± 0.8	2.5 ± 1.1	163.6 ± 1.0	174.9 ± 1.1
C			19.9 ± 0.9	178.2 ± 0.7	157.9 ± 0.9
D				161.5 ± 0.7	177.1 ± 1.1
E					20.8 ± 0.9
SP2					
A	23.5 ± 1.1	8.6 ± 1.2	27.5 ± 0.7	168.0 ± 0.9	151.3 ± 1.0
B		18.6 ± 1.1	4.5 ± 0.9	156.0 ± 1.0	174.2 ± 0.9
C			22.9 ± 0.9	173.2 ± 1.1	157.1 ± 0.9
D				151.6 ± 0.6	175.2 ± 0.5
E					27.6 ± 1.1
SP 3					
A	23.8 ± 0.8	8.2 ± 0.9	26.8 ± 0.6	167.9 ± 1.2	152.3 ± 0.7
B		18.7 ± 0.8	3.9 ± 0.8	157.0 ± 0.7	173.8 ± 0.6
C			22.5 ± 1.0	173.9 ± 0.8	158.0 ± 1.1
D				153.4 ± 0.7	175.0 ± 0.9
E					25.2 ± 0.7

Table 7.5.: Angles (in degrees, mean ± std. deviation) between the individual helices of the hexamer. Data are derived from three different simulation protocols (SP1, SP2, SP3) in the analysis simulation. There is an anti-parallel orientation of the involved helices when values are close to 180°. A, B, C and D of the hexamer are comprised of protein PDGFR β while C and D are composed of protein E5.

the helix angle between D–F remains almost the same.

Another pair of distance restraints were also applied between Gln17-Thr513, which are present in helices B–E and F–C. The angle between the helix B and E using SP1 is 163.6° with a small deviation of approximately 1°. Whereas in using SP2 and SP3, the angle between helix B and E is almost the same with a low fluctuation of approximately ~ 1°. Similarly, the angle between helix C and F is quite similar regardless of the simulation protocols, which is approximately around 157°, with a deviation of ~ 1°. So, overall it is all good that tetramer is very similar to all three simulation protocols (SP1, SP2 and SP3), which implies that the proposed model is very good.

7.3.3. Order Parameter (S)

Membrane proteins and peptides are heavily influenced by their surroundings in terms of their structure, dynamics, and function. The complex nature of the cellular membrane has been simplified by creating a number of model membrane systems by simplifying the natural membrane and studying the roles played by various components, such as supported

bilayers, micelles, and liposomes [121]. It is possible to experimentally determine the structural details of these proteins within an environment similar to that of a native lipid by using the solid-state NMR (ssNMR) method [122, 123, 124, 125]. Many solid-state NMR experiments are developed based on solid-state NMR, such as PISEMA, CPMG and CODEX [126, 127, 128].

For several decades, deuterium solid-state NMR has been used to probe sidechain mobility and to determine intermolecular contact interfaces of membrane proteins and peptides labelled with ^2H [129, 130, 131, 132]. It is possible to obtain individual signals for each atom in a protein using NMR, which is well suited to determine the structure of proteins with atomic resolution. When it comes to the determination of the structure of proteins that are not compactly folded, for example, helical membrane proteins, NMR approaches that utilise the properties of aligned samples are of paramount importance. During solid-state NMR experiments, membrane proteins in lipid bilayers are completely aligned since they are immobile on a millisecond time scale.

Deuterium NMR experiments were carried out on static membrane-reconstituted samples. An alignment similar to the one observed in single crystals of peptides can be achieved by aligning proteins between glass plates [126]. It is observed that different residues are labelled in a manner that leads to distinct dipolar splittings. By using the splitting, the order parameter (S) can be calculated, which is related to the local mobility of the sidechain (see fig. 7.9). A decrease in order parameters is generally observed with an increase in mobility but also a deviation from vertically. In these cases, a high-order parameter might indicate the presence of an intermolecular contact site.

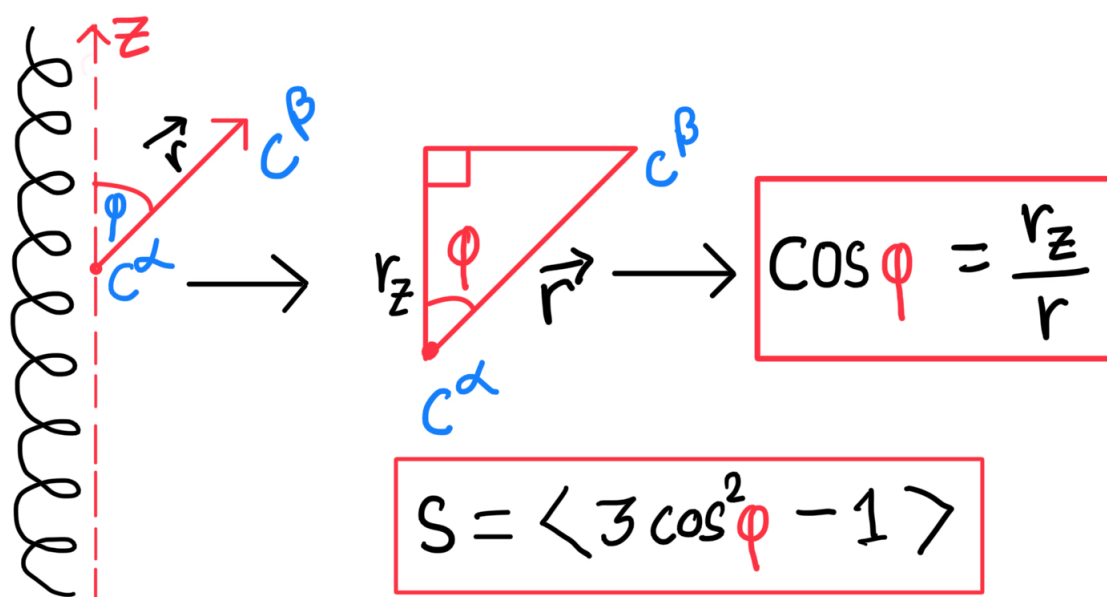


Figure 7.9.: Order parameter measurement from MD simulation.

The order parameter describes the distribution of orientations that are averaged by the motion and serves as a measure of the orientational mobility of the C–D bond. Mobility and segmental orientation affect the order parameter, but it usually decreases with mobility. Additionally, a system with an order parameter of zero can be either unordered or perfectly ordered if it is oriented at 54.7° with respect to the applied magnetic field (B_0). Consequently, changes in motion or orientation directly affect the observed dipolar splitting. The phenomenon of large splitting is often attributed to the reduced motion of a side chain, which may indicate an intermolecular contact site, as well as to orientations that are far from the magic angle.

The NMR observables, such as order parameters measured from experiments, can be measured using MD simulation of proteins embedded in the bilayer. As a result of \mathbf{S} , the helix is described by the average of all its motions. An all-atom membrane simulation provides a relatively straightforward method for calculating the order parameter [133]. The calculation describes the relative orientation of the $C_\alpha - C_\beta/N - H/C - H$ bond vector with respect to the bilayer normal (normally the z-axis in a membrane simulation). It is averaged over all of the residues for the entire sampling time.

$$S_{C_\alpha-C_\beta} = \langle 3 \cos^2 \phi - 1 \rangle / 2 \quad (7.1)$$

In the above equation, 7.1, ϕ is the angle between the desired bond vector ($C_\alpha - C_\beta/N - H/C - H$) and the bilayer normal. An angular bracket represents the average of the molecular and temporal ensembles. In my studies, I have measured \mathbf{S} along the N - H vector as well as the $C_\alpha-C_\beta$ vector for all the helices in tetrameric and hexameric complexes, with all the three simulation protocols.

We performed MD simulations for the 4-helix and the 6-helix complex using three different simulation protocols as described earlier. Further, I have computed the order parameter \mathbf{S} along the N-H bond for each residue of E5 and PDGFR as described above in this section. A helix that is completely immobile has an \mathbf{S} value of 1, which decreases as the helix tilt increase. All the simulation protocols yield very similar N-H order parameters in both 4-helix and 6-helix complexes.

In both of the complexes, E5 helices (A, B in the case of the 4-helix and A, B, C, and D in the case of the 6-helix) show similar values of the N-H order parameter. Amino acid residues at the termini in both complexes show a lower value of order parameter, which suggests that these helices are more dynamic in comparison to the amino acid residues, which are in the middle of the helices. In the case of a 4-helix complex, both of the E5

7. Structural Analysis of E5-PDGFR β Transmembrane Proteins

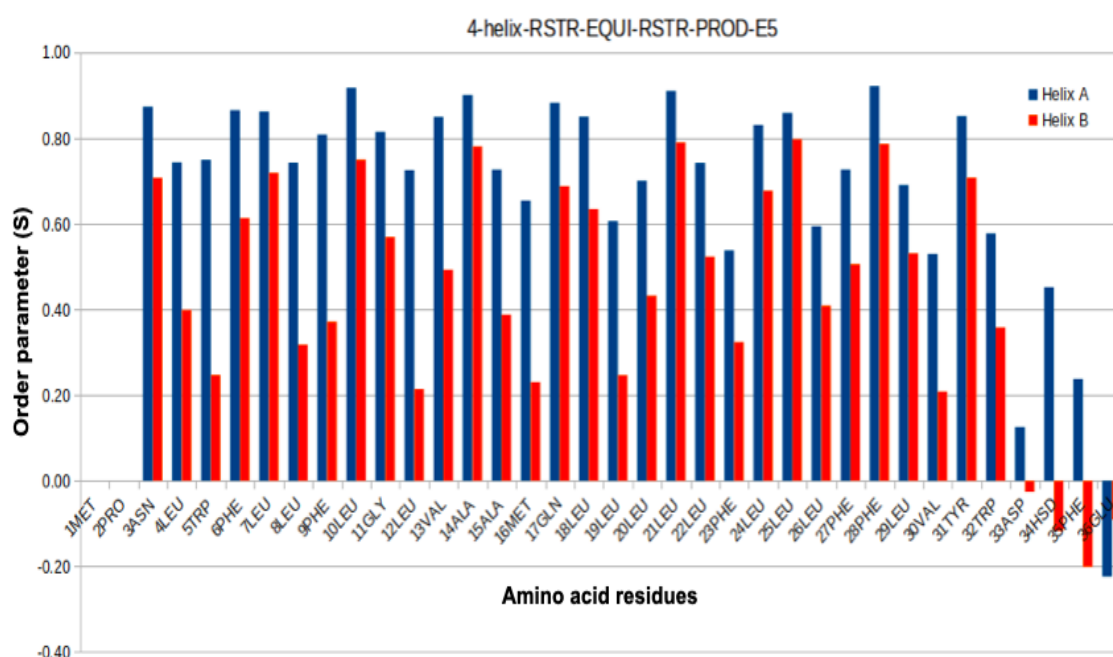


Figure 7.10.: N-H order parameter along a pair of E5 helices of the 4-helix complex from the simulation when restraints are enforced during equilibration and production simulation. Each E5 has 36 residues.

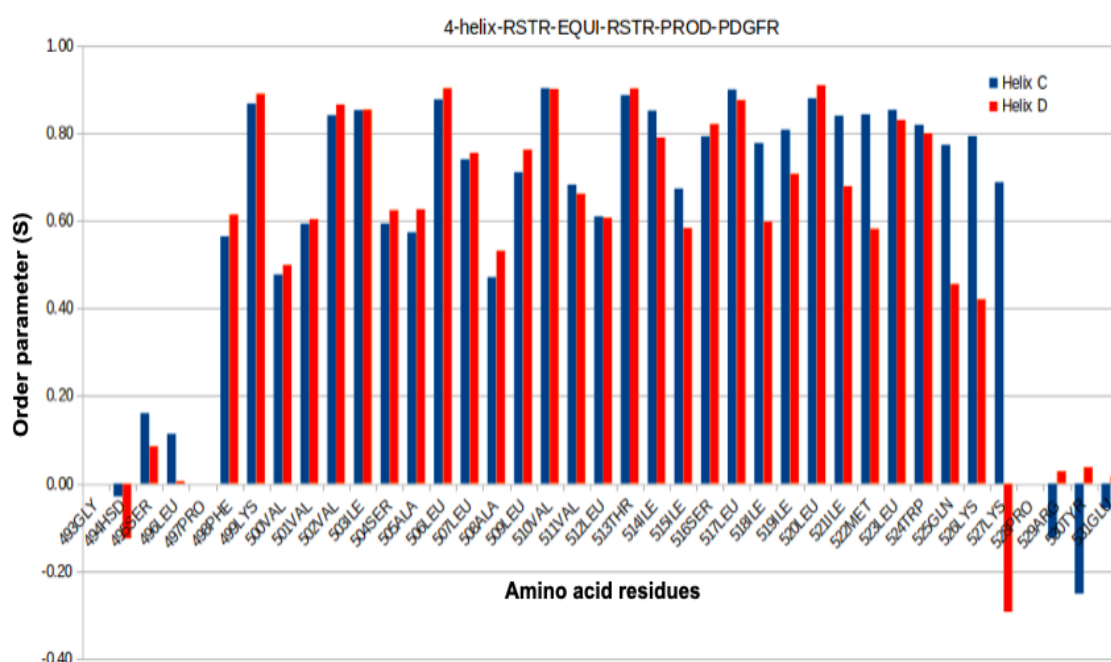


Figure 7.11.: N-H order parameter along two flanking PDGFR β helices of the 4-helix complex from the simulation when restraints are enforced during equilibration and production simulation. Each PDGFR β has 39 residues.

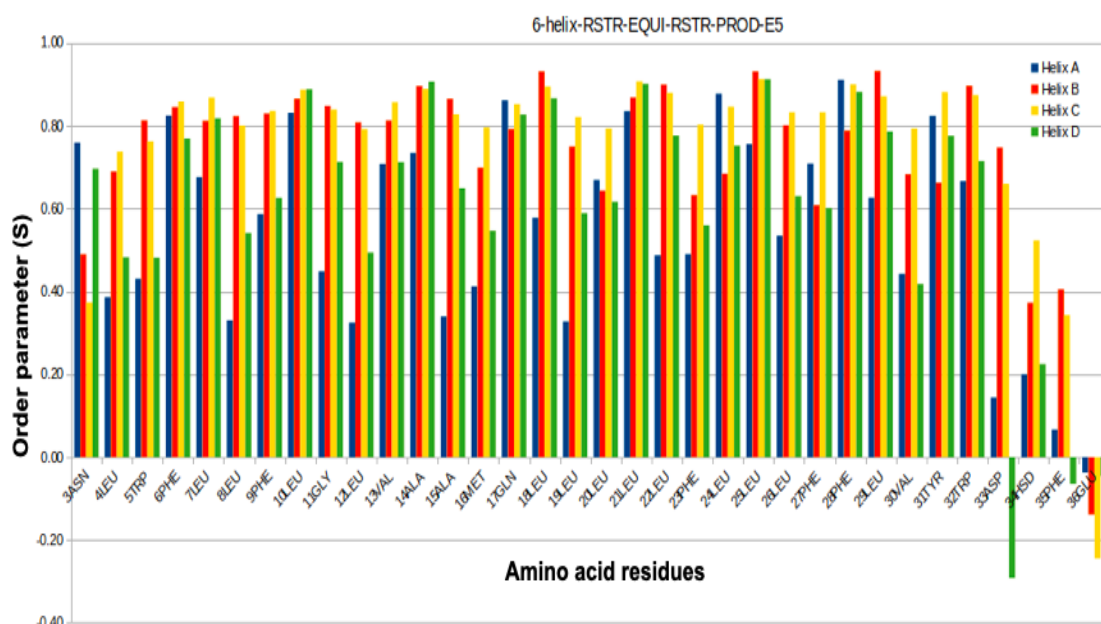


Figure 7.12.: N-H order parameter along two pairs of E5 helices of 6-helix complex from the simulation when restraints are enforced during equilibration and production simulation. Each E5 has 36 residues.

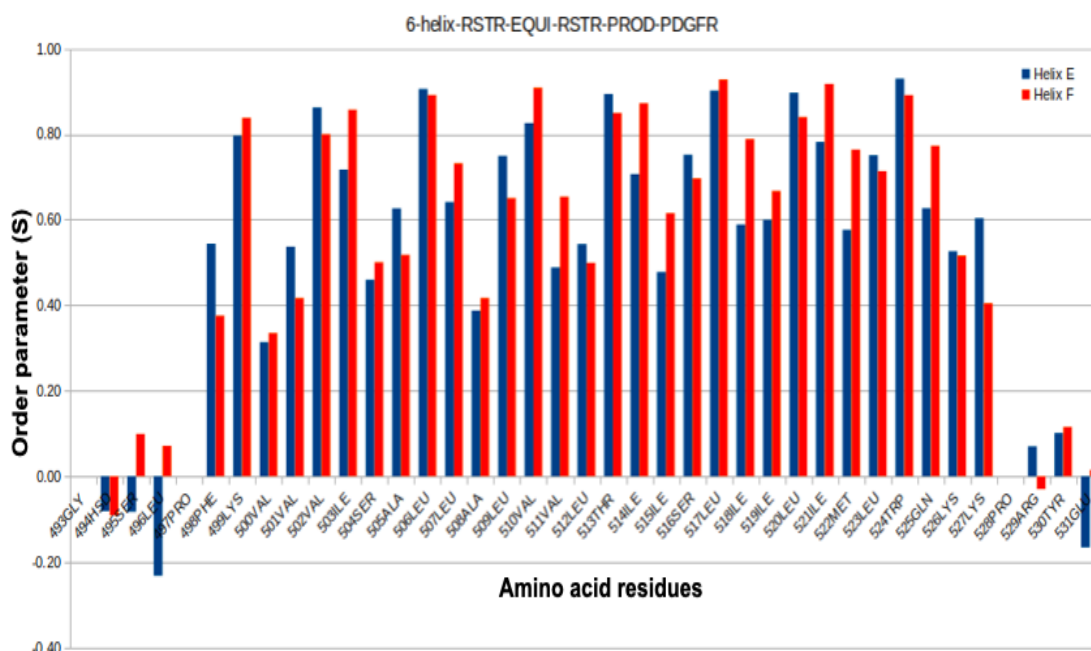


Figure 7.13.: N-H order parameter along two flanking PDGFR β helices of 6-helix complex from the simulation when restraints are enforced during equilibration and production simulation. Each PDGFR β has 39 residues.

helices show very similar values of the order parameter (see fig. 7.10). Whereas in the case of the 6-helix complex, there is little difference between the order parameter of four E5 helices (fig. 7.12). Helices A and D, which are present in the outermost part in the case of the 6-helix complex, always show lower order parameters, which suggests that they are more mobile in comparison to the other two helices (C and D). Helix B and helix C, which are present close to the PDGFR β in the case of a 6-helix complex, show similar values of order parameters and are always close to 1, which further indicates their low mobility.

Unlike, E5 helices, PDGFR β helices (C and D in the case of the 4-helix and E and F in the case of the 6-helix) in both the complexes show similar behaviour (fig. 7.11 and fig. 7.13). Most of the amino acid residues in the middle of the helices show order parameter values close to 1, which suggests that these residues are less mobile. In the case of the 4-helix complex, a similar value of the order parameter is observed, which is close to 1. This indicates that they are less mobile. In the case of the 6-helix complex, the two PDGFR β residues are present in the central core region of the complex and both of the helices show similar order parameters. But in comparison to the PDGFR β in the 4-helix complex, the order parameter values in the case of the 6-helix complex are lower, suggesting that it is more mobile in comparison to the 4-helix complexes.

Further, we can also compare these order parameters (**S**) to root means square fluctuation (RMSF). A higher-order parameter implies less tilt of the helices, which further implies less fluctuation of helices and vice-versa. Residues which are at the terminal show higher lower-order parameters than the residues which are buried in the membrane. This also corresponds to the RMSF, which is higher at the termini of the helices and lowest in the membrane-buried region.

7.3.4. PISEMA Calculation

PISEMA, which stands for Polarisation Inversion Spin Exchange At Magic Angle, relates ^{15}N chemical shift anisotropy and ^1H - ^{15}N dipolar coupling. It enhances separation-local-field (SLF) spectroscopy's resolution and sensitivity. 2D solid-state NMR can reveal structural details of molecules through resonance patterns observed in transmembrane helices and sheets. Polar Index Slant Angle (PISA) wheels are the patterns observed in PISEMA spectra for proteins that are oriented and labelled. In the backbone of a protein, the molecular weight (labelled) can easily be placed in every amide site. A two-dimensional PISEMA experiment measures simultaneously the ^{15}N chemical shift and ^1H - ^{15}N dipolar coupling of each ^{15}N -labelled amide bond. Resonance patterns formed by regular α -helices can be recognised as a two-dimensional wheel (PISA wheel). Fig. 7.14 illustrates the parameters used to calculate chemical shifts and dipolar splittings in PISA wheels.

A completely aligned sample's NMR spectrum exhibits a unique mapping between the resonance frequencies and structure as a result of nuclear spin interactions. Experimental PISEMA spectra show a wheel-like pattern of resonances, which is similar to the pattern observed for helical membrane proteins aligned with each other. As a result of these

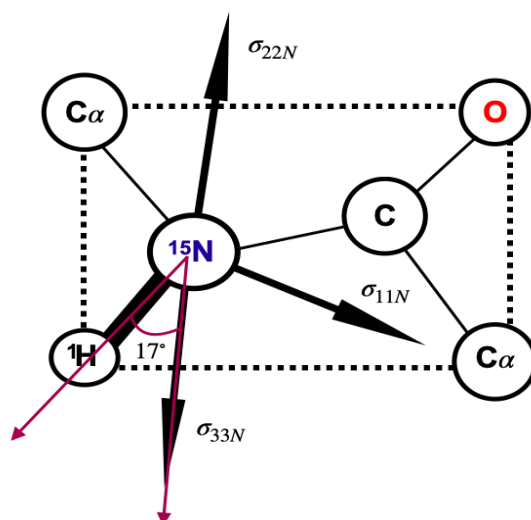


Figure 7.14.: A description of the parameters used to calculate chemical shifts and dipolar splittings for peptides. Illustration of the orientations of the principal elements of the Spin Interaction Tensors for ^{15}N , in a peptide bond. σ_{11} , σ_{22} , σ_{33} are the principal components of the ^{15}N chemical shift anisotropy vector. σ_{22} is perpendicular to the plane of the peptide. ^1H - ^{15}N is dipolar interaction along the N-H bond. σ_{33} is in the plane of the peptide and the angle between the alignments of these two interactions, in the plane of the peptide bond is 17° .

patterns, they are often referred to as PISA wheels (polarity index slant angle) [134, 135]. If the helix axis is parallel to the direction of the applied magnetic field, all amide sites will have identical orientations relative to the applied magnetic field. Hence, all resonances will overlap when the helix axis is parallel to the direction of the applied magnetic field. In the event that the helix is tilted away from the parallel orientation, this breaks the symmetry and introduces variation in the orientation of the amide NH bond vectors in relation to the field of force. There is a difference between the direction of the principal element in the chemical shift tensor of ^{15}N amide versus the direction of the NH bond vector that allows us to resolve a large number of resonances found in helical membrane proteins.

When it comes to a typical helical peptide, the PISEMA spectrum is directly related to the orientation of the helix, specifically its tilt angle. Changing the tilt angle of the peptide causes the PISA wheel to move and the size of the wheel to change. Thus, in the case of a completely static peptide at a smaller tilt angle in the membrane, all the data points are in the top left corner of the PISEMA map. A greater tilt angle causes the signals to migrate to the bottom of the map, and the wheel size will enlarge, which will also improve the spectral resolution of α -helix. A large tilt will result in the wheel reducing again in size as it moves towards the right corner of the map. Fig. 7.15 shows the results of experimental measurements of the PISA wheel of E5 and PDGFR β proteins [101].

The trajectory of the 4-helix and the 6-helix complexes was taken from all molecular dynamics simulations with three different protocols. For each simulation, a PISA wheel

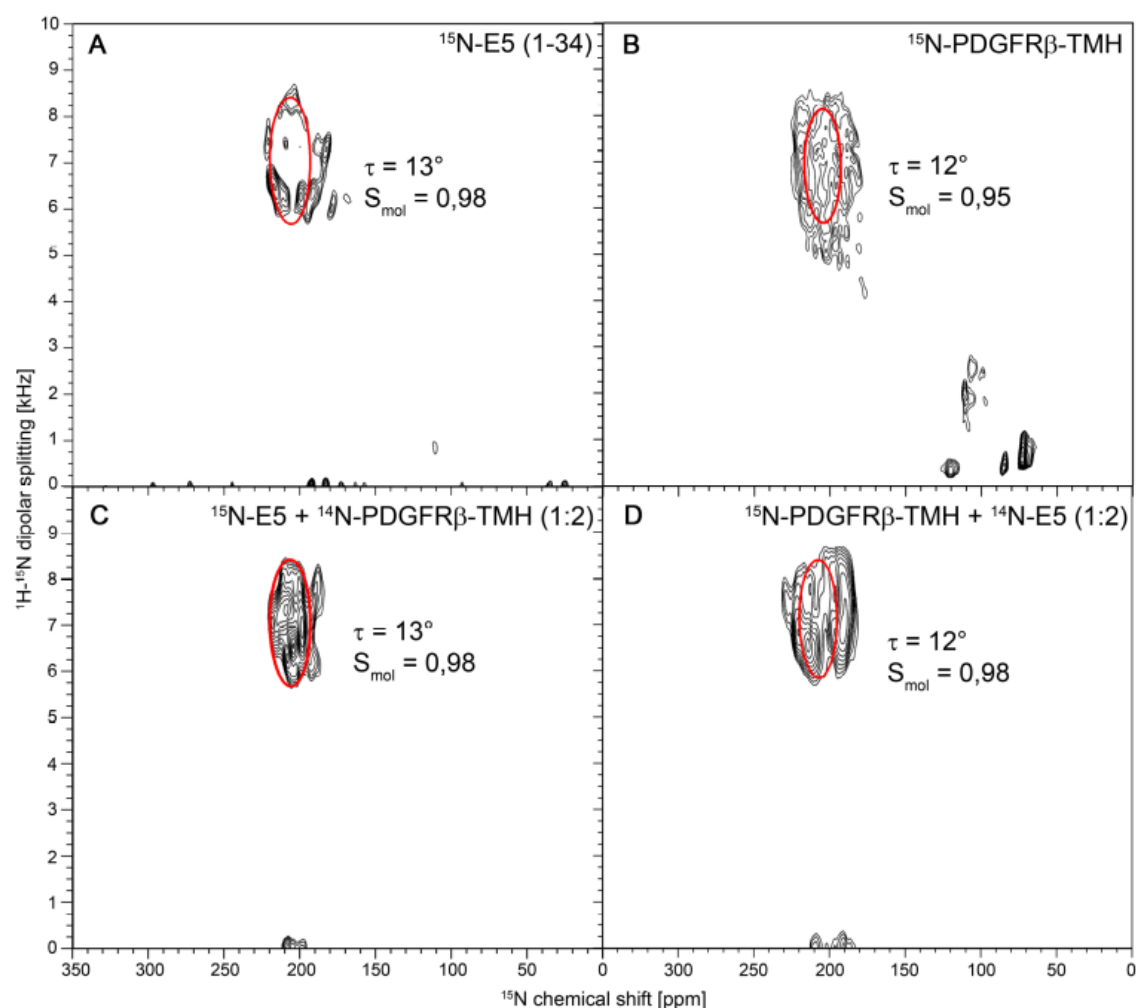


Figure 7.15.: Experimental measurements of the PISA wheel of E5 and PDGFR chain from uniformly ^{15}N -labelled. A) Solid-state 2D-NMR ^1H - ^{15}N SAMMY spectra of E5. B) uniformly ^{15}N -labelled PDGFR β . C) uniformly ^{15}N -labelled E5 with unlabelled PDGFR β D) uniformly ^{15}N -labelled PDGFR β .

was generated for all E5 and PDGFR β pairs in both the 4-helix and the 6-helix complexes. In order to get the PISA wheel from our simulation, firstly we constructed a vector at an angle of 17° in the plane of the peptide, as shown in fig. 7.14, which is σ_{33N} . Further, order parameters along this bond and the N–H bond are calculated. The order parameter of these bonds lies in the range of -0.5 and 1. Thus, in order to compare the PISA wheel from simulation and experiment, it is important to scale the obtained order parameter, so that a direct comparison can be made. Keeping in mind the maximum possible range of the chemical shift anisotropy is 64 - 222 ppm and dipolar coupling theoretically up to 10 kHz (in my experimental collaborator measurement, this value was 8.4 kHz), the order parameter of these bonds was scaled. In fig. 7.16 and fig. 7.17 the computed results are shown.

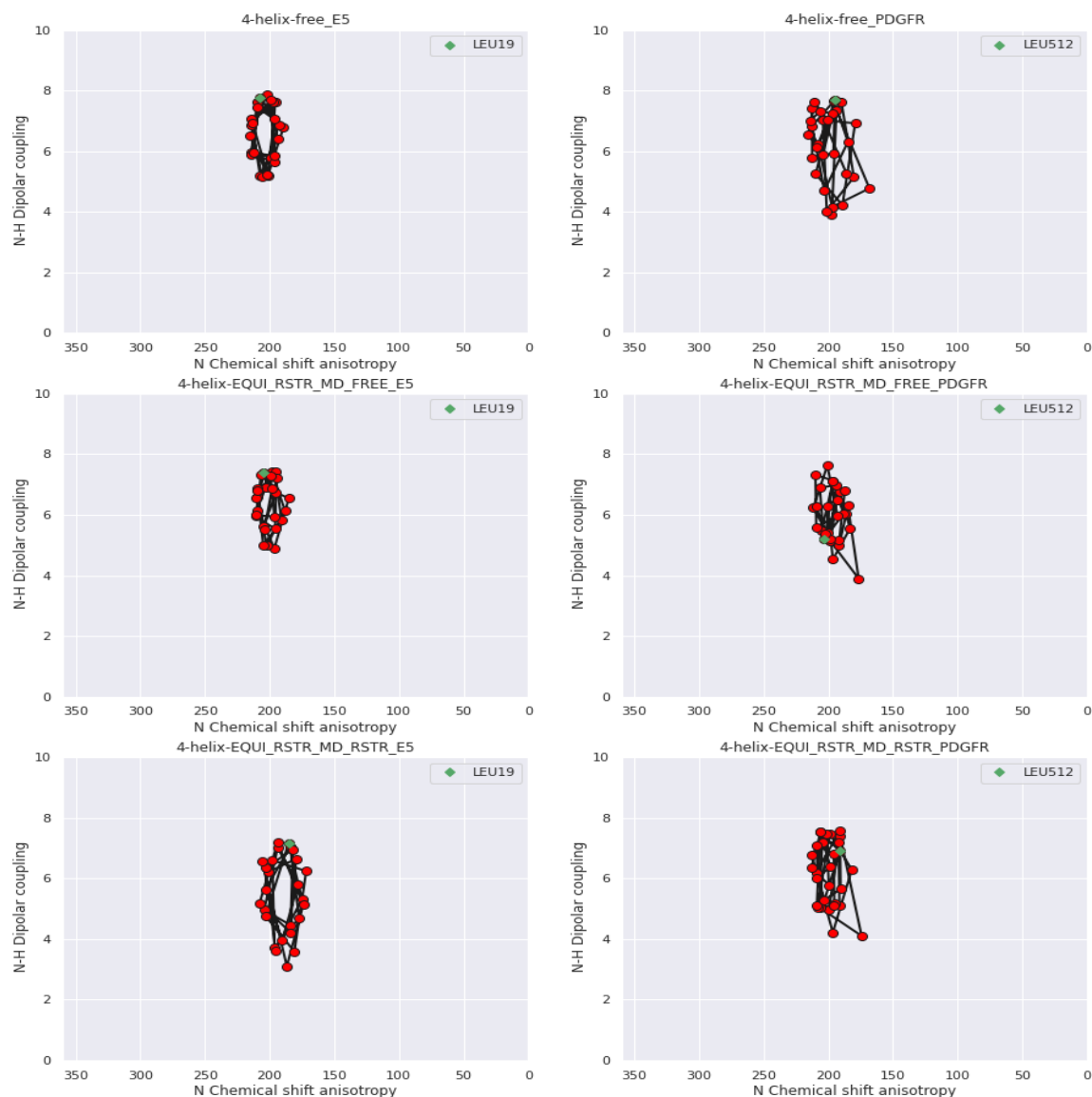


Figure 7.16.: PISA wheel analysis of the 4-helix complex measured from MD simulation with all three simulation protocols. The left column illustrates the PISA wheel of the E5 helices in all the simulations. The right column illustrates the PISA wheel of PDGFR helices in all the simulations. The 4-helix complex has one pair of E5 helices and two flanking PDGFR helices.

There is a perfect resemblance between the PISA wheel obtained from the experiment analysis and the PISA wheel obtained from the MD simulation using all three simulation protocols for both the tetrameric as well as hexameric complexes. In our 4-helix simulation (fig. 7.16), the position of E5 and PDGFR β PISA wheel is centred around 200 ppm in all the simulation protocols. Moreover, in the case of 4-helix simulation using protocol 3, slightly

shifted below in comparison to the simulation of 4-helix using the other two protocols. A larger size of the PISA wheel corresponds to a larger tilt angle, whereas smaller wheels refer to a smaller tilt angle and perfect helices. If the PISA wheels are not perfect, it implies that the helices are not perfect.

PDGFR β chains show very similar positioning of the PISA wheel with all three simulation protocols. Nevertheless, the PISA wheel in all cases is slightly shifted below with all simulation protocols as compared to the experimental measurements. The overall 4-helix complex seems to perfectly match the experimental observation. It can be said that the proposed model by Dr. Tian Li and Dr. Stephan Grage (Karlsruhe Institute of Technology, Karlsruhe) was very good.

The 6-helix model proposed by our collaborator also seems to perfectly match the PISA wheel from the experiments. All E5 chains show smaller wheels (compare fig. 7.17) and more or less at the same position with all three simulation protocols. Nevertheless, the E5 PISA wheel using simulation protocol 2 is slightly shifted below and the wheels are bigger than the other E5 wheels from the other two simulation protocols. PDGFR β in the case of a 6-helix complex shows similar results to all the simulations and the results are quite similar to the experiments. A slight difference can be seen in the PDGFR PISA wheel of the 4-helix complex and the 6-helix complex and i.e., the wheel in 6-helix complexes is slightly expanded.

7.3.5. Distance Measurement

Several studies have demonstrated that oncoprotein E5 from the papillomavirus activates PDGFR β in the membrane through a high-affinity interaction between the transmembrane helices. Nevertheless, the pattern of interaction between E5 and PDGFR β still remains unclear. In experiments, it was demonstrated that hydrogen bonds between Gln17 in E5 and Thr513 in PDGFR β and electrostatic interactions between Asp33 in E5 and Lys499 in PDGFR β stabilised the complex. Due to the highly hydrophobic nature of transmembrane helices, the hydrophobic contacts between the transmembrane helices also contribute to the stabilisation of the E5 / PDGFR β complex, but direct evidence was lacking.

Solid-state NMR measurements were performed by our collaborator in order to measure the distance between the residues. For the measurement of the distance between two different labels, Rotational Echo Double Resonance (REDOR) is one of the most commonly used NMR technologies. REDOR has been used to establish direct evidence of intermolecular distance constraints between E5 and PDGFR β , which was demonstrated by our collaborators Dr. Tian Li and Dr. Stephan Grage (Karlsruhe Institute of Technology).

The interaction between the E5 and the PDGFR β proteins has been investigated through mutational experiments [113]. The study hypothesised that mutations of the amino acids Val502, Ile506, Thr513, Leu517, Leu520, and Trp524 in PDGFR β (see fig. 7.18) result in a reduction in the activity of the E5 protein in cells that express the mutant proteins. Tian and Grage proposed several amino acids, such as Phe28, Phe27, Leu24, Val13, and Leu10,

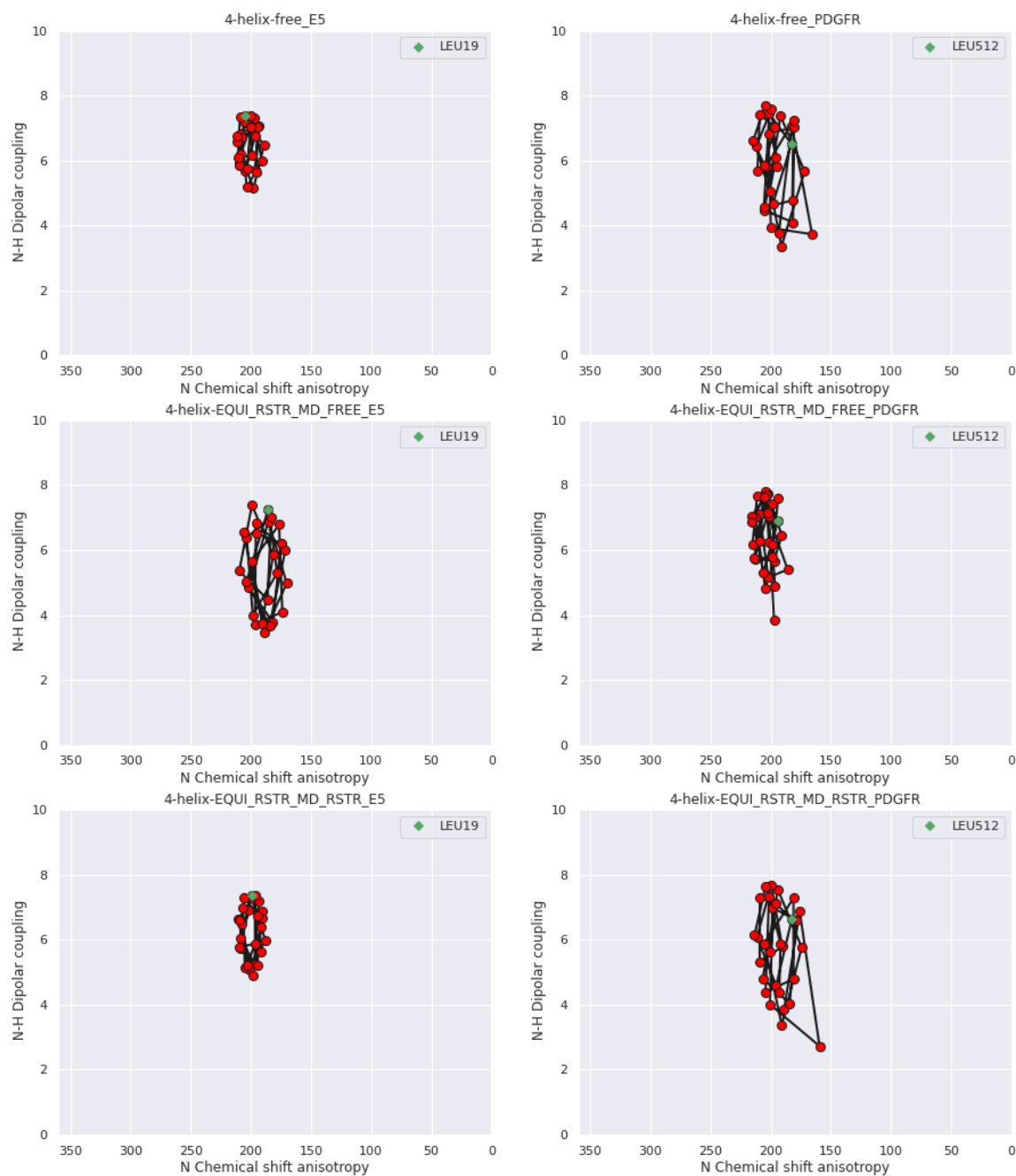


Figure 7.17.: PISA wheel analysis of the 6-helix complex measured from MD simulation with all three simulation protocols. The left column illustrates the PISA wheel of the E5 helices in all the simulations. The right column illustrates the PISA wheel of PDGFR helices in all the simulations. The 6-helix complex has two pairs of E5 helices and a pair of PDGFR helices.

that are involved in the interface in E5 based on the proximity principle and the constraints resulting from the interaction of Thr513 and Glu17. Many potential contact pairs between the PDGFR-TMD and E5 have been proposed and measured using REDOR measurements.

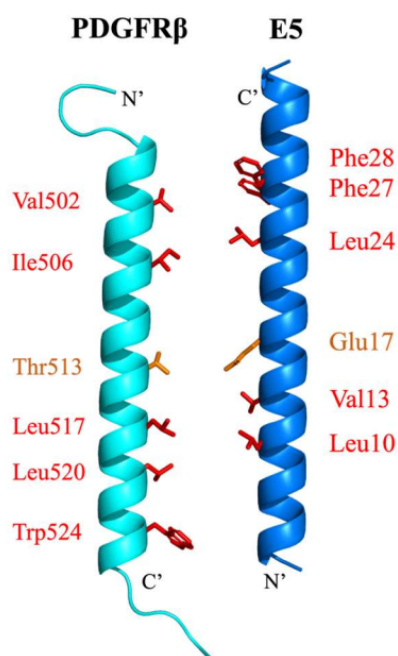


Figure 7.18.: Labelling strategy of E5 and PDGFR β -TMD. Labelled residues that may play a role in the formation of the E5-PDGFR complex are shown as sticks (red). In addition, the residues Thr513 and Glu17 in PDGFR and E5 are identified by the orange colour. [136]

Distance measurements between heteronuclear spins have been widely employed in biological samples using the REDOR method. The technique is based on the magic spinning angle (MAS) technology, which involves spinning a sample at 54.7° to average the chemical shift anisotropy (CSA) and all dipolar couplings.

Based on the membrane orientation of E5 and PDGFR in the complex, which has been proposed by Prof. Ulrich's group [101], a 4-helix (a pair of E5 dimer and two flanking PDGFR β) and a 6-helix (two pair of E5 dimers and a pair of PDGFR β) complex were proposed. The proposed structure was subsequently subjected to MD simulation using three different simulation protocols, which are described in section 8.3. As a final step, the intermolecular distances that were determined experimentally were calculated from the MD simulation, summed, averaged, and compared with those determined experimentally.

Firstly, we examined the distances between two known interacting pairs of residues (Thr513 - Glu17 and Lys499 - Asp33) with simulation protocol 3, where restraints are applied throughout the equilibration and productive simulation, in both the complexes (Table 7.6). Distance measurements with the other two simulation protocols are shown

in the appendix (Tables 11.1 and 11.2). A short distance range (3.0 ~ 5.0 Å) for Thr513 - Gln17 was found between PDGFR β (C) and E5 (B), as well as between PDGFR β (D) and E5 (A). Also, a short distance range (2.5 ~ 5.0 Å) for Lys499 - Asp33 was found between PDGFR β (C) and E5 (A) and PDGFR β (D) and E5 (B). It can be concluded from this, that the constraints were maintained and not disrupted during the simulation. There are two interactions that limit the E5 - PDGFR complex: Thr513 and Lys499 of a PDGFR bind simultaneously to Gln17 on one E5 helix and Asp33 on the other.

In the case of the 6-helix complex (Table 7.7), a short distance range (3.0 ~ 5.0 Å) for Thr513-Gln17 was also found between PDGFR β (E) and E5(B), as well as between PDGFR β (F) and E5 (C). Further, a short distance range (3.0 ~ 4.0 Å) for Lys499-Asp33 was found between PDGFR β (E) and E5 (A) and PDGFR β (F) and E5 (D). These constraints were maintained throughout the simulation, which maintains the stability of the 6-helix complexes due to these two interactions.

A long-distance range of 10 ~ 15 Å has been observed in the complex for Gln17 - Gln17, which is thought to stabilise the E5 dimer in the membrane center. Gln17 appears to adopt a new orientation within the complex, in which it interacts primarily with Thr513. The distance between Gln17 - Gln17 in all other simulation protocols in the 4-helix complex also remains similar, which further suggests the stability of the E5 - PDGFR β complex while forming a hydrogen bond with Thr513.

In addition to these two interactions, I have also measured other distance pairs that have been labelled in the experiment for distance measurements (Table 7.6). In the 4-helix complex, there are two E5 helices (A and B), and two PDGFR β helices (C and D). To determine, whether the arrangement was symmetric during the simulation, we analysed the distance between A-C, A-D, B-C, and B-D. Ideally, the distance between the A-C and B-D should show similar behaviour, whereas A-D and B-C should show similar behaviour. Similarly, in the case of the 6-helix complex (Table 7.7), there are four E5 helices (A, B, C and D) as well as two PDGFR β helices (E, F). In this case, the distance between A-E and D-F should show similar behaviour, whereas B-E and C-F should show similar behaviour. Due to the construction of the model, there are no close contacts between pairs E-C, E-D, F-A, and F-B in the hexamer. Therefore, these data are not shown.

We calculated all the experimentally measured distances from MD simulation, averaged them, and compared them to the initial structure to determine the extent of structural change. Based on these results, it appears that the distances of membrane-buried regions observed in the MD simulation are in agreement with those predicted experimentally. It is apparent that the arrangement of the PDGFR β dimer in relation to the E5 dimer did not change much during the simulation of the 4-helix as well as the 6-helix complex using all restrained simulations. At least in the centre region of the membrane, based on the distances between pairs of Leu24-Leu506, Leu10-Leu517, Val13-Leu517 and Leu10-Leu520 remained the same.

E5	PDGFR β	A - C	A - D	B - C	B - D	Experiment
Phe28 (F)	(C ¹ H ₃) Val502	10.9	14.8	18.3	8.4	7.4 ~ 9.8
Phe28 (F)	(C ² H ₃) Val502	12.7	16.5	20.0	9.8	
Phe28 (F)	(C ¹ H ₃) Leu506	12.9	10.7	8.4	16.8	8.3 ~ 11.3
Phe28 (F)	(C ² H ₃) Leu506	14.8	10.3	10.5	16.4	
Phe27 (F)	(C ¹ H ₃) Leu506	12.3	17.1	16.8	11.2	> 11.3
Phe27 (F)	(C ² H ₃) Leu506	14.1	16.6	18.5	10.6	
Leu24 (D ₃ C ¹)	(C ¹ H ₃) Leu506	11.8	6.8	6.9	11.9	8.0 ~ 10.0
Leu24 (D ₃ C ²)	(C ² H ₃) Leu506	12.8	6.6	8.6	11.7	
Leu24 (D ₃ C ²)	(C ¹ H ₃) Leu506	11.6	7.1	7.1	12.2	
Leu24 (D ₃ C ¹)	(C ² H ₃) Leu506	12.9	6.5	8.2	11.6	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu517	11.4	6.8	6.9	11.9	6.0 ~ 7.0
Leu10 (D ₃ C ²)	(C ² H ₃) Leu517	11.3	6.6	8.6	11.7	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu517	11.2	7.1	7.1	12.2	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu517	11.2	6.5	8.2	11.6	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu520	14.3	4.7	6.7	8.2	-
Leu10 (D ₃ C ²)	(C ² H ₃) Leu520	13.3	6.6	6.1	8.6	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu520	12.8	7.2	6.1	10.2	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu520	14.7	4.3	6.7	6.8	
Ala14 (D ₃ C)	(C ¹ H ₃) Leu517	7.3	7.6	5.3	4.3	
Ala14 (H ₃ C)	(C ² H ₃) Leu517	7.6	7.5	5.1	4.2	
Val13 (D ₃ C ¹)	(C ¹ H ₃) Leu517	14.3	11.9	6.9	10.2	6.0 ~ 7.0
Val13 (D ₃ C ²)	(C ² H ₃) Leu517	13.3	10.1	4.8	10.0	
Val13 (D ₃ C ²)	(C ¹ H ₃) Leu517	12.8	10.2	5.0	10.1	
Val13 (D ₃ C ¹)	(C ² H ₃) Leu517	14.7	11.8	6.4	10.2	
Leu10 (H ₃ C ¹)	(F) Trp524	9.9	16.4	6.0	12.7	7.4 ~ 8.3
Leu10 (H ₃ C ²)	(F) Trp524	8.3	14.4	4.5	10.9	
Gln17 (C)	(N) Thr513	14.2	4.0	3.9	12.6	
Asp33 (O)	(N) Lys499	3.5	20.6	19.2	3.4	
E5 – E5	Gln17(A) – Gln17(B)				12.7	

Table 7.6.: The distance (in Å) between the selected pairs of amino acids side chains of the tetramer from the simulation, in which the proteins were restrained throughout the equilibration and productive simulation (SP3). The tetramer is composed of two E5 helix (A, B) and two PDGFR β (C, D). The distances are compared with the experiment, which is measured by the REDOR method.

However, as distances increase in the membrane-surface region, MD simulations reveal even greater distances than expected, such as those between Phe28-Leu506, Phe27-Leu506, and Leu10-Trp524. The MD simulation for both the 4-helix and 6-helix complexes shows one exception in the pair of Phe28-Val502, whose distance remains the same or decreases slightly. In the MD simulation, distances may appear to be more variable because aromatic amino acids have a larger side chain, so the label may move more, resulting in longer

E5	PDGFR β	A - E	B - E	C - F	D - F	Experiment
Phe28 (F)	(C ¹ H ₃) Val502	10.3	16.5	17.8	9.2	7.4 ~ 9.8
Phe28 (F)	(C ² H ₃) Val502	9.9	15.9	16.0	7.6	
Phe28 (F)	(C ¹ H ₃) Leu506	13.4	16.5	14.5	8.3	8.3 ~ 11.3
Phe28 (F)	(C ² H ₃) Leu506	12.5	15.2	14.4	8.2	
Phe27 (F)	(C ¹ H ₃) Leu506	17.3	14.1	8.5	16.7	> 11.3
Phe27 (F)	(C ² H ₃) Leu506	15.8	13.6	8.4	16.6	
Leu24 (D ₃ C ¹)	(C ¹ H ₃) Leu506	11.6	8.1	5.7	11.5	8.0 ~ 10.0
Leu24 (D ₃ C ²)	(C ² H ₃) Leu506	9.4	6.5	6.2	10.1	
Leu24 (D ₃ C ²)	(C ¹ H ₃) Leu506	10.4	7.3	5.5	11.4	
Leu24 (D ₃ C ¹)	(C ² H ₃) Leu506	10.3	7.3	6.3	10.1	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu517	9.5	6.3	7.9	12.7	6.0 ~ 7.0
Leu10 (D ₃ C ²)	(C ² H ₃) Leu517	9.4	7.2	7.9	12.3	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu517	9.5	7.4	8.7	12.9	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu517	9.4	6.1	7.1	12.2	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu520	13.3	5.7	6.4	12.3	-
Leu10 (D ₃ C ²)	(C ² H ₃) Leu520	13.2	5.8	8.0	12.9	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu520	13.7	6.1	6.9	11.9	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu520	12.9	5.4	7.7	13.6	
Ala14 (D ₃ C)	(C ¹ H ₃) Leu517	4.9	7.8	10.2	6.8	
Ala14 (H ₃ C)	(C ² H ₃) Leu517	10.3	6.3	6.1	12.2	
Val13 (D ₃ C ¹)	(C ¹ H ₃) Leu517	10.4	6.3	6.5	12.2	6.0 ~ 7.0
Val13 (D ₃ C ²)	(C ² H ₃) Leu517	11.4	8.3	7.9	13.3	
Val13 (D ₃ C ²)	(C ¹ H ₃) Leu517	11.5	8.2	8.5	13.3	
Val13 (D ₃ C ¹)	(C ² H ₃) Leu517	10.3	6.3	6.1	12.2	
Leu10 (H ₃ C ¹)	(F) Trp524	12.4	10.6	9.5	13.7	7.4 ~ 8.3
Leu10 (H ₃ C ²)	(F) Trp524	13.1	9.2	9.2	12.8	
Gln17 (C)	(N) Thr513	11.1	4.2	4.2	12.2	
Asp33 (O)	(N) Lys499	3.3	23.2	16.6	3.5	

Table 7.7.: The distance (in Å) between the selected pairs of amino acids side chains of the hexamer from simulation, in which the proteins were restrained throughout the equilibration and productive simulation (SP3). There are no close contacts in the pairs E-C, E-D, F-A, and F-B in the hexamer due to the construction of the model, so these data are not shown here. The distances are compared with the experiment, which is measured by the REDOR method.

distances.

Interestingly, in the simulation, the distance between the two PDGFR β molecules and the E5 dimer is not identical, as the molecules are not symmetrical. Specifically, distances such as Phe28-Val502, Phe28-Leu506, and Leu10-Trp524 in the terminal regions of the helix showed different values, indicating greater fluctuations in amino acids. In spite of this, it appears that the N-terminal region (Val502, Leu506) of PDGFR β is farther away

from the E5 dimer than the C-terminal region (Trp524). Comparing our simulation to experimental data obtained by REDOR, our average distances are mostly in agreement. For the pair of Phe28–Val502, the experimental distance (7.4 ~ 9.8 Å) agrees with that only with chain A–D in the case of the 4-helix complex, and in the case of the 6-helix complex it is in agreement with the pair A–E and D–F.

Among the pairs, Phe28 - Leu506, the distances between A–D and B–C are in good agreement with the experiment as regards the 4-helix complex, but in the case of the 6-helix complex, only D–F is comparable to the experimental range (8.3 ~ 11.3 Å). There was, however, a closer distance found between helix pairs A - C and B - D in the 4-helix complex, rather than between two adjoining helices as in the starting model. Consequently, Leu506 in PDGFR β has a preference for being close to Phe28 of the neighbouring E5, which is different from the initial model but still it is in good agreement with the experimental result. As the experimental proximity (> 11.3 Å) of Phe27 - Leu506 represents all possible locations between the two helices, it represents all possibilities in the 4-helix and the 6-helix complexes.

The distances between the amino acids Leu24 - Leu506 as well as Leu10 - Leu517 seem to be in good agreement with the experimental measurements for both tetrameric and hexameric complexes. But on the other hand, in the case of Val13 - Leu517, only the distances between B - C fit the experimental data for the tetrameric complex, whereas in the case of the hexameric complex the helix pair B–E and C–F are in agreement with the experimental REDOR measurements. Likewise, in the case of Leu10 - Trp524, distances obtained from MD simulation are slightly smaller than the experimentally reported in the case of the tetrameric complex, but in the case of the hexameric complex, the distances between these two pairs are significantly larger than experimental values. These results indicate that the large sidechains of two Trp524 in the complex have completely different orientations. Therefore, two PDGFR β molecules are not arranged in a symmetrical manner with regard to the E5 dimer.

The comparison of our simulation distance measurements with those obtained by REDOR results in three types of situations. In the first situation, the two values obtained from the simulation for the two helix-helix contacts agree with the experimental values, such as the pair of Phe28 - Leu506, Leu24 - Leu506, and Leu10 - Leu517 in 4-helix complexes. Whereas in the case of the 6-helix complexes, both values obtained from the simulation of helix-helix contact Leu24 - Leu506, Leu10 - Leu517, and Val13 - Leu517 agree well with the experimental distance measurement. In the second situation, the distances in half of the complex are consistent with those measured by REDOR, such as the pairs of Phe28 - Val502, Val13 - Leu517 and Trp524 - Leu10 in the case of the 4-helix complex. A third situation occurs when the distances between the helices agree with the experiment, but the agreement is found between different helix pairs, as was initially assumed, only for Phe28 - Leu506, in the case of the tetramer complex.

A key point to remember is that some of the measured average distances derived from the simulation used in the comparison with the experimental results are not in agree-

ment, which might be misleading. Differences in the experimental measured distances and distances measured from simulation can happen due to the average based on several distances, but only one of them agrees with the experiment. Also, it is possible that the inhomogeneity of the sample is responsible for the distance discrepancies between MD simulation and REDOR. MD simulations were conducted on a homogeneous object in which all helices of the E5 and PDGFR contribute to the formation of the heterocomplex. There is probably no such thing in an actual experiment, where proteins are distributed in an inhomogeneous manner in a real sample. This results in fewer protein signals derived from well-formed complexes due to an inhomogeneous distribution consisting of proteins in different states of oligomerisation and aggregation.

8. Water and Ammonia Selectivity Through Aquaporin *AtTIP2;1*

8.1. Introduction

Aquaporins (AQP) are tetrameric transmembrane proteins, which are found in bacteria, plants, and animals. Cell membranes are formed by the phospholipid bilayers, which are highly hydrophobic in nature. Thus, it is nearly impossible for water molecules to permeate through it. Nevertheless, there are many metabolic processes that critically depend on the efficient exchange of water and another solute across the cell membrane. Aquaporins are general categories as the water channel proteins, which form the pore in the membranes of the biological cells [18, 137]. The formation of this pore mainly facilitates the transport of water and some other polar solutes such as carbon dioxide, hydrogen peroxide, glycerol, and ammonia [138] under certain conditions, between cells. Aquaporins enable the passive permeation of water and other solutes across the membrane, which is remarkably efficient for the water balance and other cellular processes. Up to now, more than 350 aquaporins are known; in humans alone, there are 13 aquaporins that have been isolated, which shows specialised permeation features [139]. The structure of the aquaporin *AtTIP2;1* is shown in fig. 8.1 exemplary for homotetrameric aquaporins and aquaammoniaporins.

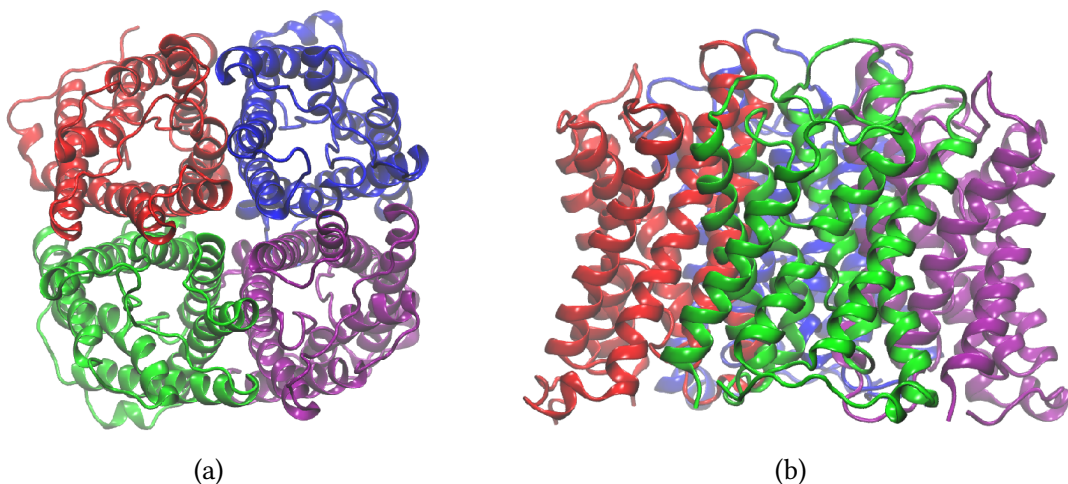


Figure 8.1.: The structure of the *AtTIP2;1*. It is important to note that aquaporins and aquaammoniaporins are homotetramers with four monomeric channels each, as shown in the (a) top view and (b) side view.

Essentially, aquaporins exist as tetramers, with each monomer forming a functional pore [140]. In spite of the fact that the inside of the pores is surprisingly hydrophobic, some sites within the pores are hydrophilic. The top to bottom of the molecule is lined with carbonyl oxygen backbones that accept hydrogen bonds from the water molecules. In the centre of the pore, two highly conserved asparagines, called NPA-motifs, donate hydrogen bonds to water, thereby stabilising the chain's bipolar orientation [141]. The channel structure of *AtTIP2;1* is shown in detail in fig. 8.2.

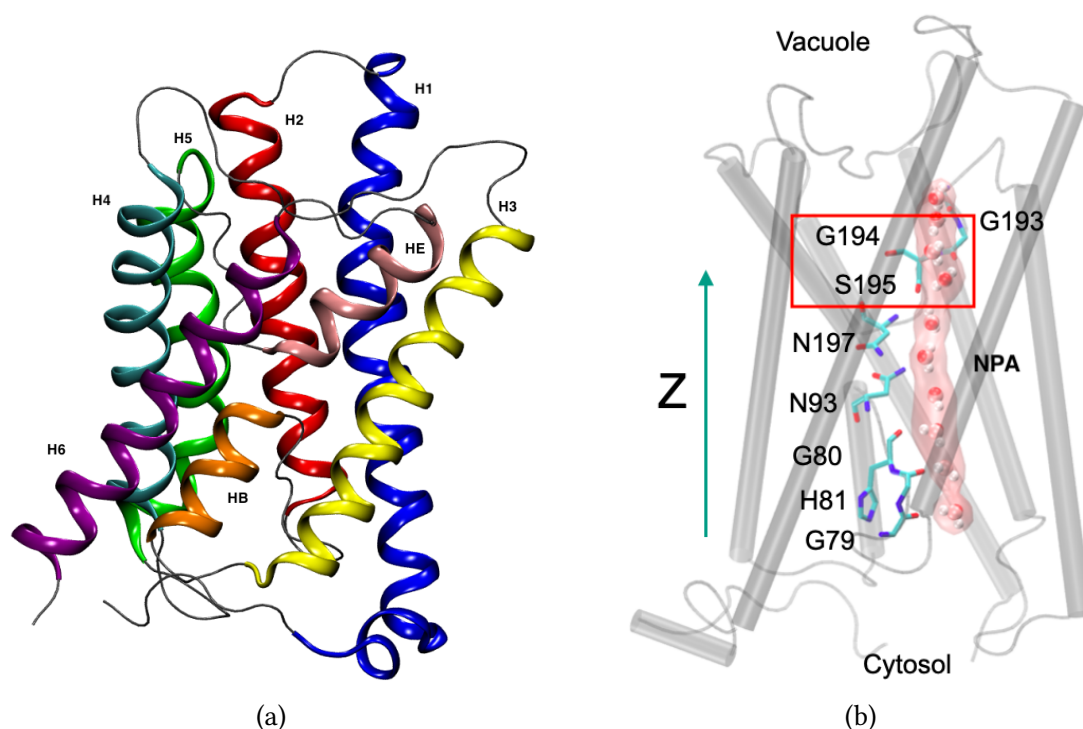


Figure 8.2.: Channel structure of *AtTIP2;1*. Each monomeric channel in (a) consists of six transmembranes (H1 - H6) helices connected by five loops. There are two shorter helices HE and HB, both displaying the AQP-signature motif Asn-Pro-Ala (NPA). In the central NPA region, backbone carbonyl oxygens and two asparagines provide hydrogen bonding sites for the line of water molecules entering the pore. (b) illustrates the direction of the channel axis z, which runs from the cytosolic side of the vacuole (bottom) to its interior (top).

The solute selectivity of the channel is highly dependent on the composition of the selectivity filter (SF), which is located at the entrance of the pore of the aquaporins. The SF is usually the narrowest part of the pore and, therefore, is able to sterically block the passage for large solutes. Human AQP1, which is generally reported to be water-only permeable, has been suggested to rely on the highly conserved SF arginine [142] as a key residue.

There is no doubt that the more similar the solute is to water, the more challenging it becomes to achieve high selectivity without also affecting the water's permeability. Considering that ammonia (NH_3) is a polar molecule with a similar size and dipole moment to water, the selectivity of ammonia is particularly interesting. It is expected that ammonia will fit through narrow constrictions, just as water does, and that it will form hydrogen bonds readily with both water and the polar sites within the pore just as water does.

Tonoplast Intrinsic Proteins (TIPs), which belong to the aquaporin superfamily, facilitate the permeation of water as well as ammonia. TIP2;1 from the wheat plant is an example of an ammonia-permeable aquaporin found in the membrane surrounding the vacuole organelle. It has been established by crystal structures that AQPs are homotetramers, where each monomer contains a functional pore composed of six membrane-spanning helices (helix 1–helix 6), five connecting loops (loop A–loop E), and two shorter helices (helix B and helix E).

It is noteworthy that both short helices display the AQP-signature motif Asn-Pro-Ala (NPA) [143, 144, 145]. These two NPA motifs in the middle of AQPs form a bipartite trans-membrane segment. A variety of small, uncharged polar molecules can permeate through different AQP isoforms. A macro dipole of the short helices is able to effectively exclude protons from the pore in part by concentrating the positive charge at the NPA region [146]. An aromatic/arginine selectivity filter is believed to achieve substrate specificity, which is composed of four residues situated at the non-cytosolic end of the pore. There are two types of residues preserved in most AQPs: Arginine, which contributes to proton exclusion, and histidine, which contributes to water specificity.

8.1.1. Crystal Structure of *At*TIP2;1

In *At*TIP2;1, the pore diameter is smaller at the NPA region than in any other AQP, and remains constant throughout the pore at approximately 3 Å. It is unusual since in most open AQP structures, the aromatic/arginine selectivity filter forms the narrowest portion of the pore. It is important to note that the SF compositions HSD63, ILE185, GLY194, and ARG200 (see fig. 8.3) are similar to those of another ammonia permeable aquaporin, the mammalian AQP8 in this case. There is an unexpected position of the SF arginine R200 in the crystal structure, where, instead of forming hydrogen bonds in the pore, the arginine is occupied by hydrogen bonds with the neighbouring histidine, HSD63. The crystal structure also suggests extending the SF to include a fifth residue, H131, to sterically enforce R200's special position and interact with the water molecules entering the pore.

An experimental study enables the modulation of the unitary water and ammonia permeability of the *Arabidopsis thaliana* aqua-ammonia channel (*At*TIP2;1), by reconstituting it into OSPC containing large unilamellar vesicles. Quantitative activity characterisation utilising fluorescence correlation spectroscopy, stopped-flow methodology, and dynamic light scattering studies suggest modulating its unitary water and ammonia permeability by switching between the liquid-disordered and gel phase states of the bilayer. The perme-

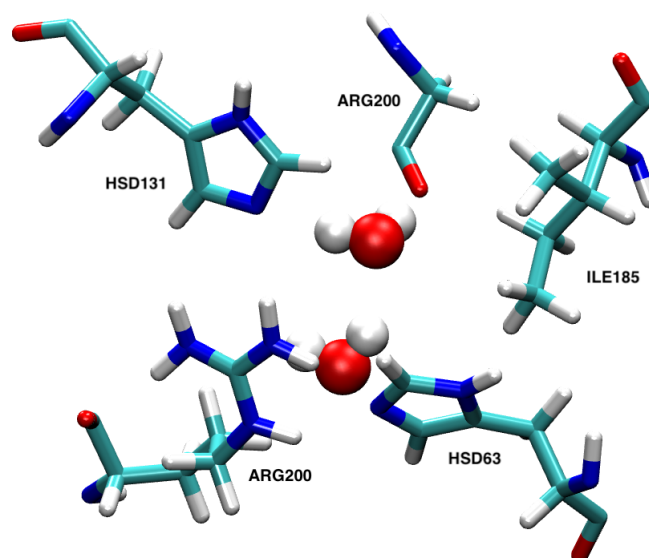


Figure 8.3.: The Selective Filter (SF) composition of TIP2;1 (top view) consists of HSD63, HSD131, ILE185, ARG200. Arginine ARG200 in the SF region is hydrogen bonded with histidine HSD63, instead of offering hydrogen bonds into the pore. The inclusion of additional histidine HSD131 in the SF region sterically enforces the special position of ARG200 and interacts with water entering the pore.

ation of water across the gel phase lipid is substantially lower in comparison to the liquid phase lipid. Also, the permeation of ammonia is significantly lower than in the liquid phase liquid. Furthermore, experimental evidence has shown that *AtTIP2;1* exhibit lower permeation in a gel state compared to a liquid-disordered state in *in vitro* experiments. As a result of this experimental finding, it becomes clear that the lipid matrix plays a very important role in the *in vitro* and *in silico* characterisation of proteins and will inevitably lead to a great deal of future research on the permeability and regulation of AQPs as well as other membrane proteins.

The results of these experiments must be explained at the atomic level in order to gain a thorough understanding of TIP2;1 permeability and selectivity. In the next section, I have explained the state-of-art computer simulation and methods which I have used to reproduce and interpreted the experimental observations.

8.2. Methodology

8.2.1. Force field, simulation setup and equilibration of aquaporin *AtTIP2;1* in liquid phase lipid

Starting from a crystal structure of *AtTIP2;1*, PDB ID 5I32 [147], its tetrameric structure was prepared with PyMol [148], and was embedded in an OSPC lipid bilayer placed in a

rectangular periodic box sized $11.93 \times 11.93 \times 9.62 \text{ nm}^3$ using the CHARMM-GUI membrane builder [114]. The initial arrangement of lipid molecules corresponded to the area of 0.585 nm^2 per lipid. The aquaporin incorporated into the OSPC bilayer was solvated by 22,235 CHARMM-TIP4P molecules, of which 16 were replaced by sodium ions to achieve electroneutrality. The entire system comprising the protein, lipids, water and counterions was described with the CHARMM36 force field [149], and the topology of OSPC molecules was adopted from the readily available SOPC topology in CHARMM-GUI (shown in fig. 8.4). Energy minimisation and equilibration were carried out with Gromacs 5.0, while Gromacs 2020.2 was used for all subsequent production simulations [59, 60, 61].

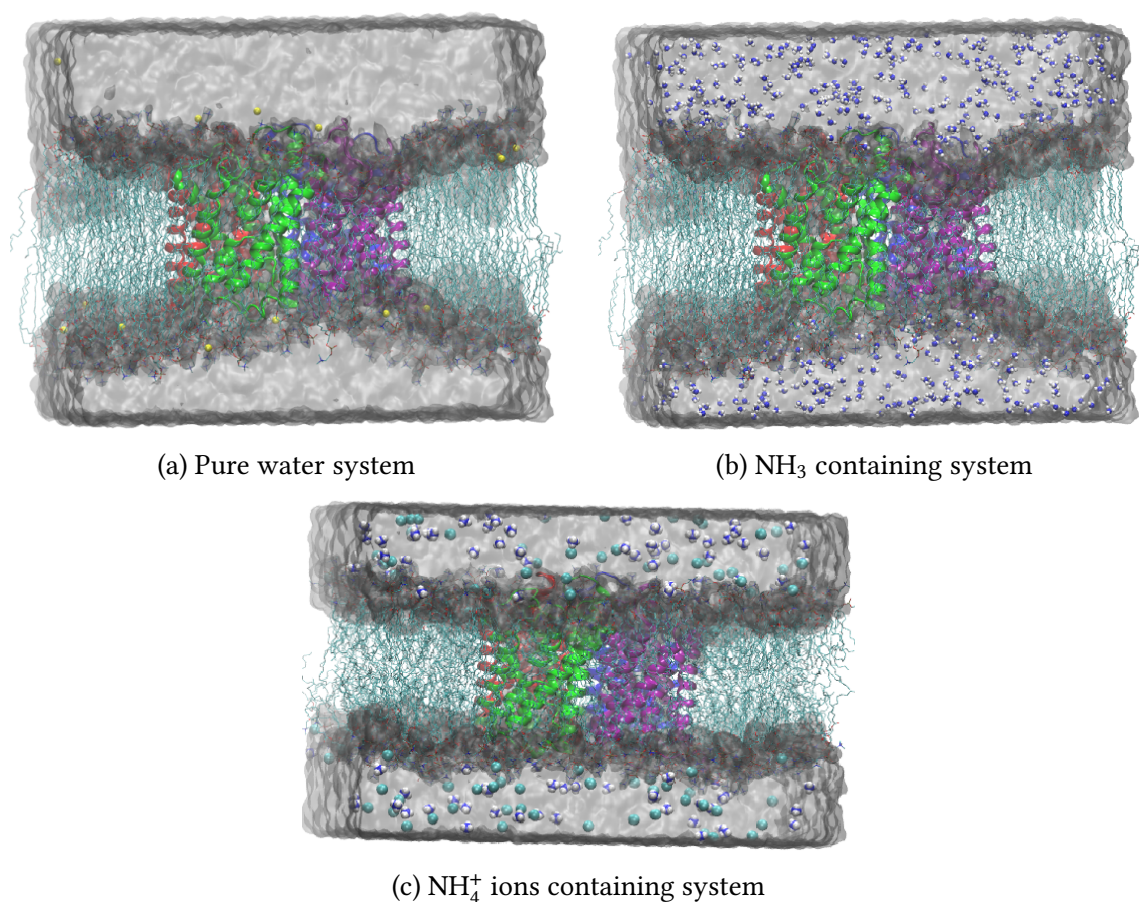


Figure 8.4.: Representative snapshots for the simulation setups in three different situations. In a pure water system (a), the aquaporin-lipid complex is solvated with water and neutralised with chloride ions. In the case of NH_3 containing systems (b), the required number of water is replaced with 1000 NH_3 . In the case of a system containing NH_4^+ ions (c), the required number of H_2O are replaced with NH_4^+ ions, which are further neutralised with an equivalent number of chloride ions.

The system was equilibrated by means of a standard procedure of CHARMM-GUI, as follows: The steepest descents minimisation algorithm was used to minimise energy until

the maximum force dropped below 1000 kJ/(mol·nm). The following position restraints were imposed on the system: 4000 kJ/(mol·nm²) on the backbone atoms, 2000 kJ/(mol·nm²) on the side chain heavy atoms of the peptide, and 1000 kJ/(mol·nm²) on the phosphorus atoms in the lipid head groups. Also, two dihedral restraints of 1000 kJ/(mol·rad²) were applied to each lipid molecule: one on the double bond torsion C38–C39=C310–C311, and one on the improper dihedral in the glycerol moiety, C1–C3–C2–O21. The energy-minimised structure was equilibrated by performing six NVT MD simulations, in which the restraints were relaxed gradually, as summarised in table 8.1.

Equilibration	Time ps	Time Step fs	Backbone kJ/(mol·nm ²)	Side Chain kJ/(mol·nm ²)	Lipid kJ/(mol·nm ²)	Dihedral kJ/(mol·rad ²)
1	125	1	4000	2000	1000	1000
2	125	1	2000	1000	400	400
3	125	1	1000	500	400	200
4	500	2	500	200	200	200
5	500	2	200	50	40	100
6	500	2	0	0	0	0
7	1000	2	0	0	0	0

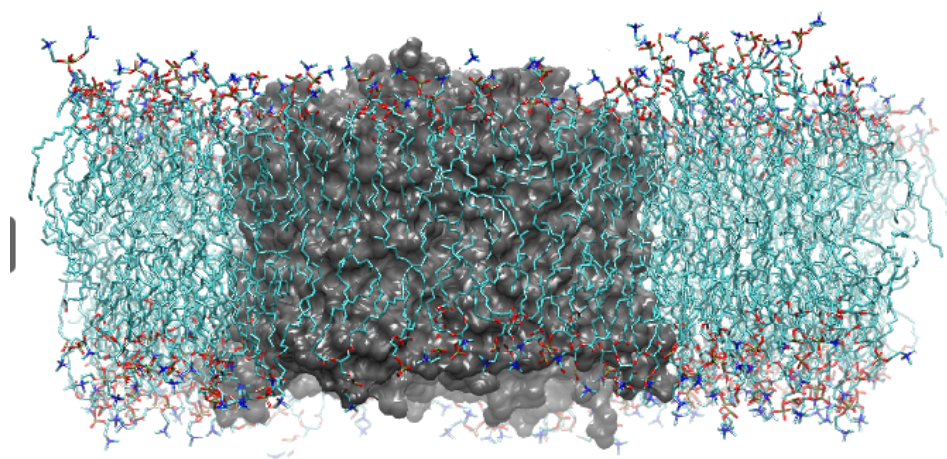
Table 8.1.: In each of the NVTs equilibration steps, lengths of simulations and force constants are provided for each kind of restraint.

The leap frog integrator was used with a time step of 1 fs for the first three equilibration steps, and 2 fs in all of the following simulations. All bonds involving hydrogen atoms were constrained to their respective equilibrium lengths with LINCS [33]. The first two equilibrations were run in an NVT ensemble, with a constant temperature of 303.15 K maintained by a Berendsen thermostat with a coupling constant of 1 ps. The equilibrations 3–6 were run in an NPT ensemble, with a Berendsen thermostat (same as equilibration 1–2) and the constant pressure of 1 bar maintained with two different Berendsen barostats in a semi-isotropic setup, controlling the pressure in the membrane plane and in the membrane normal direction separately, with coupling constants of 5 ps and compressibilities of $4.5 \times 10^{-5} \text{ bar}^{-1}$. For the equilibration step 7, we have used Nosé–Hoover thermostats at a constant temperature of 303.15 K with a coupling constant of 1 ps. The Parrinello–Rahman barostat was also applied with a reference pressure of 5 bar and coupling constant of 5 ps.

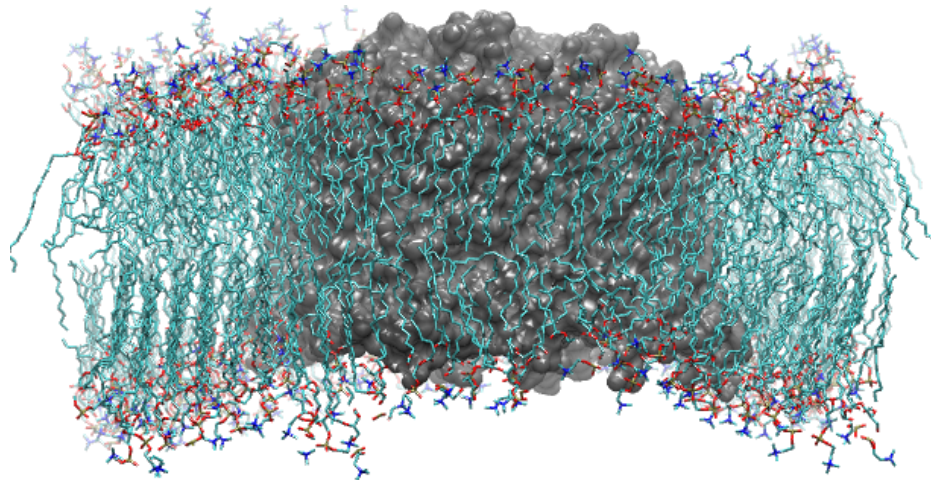
The production simulations were performed with a time step of 2 fs for a total length of 3 μs . A constant temperature of 310 K was maintained with two Nosé–Hoover thermostats [64, 65], coupling protein+lipid and water+ions to separate heat baths with coupling constants of 1 ps. An isotropic semi-isotropic pressure scheme was used to maintain constant pressure in the membrane plane so that the pressure in the membrane’s normal direction was controlled independently of the pressure in the membrane plane. The Parrinello–Rahman barostat [66] was applied with a reference pressure of 1 bar, a coupling constant of 5 ps, and compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. Long-range electrostatic interactions were treated with the particle–mesh Ewald scheme (PME), using a real-space cutoff of

1.2 nm, and Lennard-Jones potentials were switched from 0.8 nm and cut off at 1.2 nm. The coordinates of heavy atoms of the proteins, lipids and solvent were saved every 5 ps.

Trajectories were visualised and images of molecules were rendered with VMD [150]. The analyses were carried out with the GROMACS tools and Python libraries, with Matplotlib being used for data plotting [151].



(a) Aquaporin *AtTIP2;1* in liquid phase lipid



(b) Aquaporin *AtTIP2;1* in gel phase lipid

Figure 8.5.: Snapshot from the MD simulation showing the disordered lipid in liquid phase lipid (a), whereas in the case of gel phase lipid (b) the tails of the lipids are ordered and parallel.

8.2.2. Simulation setup and equilibration of aquaporin *AtTIP2;1* embedded in gel phase lipid

We also performed an unbiased simulation of *AtTIP2;1* in gel phase lipid. The simulation protocol followed that of the simulations involving liquid lipids, except that the reference

temperature was 270 K. However, the structure of the system was prepared using a two-step procedure, which started from the structure of the system following 1 μ s of MD simulation at 310 K, with lipids in the liquid phase. First, in a simulation of 200 ns, the temperature maintained with the Nosé–Hoover thermostats with coupling constants of 0.5 ps, was decreased from 310 K to 270 K, corresponding to a cooling rate of 0.2 K/ns. The desired ordering of the fatty acid chains was facilitated by imposing small “artificial” forces on the chains: on each of the fatty acid chains, the carboxyl carbon atom and the terminal carbon atom were pulled away from each other along the membrane’s normal direction by a force of 1 kJ/(mol·nm). The second phase of the equilibration consisted of an MD simulation of 100 ns performed at the temperature of 270 K with no additional forces or restraints. Structures of both, AtTIP2;1 in liquid phase lipid and gel phase lipid are demonstrated in fig. 8.5.

8.2.3. Simulations of AtTIP2;1 in the presence of ammonia or ammonium

Additionally, AtTIP2;1 embedded in a lipid bilayer, liquid as well as in the gel state, was simulated in the presence of ammonia (NH₃) or ammonium (NH₄⁺) salt. The topologies of NH₃ and NH₄⁺ were taken from the CHARMM small molecule library at CHARMM-GUI. 1000 water molecules in the structure following a previous 1 μ s simulation in a neat aqueous environment were replaced by NH₃ or NH₄⁺ by using GROMACS tools; any NH₃ or NH₄⁺ penetrating into the membrane were removed from the system manually. In the latter case, the corresponding number of water molecules was replaced by chloride counterions to preserve electroneutrality. The resulting molecular systems were energy-minimised as described above, followed by unrestrained 1 ns NPT equilibrations performed at the respective temperatures and a pressure of 1 bar with a time step of 2 fs. The production simulations of 3 μ s were performed with the same settings as described above for the simulations with a neat aqueous environment. The coordinates of heavy atoms of protein, ammonia and water were saved every 5 ps.

8.3. Results

8.3.1. Water passage in liquid and gel phase lipid in pure water system

In a pure water system, we performed 3 μ s simulations in the liquid and gel phases, which were recalculated for 1 μ s. We have simulated liquid phase lipid at 310 K and gel phase lipid at 270 K. In the liquid phase and gel phase lipid, we counted the amount of water passing through each channel. We have also calculated the average time taken by H₂O molecules to pass through any of the channels in liquid and gel phase lipid.

In the case of liquid phase lipid simulation, there are 3946 H₂O passages through all the channels combined. Whereas in the case of gel phase lipid there are 645 H₂O molecules passages across each channel combined. In the case of liquid phase lipid, each of the channels behaves in a similar way and the number of H₂O passages through these channels is \sim 1000. Also, the ratio (Up passages / Down passages) of water passing through each

channel is quite similar, which is ~ 1.2 . Whereas in the case of gel phase lipid, the number of H₂O passages through each channel shows different behaviour. All the channels behave in a similar way and the ratio of the number of H₂O molecules passing up and down is ~ 1.0 . But if we compare the number of water molecule passages through each specific channel, they are quite different, especially channel C. In the case of channel C, there are 42 water molecules passing through it in 1 μ s simulation, whereas in the case of A and D, there are 232 and 245 water molecules passing through it, either way of the channel. Channel C is different due to three major reasons. First, there is a hydrogen bond between HSD131–ARG200, which narrows the main pore and it strictly prohibits the permeation of water in the side pore. Secondly, the formation and breaking of a hydrogen bond between ASN83–MET196 at NPA regions also controls the water passage. Deformation of this hydrogen bond narrows the space and blocks the passage of water in the main channel. Finally, the deformation of the hydrogen bond between GLU24–HSD81 at the cytosolic also narrows the main channel. These three factors contribute to a fewer permeation of water across channel C. Whereas in the case of channel B, there are 135 H₂O molecule passages through each channel, refer to table 8.2.

Channel	Up	Down	Total
A	515	422	937
B	484	482	966
C	542	461	1003
D	579	460	1040
			3946
Average Time: 2.0 ns		pf: 16.4	

(a) Lipid phase lipid at 310 K

Channel	Up	Down	Total
A	112	120	232
B	68	66	135
C	20	22	42
D	127	118	245
			645
Average Time: 10.3 ns		pf: 3.3	

(b) Gel phase lipid at 270 K

Channel	Up	Down	Total
A	292	315	607
B	179	174	352
C	51	58	110
D	333	308	607
			1711

(c) Gel phase corrected to 310 K

Table 8.2.: The passage of water across the aquaporin *AtTIP2;1* in liquid (a) and gel (b) phase lipids in pure water in all channels of the system. The amount of water passing from the cytosol to the vacuole is referred to as up and the number of water going from the vacuole to the cytosol is referred to as down. (c) corrected the number of water passages across all channels in the gel phase lipid at 310 K.

We have measured the average time taken by H₂O molecules to pass through liquid and gel phase lipids. The time taken by each passage is calculated by noting the time of entering into the channels and the time of exit from the channel. The average time taken by H₂O passage is measured by the sum of time taken by each passage divided by

the number of passages. The average time taken by H₂O passage in liquid phase lipid is 2.0 ns, whereas in the case of gel phase lipid it takes 10.3 ns to pass through the aquaporin channels. The time taken for H₂O passage in the case of gel phase lipid is more than five times higher than the amount of time taken to pass through in the case of liquid phase lipid.

It is not fair to compare the number of H₂O passages caused by liquid phase lipids with those caused by gel phase lipids since the liquid phase lipid was studied at 310 K whereas the gel phase lipid was studied at 270 K. Thus, in a hypothetical situation, we calculated the number of H₂O molecules passing through the gel phase at 310 K using the temperature-dependent Arrhenius formula. The activation energy (E_A) for the H₂O passage is taken from the literature as 4 kcal/mol [147]. As a consequence, we can determine the number of passages through each channel by multiplying the H₂O passage by a factor of 2.6. As a result of incorporating corrections in gel phase H₂O passage, there are 1711 H₂O molecules that passed through all channels, which is still less than half of the water transport in liquid phase lipids.

Furthermore, the osmotic permeability (pf) of water is also measured in all systems of both phases of lipid. The osmotic permeability has a direct relationship with the "hopping rate" of the water chain in a single-file model [152]. The pf of water permeation is $16.4 \times 10^{-14} \text{ cm}^3 \text{ s}^{-1}$ and $3.3 \times 10^{-14} \text{ cm}^3 \text{ s}^{-1}$ in liquid phase and gel phase lipid, respectively.

8.3.2. Ammonia passage in liquid and gel phase lipid in the system containing 1000 NH₃ molecules

Further, we have performed 3 μs simulations in the system containing 1000 NH₃ molecules in the system in liquid as well as gel phase lipid. The number of passages is further averaged over 1 μs simulation. Our first step was to count how many molecules of ammonia passed through each channel in both phases of the lipids. In total, there are 59 NH₃ passages in the liquid phase lipid simulation, whereas in the case of gel phase lipid simulation there are only 6 NH₃ passages. Gel phase lipids have a significantly lower quantity of passages of NH₃ compared to liquid phase lipids in terms of the number of passages of NH₃ which pass through. As far as the behaviour of the channels in liquid lipids is concerned, they behave in a similar manner to those in solid lipids. In liquid phase lipid, 11 and 14 NH₃ molecules pass through channels A and B, respectively, whereas in the case of gel phase lipid there is only one NH₃ passing through these two channels 8.3. It is interesting to note that 19 NH₃ passes through channel C in liquid phase lipid, but none is transferred in gel phase lipid. There are 15 molecules of NH₃ that pass through channel C in liquid phase lipid, of which 7 molecules go up (from the cytosol to the vacuole) and 8 molecules go down (from the vacuole to the cytosol). The gel phase lipid has four passages in channel D, three of which are upward and one downward. In addition, it is interesting to note that NH₃ does not pass upward (cytosol to vacuole) in channels A, B, or C in gel phase lipid. But in the case of channel D, there are three upward passages.

Channel	Up	Down	Total
A	5	6	11
B	8	6	14
C	11	8	19
D	7	8	15
			59
Average Time: 3.5 ns		pf: -	

(a) Lipid phase lipid at 310 K

Channel	Up	Down	Total
A	0	1	1
B	0	1	1
C	0	0	0
D	3	1	4
			6
Average Time: 28.6 ns		pf: -	

(b) Gel phase lipid at 270 K

Channel	Up	Down	Total
A	1	2	3
B	1	3	3
C	0	1	1
D	8	2	10
			17

(c) Gel phase corrected to 310 K

Table 8.3.: The passage of NH_3 across the aquaporin *AtTIP2;1*, each channel in liquid (a) and gel (b) phase lipids in systems where 1000 ammonia molecules are present in the system. The amount of water passing from the cytosol to the vacuole is referred to as up and the number of water going from the Vacuole to the cytosol is referred to as down. (c) corrected number of water passages across all channels in the gel phase lipid at 310 K.

The average time taken in both phases of the lipid to pass these NH_3 passages has also been measured in both phases. In the case of liquid phase lipids, it has been observed that NH_3 molecules take on an average of 3.5 ns to pass through the aquaporin channels. In contrast, the average time taken by NH_3 to pass through channels in the gel phase is even higher and can reach a value of 28.6 ns. The NH_3 molecules are significantly blocked by the aquaporins in the gel phase. Also, even if it goes into the aquaporin channel, it takes longer to pass through it.

As discussed previously, simulations are performed at 270 K in the gel phase and at 310 K in the liquid phase. Thus we calculated the hypothetical number of NH_3 molecules passing through the gel phase at 310 K using the Arrhenius formula as described before in the previous section. After the correction, there are 17 NH_3 passages through the gel phase systems at 310 K. According to table 8.3, the ratio of NH_3 passage through gel phase lipid at 310 K to NH_3 passage through liquid phase lipid at 310 K is ≈ 0.29 .

8.3.3. Water passage in liquid and gel phase lipid in the system containing 1000 NH_3 molecules

I have also counted the number of H_2O molecule passages through aquaporin channels in both phases of the lipid in systems containing 1000 NH_3 molecules as seen in table 8.4.

8. Water and Ammonia Selectivity Through Aquaporin AtTIP2;1

For liquid phase lipids containing 1000 ammonia, there are 3467 H₂O molecules that cross aquaporin channels, which is slightly less than the 3946 H₂O molecules in the pure water system. Additionally, in the case of gel phase lipid containing 1000 ammonia, there are 491 molecules of water that cross the channel. The ratio of H₂O passage (Up/Down) in all the channels in liquid phase lipid is in the range of 0.99 ~ 1.03. In liquid phase lipid systems that contain NH₃ contents, there is always less H₂O passing through each channel than the corresponding channels in pure water systems in which NH₃ is not present. The difference is more pronounced in the case of channels A, B and C, but in the case of channel D, the difference is small.

Channel	Up	Down	Total
A	397	400	797
B	422	397	819
C	424	402	826
D	522	503	1025
			3467
Average Time: 2.3 ns		pf: 13.9	

(a) Lipid phase lipid at 310 K

Channel	Up	Down	Total
A	42	48	89
B	50	70	120
C	26	23	49
D	115	118	233
			491
Average Time: 12.6 ns		pf: 2.4	

(b) Gel phase lipid at 270 K

Channel	Up	Down	Total
A	109	125	234
B	130	183	313
C	69	59	128
D	301	308	609
			1284

(c) Gel phase corrected to 310 K

Table 8.4.: The passage of water across the aquaporin AtTIP2;1, each channel in liquid (a) and gel (b) phase lipids in systems where 1000 NH₃ molecules are present in the system. The amount of water passing from the cytosol to the vacuole is referred to as up and the number of water going from the vacuole to the cytosol is referred to as down. (c) corrected number of water passages across all channels in the gel phase lipid at 310 K.

In the same way, in the case of gel phase lipid containing 1000 NH₃, there is a fewer H₂O passage in comparison to the pure water system in the gel phase. Each channel behaves differently in the case of gel phase lipids. Channel D shows the highest permeation of H₂O, whereas channel C shows the lowest permeation of H₂O in gel phase ammonia systems. There are a few differences between the H₂O passage through the pure water system and NH₃ containing systems. In pure water systems, there are 232 H₂O molecules passing through channel A, whereas in the case of ammonia-containing systems there are 89 water molecules that pass through it. The remaining channels B, C and D behave in the same manner in the pure water system as well as the ammonia-containing system in the gel phase. The ratio of H₂O passage (Up/Down) in all the channels in gel phase lipid is in the

range of 0.70 ~ 1.15.

I have also measured the average time taken by H₂O molecules to pass through these channels. In the case of liquid phase lipid containing NH₃, the average time taken by H₂O is 2.3 ns, which is slightly more than the time taken in the case of the pure water system which is 2.0 ns. Also, in the case of gel phase lipid containing NH₃, the average time taken by H₂O is 10.3 ns, which is also less than the time taken in the case of the pure water system which is 12.6 ns.

Consequently, I have also calculated the corrected H₂O permeation in the gel phase at 310 K using the Arrhenius formula. Thus, in corrected gel phase permeation, 1284 H₂O molecules passed through all the channels combined. The overall permeation of H₂O in the corrected gel phase ammonia system is almost three times less than the H₂O permeation in the liquid phase ammonia system at 310 K.

8.3.4. Water passage in liquid and gel phase lipid in the system containing 1000 NH₄⁺ molecules

In addition to that, we have also performed 3 μ s of calculation for the system that contains 1000 NH₄⁺ ions in liquid and gel phase lipids. As reported before, charged ions cannot pass through aquaporin channels [153, 154, 155, 156, 157]. During a simulation of 3 μ s it is not surprising that we find no NH₄⁺ ions transfer in both phases of lipid. It should be noted that the presence of NH₄⁺ ions in the systems severely affects the H₂O permeation in liquid phase lipid systems, but it does not affect the H₂O permeation in gel phase lipid systems. As a result of liquid phase lipid simulation containing NH₄⁺ (shown in table 8.5), we find there are roughly 1987 molecules of H₂O passing through all of the channels combined. This is nearly half the amount of water molecules flowing through pure water simulation which is based on liquid phase lipids. Whereas in the gel phase lipid containing NH₄⁺ ions, there are 544 molecules H₂O which passed across all the channels. This is quite similar to the number of H₂O (645 molecules) which passed through the pure water system in gel phase lipid.

Channels A and D show the lower number of passages of H₂O for liquid phase lipid containing NH₄⁺, which is 258 and 187, respectively, compared with the number of passages for the channels B and C, which is 721 and 820, respectively. In channels A, B, and C, the ratio (Up/Down) of H₂O permeation is approximately ~1 across each of the three channels. However, the ratio (Up/Down) of the passage of H₂O by channel D is approximately 1.56, which is much higher than that of channels A, B, and C. Additionally, it is worth mentioning that the average time taken for the water to move across the channels is 3.8 ns on average. This is twice as long as the average time taken for pure water systems, which is 2 ns on average.

8. Water and Ammonia Selectivity Through Aquaporin AtTIP2;1

Channel	Up	Down	Total
A	134	124	258
B	371	350	721
C	410	411	820
D	114	73	187
			1987
Average Time: 3.8		pf: 11.6	

(a) Lipid phase lipid at 310 K

Channel	Up	Down	Total
A	18	22	39
B	78	75	153
C	128	121	248
D	53	50	103
			544
Average Time: 10.6		pf: 3.4	

(b) Gel phase lipid at 270 K

Channel	Up	Down	Total
A	46	57	103
B	204	197	401
C	334	316	650
D	139	131	270
			1424

(c) Gel phase corrected to 310 K

Table 8.5.: The passage of water across the aquaporin *AtTIP2;1*, each channel in liquid (a) and gel (b) phase lipids in systems where 1000 NH_4^+ molecules are present in the system. The amount of water passing from the cytosol to the vacuole is referred to as up. The number of water going from the vacuole to the cytosol is referred to as down. (c) corrected number of water passages across all channels in the gel phase lipid at 310 K.

Interestingly, water across ammonium-containing systems in the gel phase is more or less similar to the pure water system. The ratio (Up/Down) of H_2O passage across channels B, C, and D is approximately ~ 1.05 . On the other hand, the ratio between the Up and Down passage of H_2O for channel A is approximately 0.82. Moreover, channel C exhibits the highest number of H_2O passages across gel phase lipids containing NH_4^+ ions, in contrast with pure water, which exhibits the lowest number of H_2O passages across the pure water gel phase system. The gel phase ammonium system has 248 water molecules passing through channel C, whereas in the case of the gel phase pure water system, there are 42 water molecules passing through channel C. Also, channel A shows the lowest number of H_2O permeation in the case of the gel phase lipid containing ammonium ions, which shows the 39 H_2O passages over the length of 1 μs simulation in the pure water system.

Further, it is interesting to note that the average time of passage of H_2O across the channels of ammonium ions is quite similar to that of the average time taken for the passage of H_2O across the channels of a pure water system, which is 10.3 ns on average. The bottom line is that the presence of ammonium ions in the gel phase has no effect on the water permeation, but on the other hand, the water permeation goes down to almost half in the case of liquid phase lipid simulation when the ammonium ions are present in the system. A gel phase ammonia system allows the passage of a total of 491 molecules of H_2O ,

unlike a gel phase ammonium system on other hand allows the passage of 544 molecules of H₂O. As compared to the permeation of H₂O in the case of gel phase ammonium systems, which takes an average time of 10.3 ns for the passage of H₂O, the average time for the passage of H₂O in the case of the ammonia systems is 12.6 ns. Channel C shows the lowest amount of permeation of water in the ammonia system, which is 49, which is much lower than channel C corresponding to the ammonium system, which shows a high amount of permeation of 248. Compared to the ammonia system, the system containing ammonium ions has a greater number of water permeation than the channel containing ammonium ions. Further, we have also calculated the gel phase and corrected the number of H₂O passage at 310 K using the Arrhenius formula at activation energy (E_A) 4 kcal/mol. There are 1424 H₂O molecules passages through the hypothetical situation in gel phase lipid at 310 K.

It can be seen from the water passage through the NH₃ and NH₄⁺ system, that pf is inversely proportional to the amount of water average time taken by water molecules to pass through it. The pf water passes through the NH₃ system in liquid phase is $13.9 \times 10^{-14} \text{cm}^3 \text{s}^{-1}$ and $2.4 \times 10^{-14} \text{cm}^3 \text{s}^{-1}$ in gel phase lipid. At the same time, in NH₄⁺ systems, pf of water in the liquid phase is similar to that of NH₃ system, which is $11.6 \times 10^{-14} \text{cm}^3 \text{s}^{-1}$. The pf is less in comparison to the pure water system in liquid phase lipid. The pf of water in NH₃ system is $2.4 \times 10^{-14} \text{cm}^3 \text{s}^{-1}$, which decreased marginally in comparison to the pure water system. Whereas the pf remains similar in the gel phase H₂O and NH₄⁺ system. The pf of water in NH₄⁺ system is $3.4 \times 10^{-14} \text{cm}^3 \text{s}^{-1}$.

8.4. Discussion

8.4.1. Water Permeability and Selectivity

Using classical MD simulation, we investigated how the plant aquaporin TIP2;1 affects the selectivity and permeability of water and ammonia in liquid and gel phase lipids in plants. We have calculated water and ammonia permeabilities for all the simulated systems in liquid and gel phase lipid. In all the simulated systems, the results show that there is a decrease in the permeability of the water across the aquaporin TIP2;1 in the gel phase, to various degrees. In a pure water system, there are overall 3946 H₂O molecules that pass through a liquid phase lipid, whereas there are 645 H₂O molecules that pass through a gel phase system. Furthermore, in the NH₃ containing system, there are 3467 molecules that cross in the liquid phase lipid, and 491 molecules passed in the gel phase lipid. Furthermore, it is even more interesting that there are 1987 H₂O molecules that pass across the aquaporin channels in case of the liquid phase lipid NH₄⁺ ions containing system in comparison to the 3946 pure water system in the liquid phase. In comparison with pure water systems in the liquid phase, the permeation of water is reduced almost by a factor of 2 in NH₄⁺ systems. In contrast, water permeation in the NH₄⁺ system does not differ much from that in the pure water system in gel phase lipid.

On the other hand, while comparing the gel phase corrected value with the liquid phase, water permeation also shows a similar pattern in all the systems. There are 1711, 1284, and 1424 H₂O molecules passing across the aquaporin in pure H₂O, NH₃ and NH₄⁺ systems, respectively, in the corrected gel phase systems. It has little or no influence on H₂O permeation in NH₃ and ammonium ion systems in gel phase lipid.

In order to better understand how water is distributed and interacts in the pore before adding the complexity of ammonia to the model, we first try to understand the distribution of water alone in the pore and the effects it will have on the selective filter region (SF) region. The location of the water molecules within the pore indicates hydrogen bonding between them and the main sites of pore-lining that are formed through a chain of carbonyl oxygen atoms. There are three major categories of amino acid residues that contribute to the stability of the water molecules along the pore of each channel. In the first category, the residues GLY79, GLY80, HSD81, GLY193, SER195, and GLY194, are responsible for stabilising the water through the carbonyl oxygen of the backbone. Among the second category are the N_δ donors of the NPA asparagine ASN83 and ASN197. Finally, the third category belongs to the stability provided by the SF residues such as HSD131 and ILE185.

HSD63(N _ε 2) – ARG200(N _ε)						
Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.04 ± 0.18	3.01 ± 0.14	3.13 ± 0.25	2.97 ± 0.12	3.21 ± 0.45	3.0 ± 0.2
B	3.05 ± 0.18	3.02 ± 0.14	3.06 ± 0.19	3.02 ± 0.14	3.03 ± 0.16	3.02 ± 0.13
C	3.47 ± 0.8	3.09 ± 0.21	3.31 ± 0.52	3.06 ± 0.17	3.13 ± 0.39	3.26 ± 0.7
D	3.19 ± 0.24	3.26 ± 0.28	3.19 ± 0.26	3.24 ± 0.28	3.41 ± 0.35	3.94 ± 0.63

Table 8.6.: Average distance (in Å, mean ± std. deviation) measured from 3 μs simulation between atom N_ε2 of HSD63 and atom N_ε of ARG200.

In contrast to the other SF residues, HSD63 is mainly locked into hydrogen bonding with ARG200. This hydrogen bonding is conserved in SF arginine throughout the simulation in all the systems in both phases of lipid. The average distance between N_ε2 of HSD63 and N_ε of ARG200 in all the systems in both phases of lipid (see table 8.6) is in the range of 3.0 ~ 3.2 Å, except NH₄⁺ ions containing systems in both phases of liquid which show larger distances as well as larger standard deviations. Whereas the distance between the heavy atom (N_ε2) of HSD131 and the NH1 atom of ARG200, shows very larger distances with a larger standard deviation, in all the systems in both phases of the lipid (see table 8.7). It is noteworthy that in all the systems of liquid phase lipids (compare fig. 8.6), the distances are quite large in channel C with a large deviation, whereas channels A, B, and D show stable interaction throughout the simulation through small deviation. Also in the case of gel phase lipid, all the other channels show a stable hydrogen bonding interaction between HSD131 and ARG200 in the pure water system. But, this interaction pattern is changed with the addition of NH₃ and NH₄⁺ ions in all the systems.

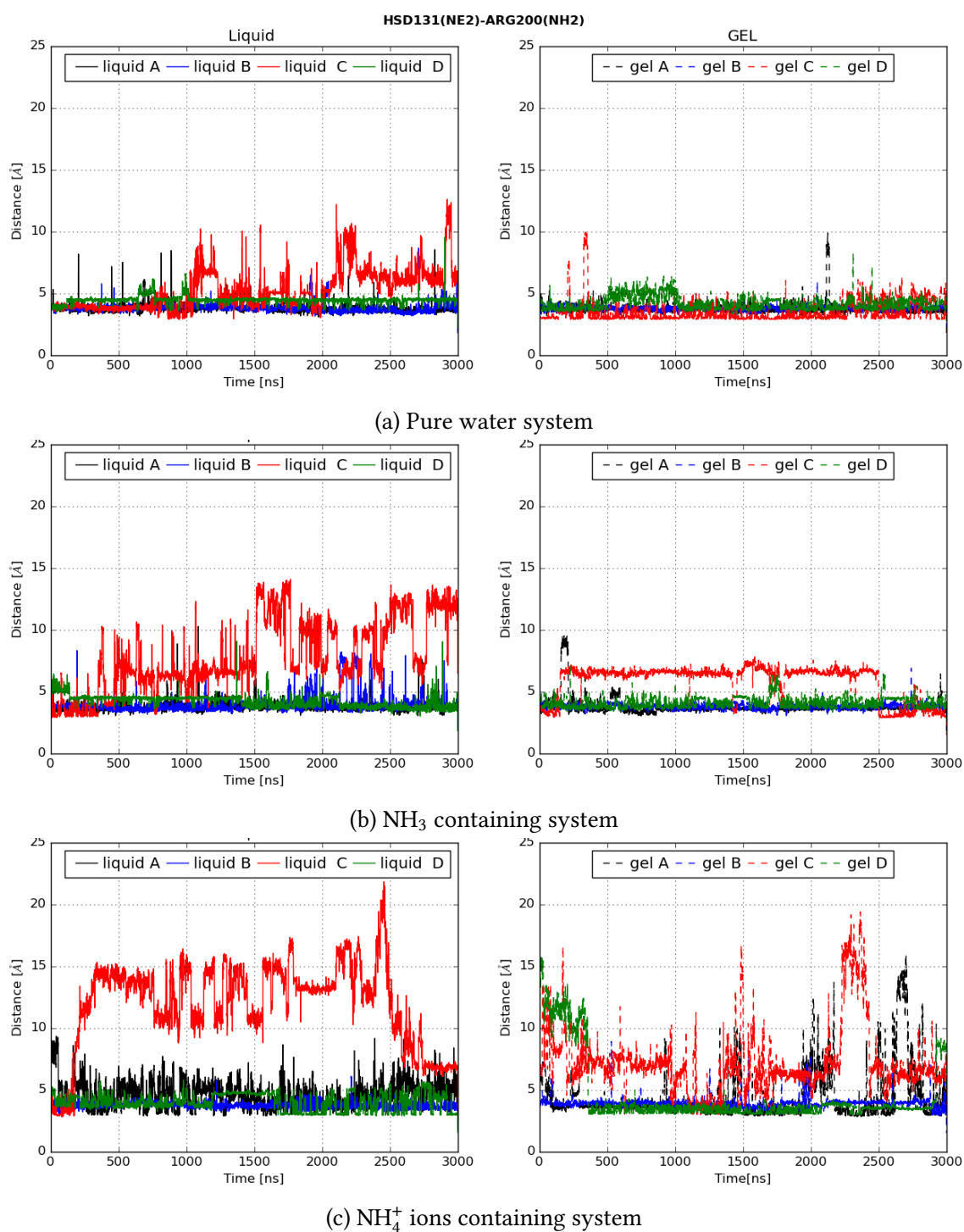


Figure 8.6.: The average distance (in Å) is shown as the evolution of time in liquid phase lipid (on the left side) and gel phase lipid (on the right side) in all the systems for 3 μs . Channel C (red) shows higher fluctuation in all the systems in both phases of lipid except the pure water system in the gel phase.

HSD131(N ϵ_2) - ARG200(NH1)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.61 ± 0.68	3.68 ± 0.83	3.46 ± 0.71	3.36 ± 1.04	4.61 ± 1.27	4.43 ± 2.24
B	3.52 ± 0.55	3.41 ± 0.37	3.83 ± 1.05	3.44 ± 0.39	3.78 ± 0.45	3.91 ± 0.6
C	5.2 ± 2.21	4.18 ± 1.09	7.60 ± 3.10	4.69 ± 0.84	11.95 ± 3.57	7.24 ± 2.88
D	3.27 ± 0.48	4.54 ± 0.77	3.49 ± 0.67	4.32 ± 0.73	4.04 ± 0.77	4.54 ± 2.7

Table 8.7.: Average distance (in Å, mean ± std. deviation) measured from 3 μ s simulation between the atom N ϵ_2 of HSD131 and atom NH1 of ARG200.

Specifically, in channel C of the ammonium system, the distance between HSD131 and ARG200 is very high in both phases of lipids. It is possible that this contributes to the greater number of water passages across channel C, in both phases of the lipid. In the ammonium ions system, 820 water molecules passed through channel C in liquid phase lipid, which is the highest among all the other channels. On the other hand, there are 248 water molecules passing through channel C in the gel lipid containing ammonium ions in the systems, which is also the highest among all the other channels.

Hydrogen bonding between HSD131 and ARG200 does not necessarily prevent water from passing through the channels in two phases of lipid, but it can restrict the pathways by narrowing the size of the pore. Furthermore, HSD131 is also involved in forming hydrogen bonds between the GLY193 or GLY194. This interaction is not stable and often breaks in all the systems, in both phases of lipids, whenever the hydrogen bond is formed between HSD131 and GLY193/GLY194 the side pore is restricted and the pore gets narrower, which makes water or ammonia to pass through it.

Viveca Lindahl and coworkers have already suggested the HSD81 on the cytosolic side can act as a gatekeeper and can affect the passage of water [158]. Histidine HSD81 shifts and closes the pore, which decreases the permeability of water. Kaptan and his coworker studied the pH regulation of permeability through AQP4 by protonating the histidine on the cytosolic side [159].

The swinging motion of HSD81 is described by the dihedral angle $C_\alpha - C_\beta - C_\gamma - N_{\delta_1}$, which is crucial for the opening and closing of the pore (Fig. 8.8). All the simulated systems irrespective of the lipid, show a high probability distribution density at $\sim 95^\circ (\pm 10)$, which corresponds to the open conformation of the aquaporin channel. Whereas all the simulated systems in both phases of the lipid show a small probability distribution density near $\sim -50^\circ (\pm 10)$. In liquid phase lipid, all three simulated systems are almost similar in dihedral angle distribution, having a high probability distribution at $\sim 95^\circ$ and a little hump near $\sim -50^\circ$. Moreover, in the case of the gel phase, the high probability distribution remains unchanged as in the liquid phase lipid, but the slightly larger hump can be seen at -50° . HSD81 shows similar trends in channel B in gel phase lipid across all the systems, which can be seen from the number of water passages, which remains in the range 300~400. Moreover, in the gel phase ammonium system, the slightly bigger hump can be found at

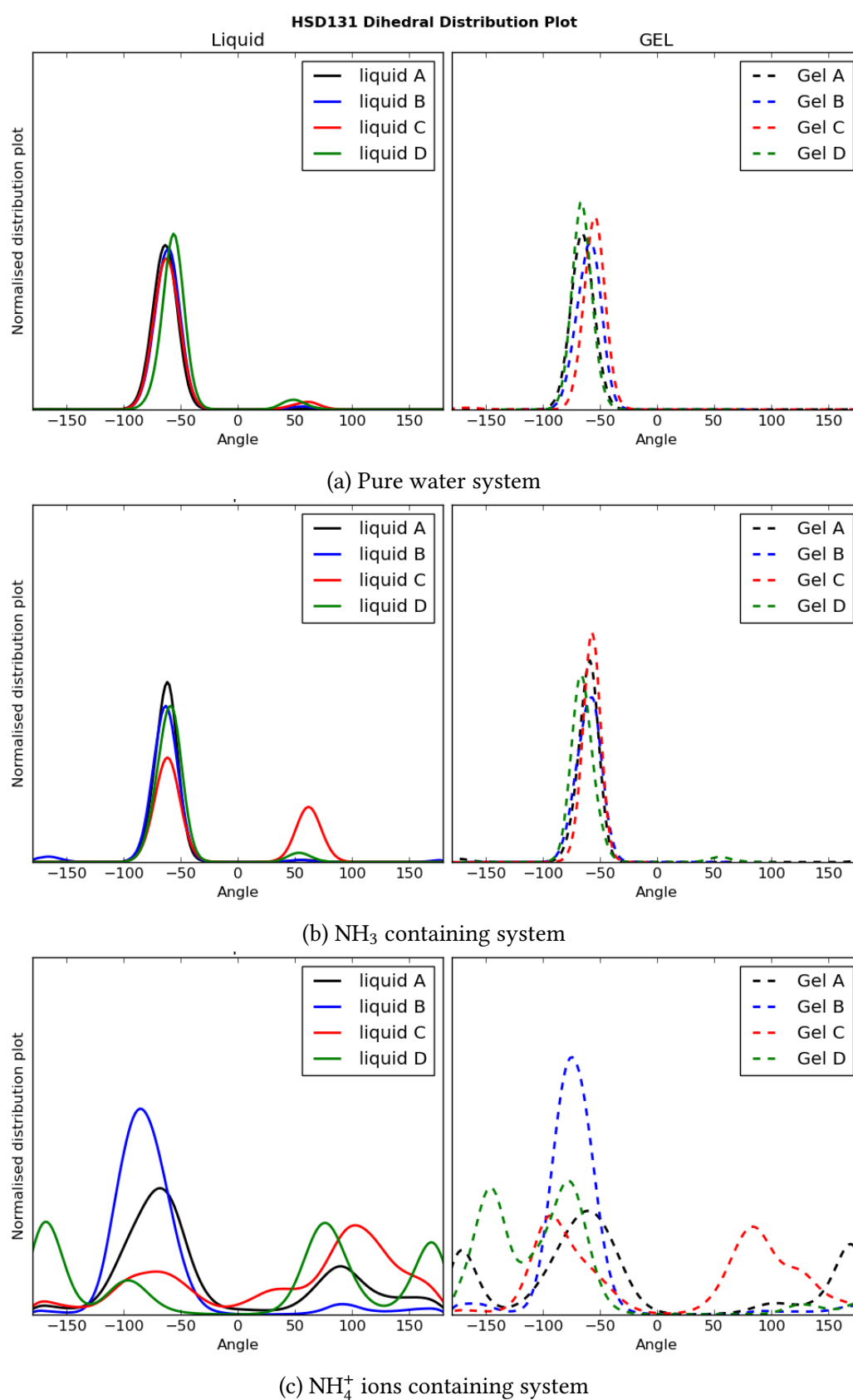
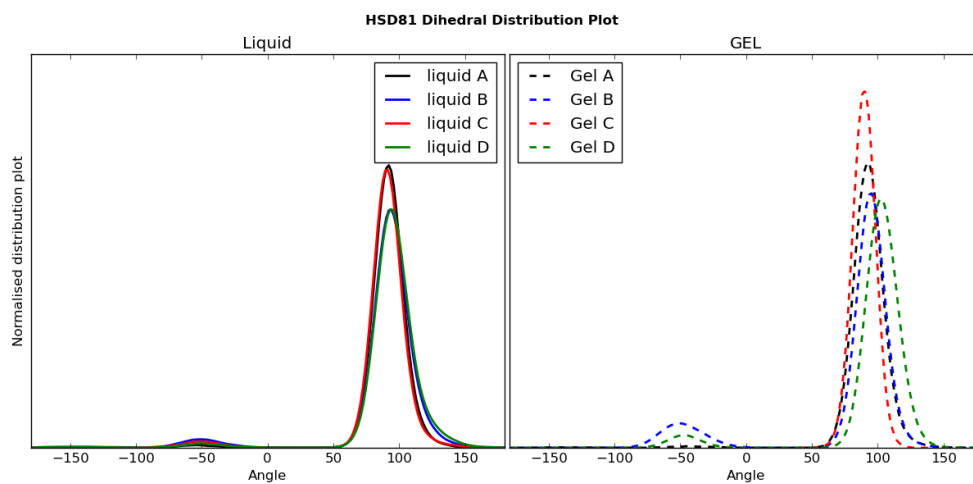
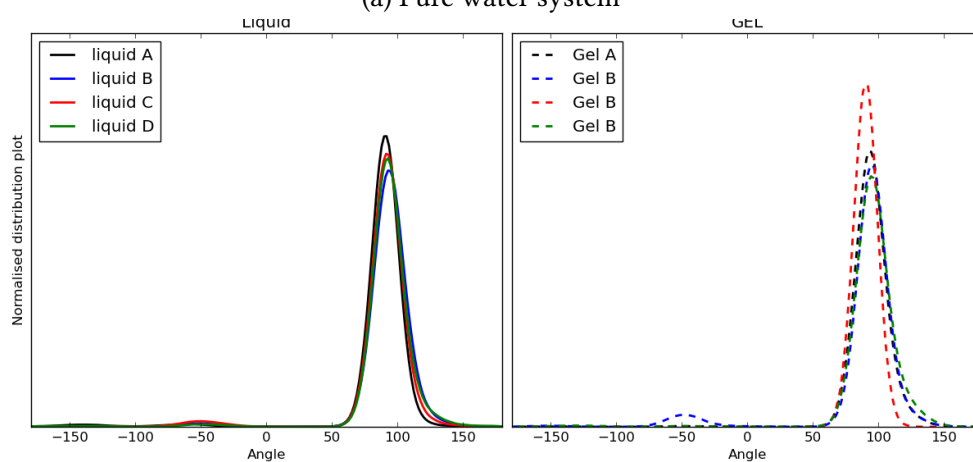


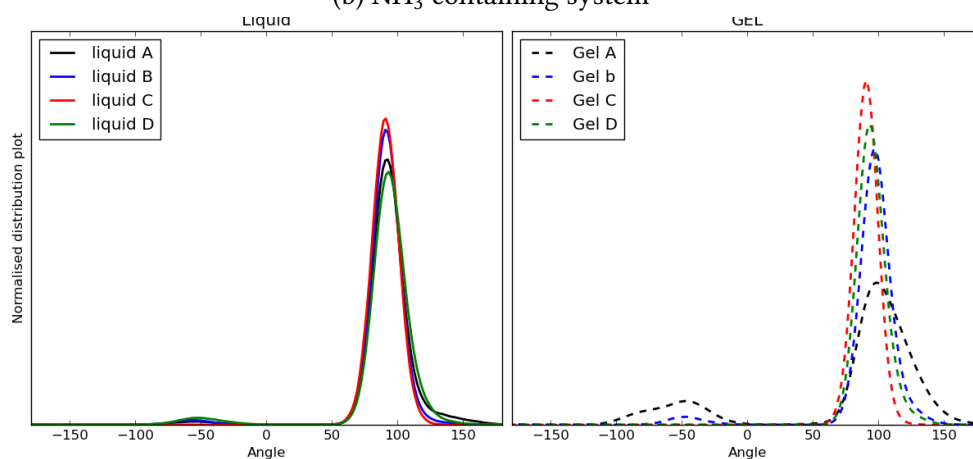
Figure 8.7.: Distribution of HSD131 dihedral angle characterising the formation and deformation of the hydrogen bond with ARG200. The dihedral is defined by $C_\alpha-C_\beta-C_\gamma-N_{\delta_1}$.



(a) Pure water system



(b) NH₃ containing system



(c) NH₄⁺ ions containing system

Figure 8.8.: Distribution of HSD81 dihedral angle characterising the closing and opening of the pore. The dihedral is defined by $C_{\alpha}-C_{\beta}-C_{\gamma}-N_{\delta_1}$.

$\sim -50^\circ$ in channel A. As a consequence of that, the number of passages through channel A in the gel phase ammonium system is significantly lower.

GLU24(CD)–HSD81(ND1)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.55 ± 0.25	3.78 ± 0.3	3.48 ± 0.22	3.74 ± 0.24	3.51 ± 0.21	4.7 ± 0.87
B	3.77 ± 0.31	3.74 ± 0.26	3.84 ± 0.3	3.68 ± 0.2	3.5 ± 0.2	3.82 ± 0.23
C	3.5 ± 0.25	3.44 ± 0.12	3.52 ± 0.25	3.42 ± 0.11	3.43 ± 0.14	3.45 ± 0.13
D	3.83 ± 0.26	4.09 ± 0.38	3.63 ± 0.3	3.76 ± 0.26	3.67 ± 0.33	3.63 ± 0.27

Table 8.8.: Average distance (in Å, mean ± std. deviation) measured from 3 μ s simulation between the atom CD of GLU24 and atom ND1 of HSD81.

GLU24 (CD) – SER78(OG)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	7.77 ± 3.16	3.66 ± 0.17	10.02 ± 1.1	3.67 ± 0.16	9.72 ± 1.82	9.89 ± 1.55
B	5.9 ± 3.35	3.61 ± 0.19	3.72 ± 0.24	3.62 ± 0.18	9.52 ± 2.0	3.65 ± 0.17
C	10.13 ± 1.01	10.03 ± 0.76	10.35 ± 0.81	10.01 ± 0.76	10.14 ± 0.79	10.14 ± 0.8
D	3.71 ± 0.21	9.9 ± 3.17	7.4 ± 3.19	3.7 ± 0.28	9.6 ± 2.36	8.12 ± 3.3

Table 8.9.: Average distance (in Å, mean ± std. deviation) measured from 3 μ s simulation between the atom CD of GLU24 and atom OG of SER78.

The conserved position of glutamic acid on the cytosolic side, here GLU24, suggested that the histidine could be trapped and prevented from closing the pores. In all the systems in both phases of lipid HSD81 is involved in forming hydrogen bonds between GLU24, which remains stable most of the time in the simulation in all the systems. The distance between atom ND1 of HSD81 and atom CD of GLU24 remains in the range of 3.5 ~ 4.0 Å with a small deviation (± 0.2 Å) in most cases (see table 8.8 and fig. 8.9). In the ammonium system of the gel phase, channel A has a higher value for the average distance between HSD81 and GLU24, which is in the range of $\sim 4.7 \pm 0.87$ Å. The deformation of a hydrogen bond between HSD81 and GLU24 in channel A leads to the blocking of a water passage across channel A.

Overall, there are several factors that control the movement of water across each channel of the aquaporin. The stronger hydrogen bond between HSD81 and GLU24 determines the flow of water along the channel. The formation of hydrogen bonds restricts the movement of HSD81 from going into the channel and thus the channels are open. Hydrogen bonds are disrupted and reformed between GLU24 and the adjacent serine SER78 (see table 8.9), which is also indicative of the closing and opening of the aquaporin channels. As long as GLU24–SER78 is connected by a hydrogen bond, HSD81 remains free to move into the channel restricting the water passage. The hydrogen bond between GLU24–SER78 is highly unstable and remains deformed in most parts of the simulation. Representative

8. Water and Ammonia Selectivity Through Aquaporin AtTIP2;1

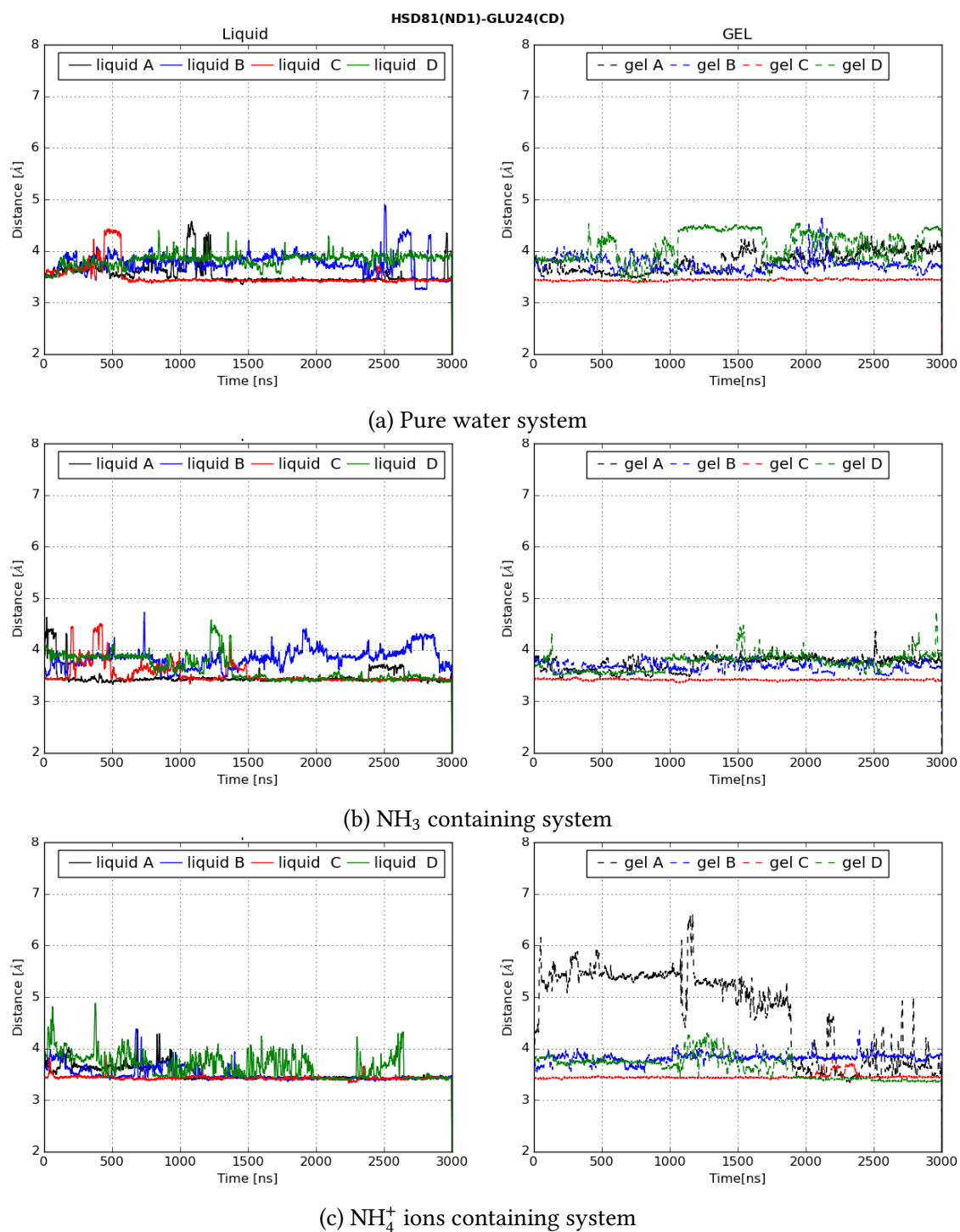


Figure 8.9.: The average distance (in Å) is shown as the evolution of time in liquid phase lipid (on the left side) and gel phase lipid (on the right side) in all the systems for 3 μ s. The distance between heavy atom N _{δ 1} of HSD81 and atom CD of GLU24 is measured.

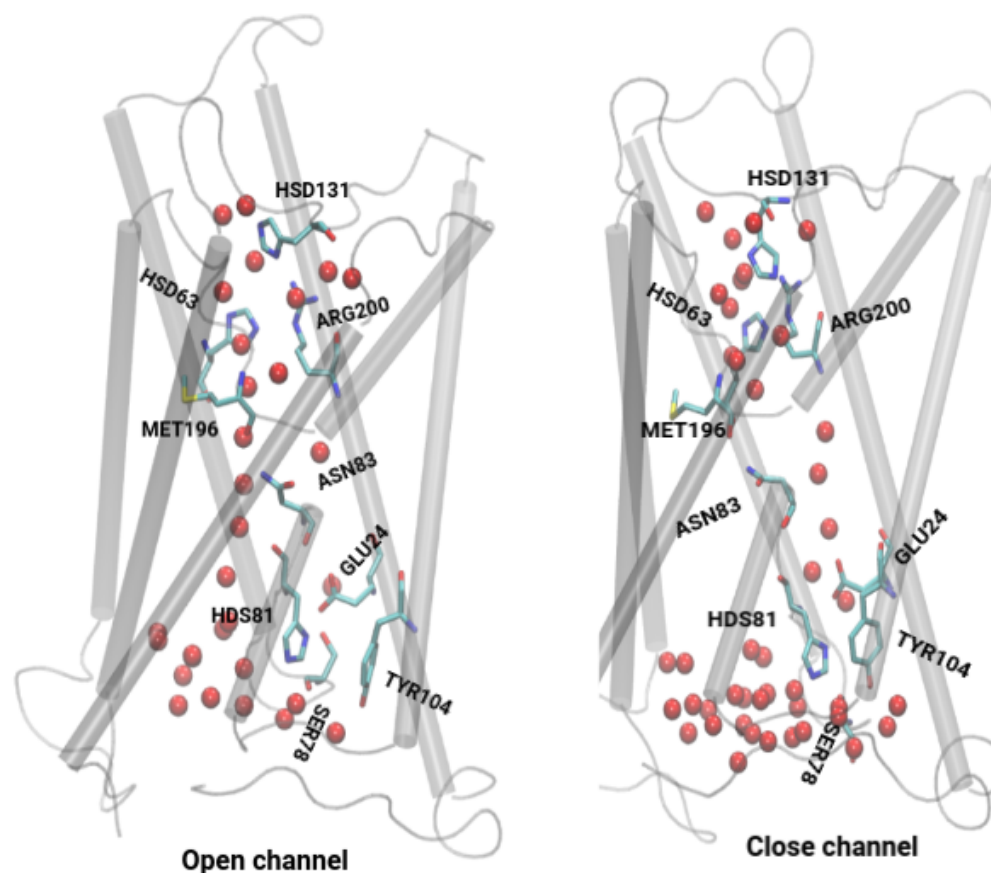


Figure 8.10.: Representative snapshots of the open and closed state of the channel from $3 \mu\text{s}$. The image on the left shows the open state of the channel where HSD131 is directed outward and the HSD81 is bonded with GLU24. In the closed state (right image) the HSD131 is directed inward into the channel and HSD81 is free from hydrogen bonds with GLU24.

snapshots of the open and closed state of the channel are shown in fig. 8.10.

At the NPA motif region of the aquaporin, there are four hydrogen bonds that were conserved in the crystal structure. The hydrogen bond between the ASN197–ALA199, which are on the same chain, remains intact throughout the simulation. The average distances are shown in the appendix section of aquaporin. ASN83 and MET196 form hydrogen bonds across the two chains, which remain stable throughout the simulation in all the systems in both phases of the lipid. The hydrogen bond between ASN83–MET196 and VAL82–ASN197 is stronger when the channels are closed.

8.4.2. Ammonia Selectivity

It has been demonstrated experimentally that the aquaporin *AtTIP2;1* significantly reduces ammonia permeability in gel phase lipids as compared to liquid-phase lipids. The previous

simulation by Groot and his coworker also supports the low permeability of NH_3 across the bilayer facilitated by aquaporins [147]. As reported earlier, ammonia interacts with HSD131 on the vacuole side and in a similar fashion as water in the pores of AQP's [160].

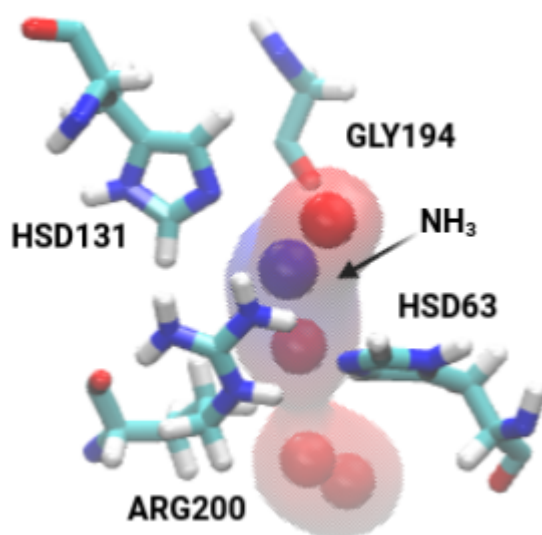


Figure 8.11.: Snapshots from the simulation of the NH_3 system at the selective filter region. NH_3 (blue) is stabilised by forming a hydrogen bond with three polar residues HSD63, HSD131, and ARG200.

In liquid phase lipid at 310 K, there are 59 NH_3 molecules passing through the AQP's over the period of 1 μs MD simulation. Whereas in the case of gel phase lipid at 270 K, 6 molecules of NH_3 passed through the AQP's. For every 10 molecules passed through the AQP's in the liquid phase, only one NH_3 molecule passed through the gel phase lipid at 270 K. The gel phase NH_3 passage corrected for 310 K simulation, there are 17 NH_3 molecules passed through. The ratio for the passage of NH_3 in liquid at 310 K to gel phase at 310 K is 3.47.

In our simulation containing 1000 NH_4^+ in the liquid phase and gel phase lipid, it is often seen that the NH_3 goes into the channel but did not manage to pass through the channel. This situation is more frequent in the gel phase lipid simulation. Also, there are some occasions when NH_3 goes deep into the channel and stays along the chain of water but still did not make it through in the gel phase. And the permeation of NH_3 is controlled by the direction of the flow of the water in the channel in both phases of the lipid. Earlier work from our experimental collaborator suggests that the positioning of phenylalanine (PHE192) can block the pore [147]. Our MD simulation also shows the hydrogen bond between PHE192–ALA146, see fig. 8.12, which varies to various degrees across the channel in all the systems of both phases of lipid. Due to the large size of the phenylalanine and its presence on the periphery of the transmembrane domain, it can be vital for the selectivity of the NH_3 .

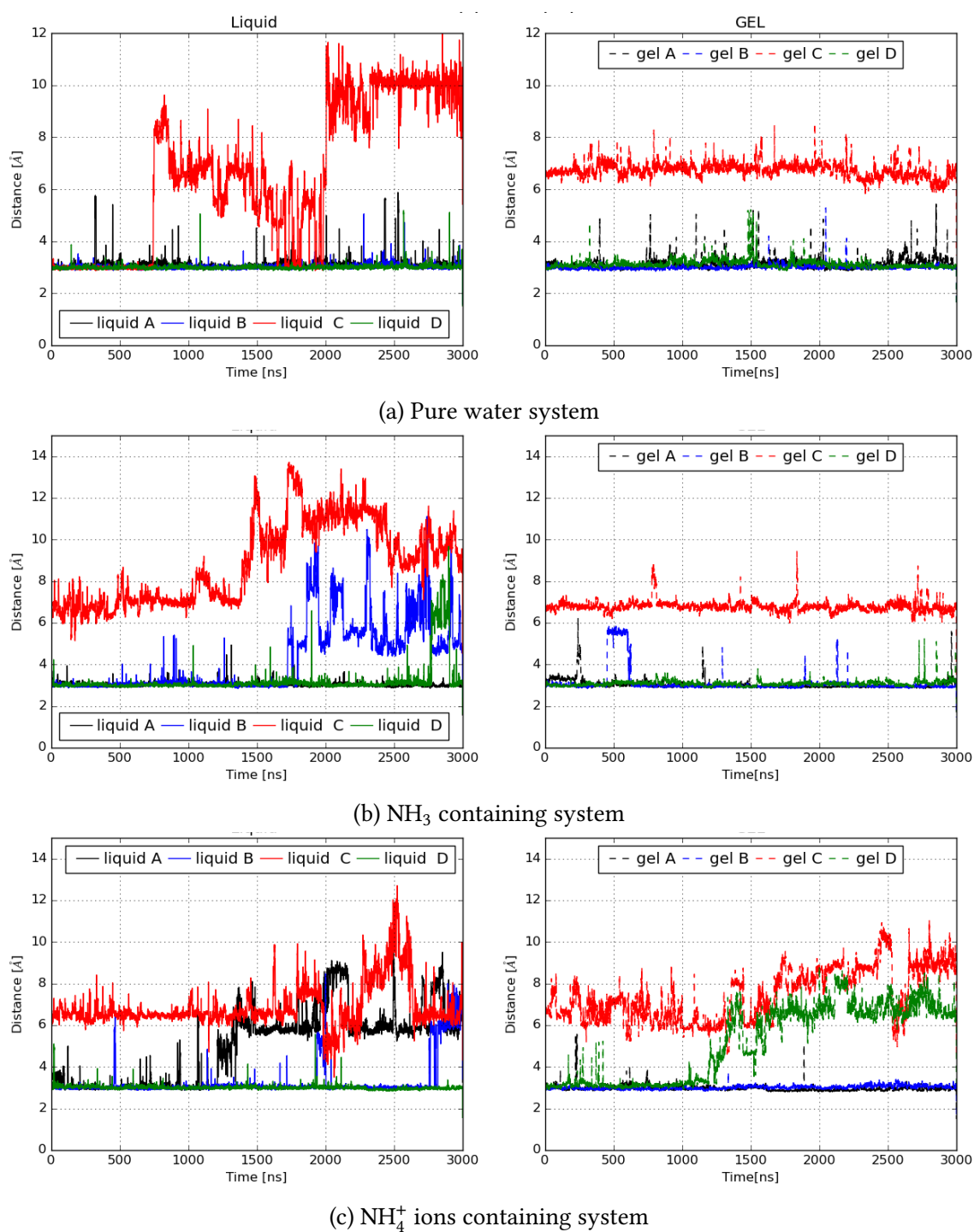


Figure 8.12.: The average distance (in Å) is shown as the evolution of time in liquid phase lipid (on the left side) and gel phase lipid (on the right side) in all the systems for $3 \mu\text{s}$. The distance between heavy atom N_{δ_1} of O of ALA134 and atom OH of PHE192 is measured.

8.4.3. Ammonium Ion Blocking

It is well known that aquaporin does not allow the passage of charged ions across the channel. It is only the passage of a neutral or polar group through the AQPs. Moreover, there is no experimental evidence for the passage of NH_4^+ ions across the aquaporin channels, though the size of the NH_3 and NH_4^+ is not much of a difference. In order to understand why NH_4^+ ions are blocked through this aquaporin, MD simulation of a system containing 1000 NH_4^+ ions is performed in both phases of the lipid for 3 μs . Analysis of these trajectories in both phases of lipid shows that the NH_4^+ ions are completely blocked by the aquaporin AtTIP2;1. Which was not a surprising revelation to find. As expected, MD simulations containing ammonium ions did not exhibit spontaneous permeation, which may be due to electrostatic repulsion and desolvation inside the pore of the aquaporin channel. Further, the passage of water is analysed in both phases of lipid and the gel phase water passage is further corrected to 310 K.

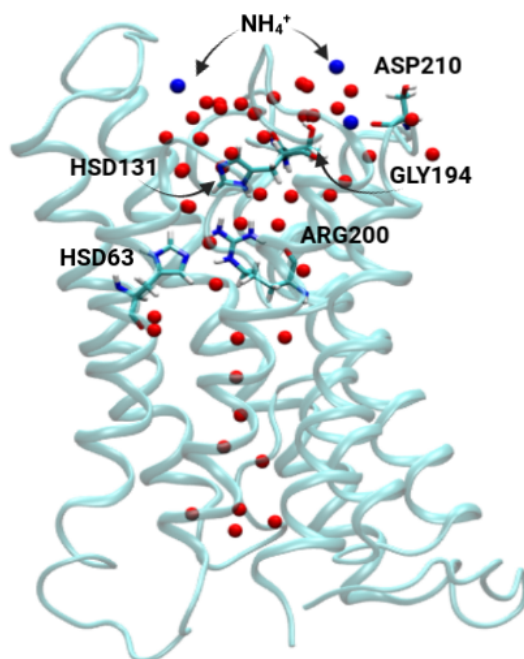


Figure 8.13.: Snapshots of NH_4^+ ions accumulation on the vascular surface. The residue of the selective filter region is shown in the liquorice representation.

There are 1987 H_2O molecules passed through all the channels combined in liquid phase lipid. Which is a surprising revelation from the NH_4^+ ions containing system in liquid phase lipid. As it is almost half the number of water molecules passed through the pure water system in the liquid phase. Further, it is worth noting that the number of water passed through the gel phase in the NH_4^+ ions system is pretty much similar to the number of water passed in the pure water system. But there was experimental evidence to prove this observation from MD simulations at that point in time. After our MD simulation input,

Thomas Barta and Dr. Andreas Horner, our experimental collaborator from Johannes Kepler University, Linz, conducted experiments in order to confirm the results obtained from the simulation. It is important to note that this experiment shows similar results to our MD simulations.

Even though the NH_4^+ ions are blocked from passing across the aquaporin channel, it has been suggested that the accumulation of NH_4^+ ions on the surface of the vacuolar surface of *AtTIP2;1* can increase the permeation of water. As compared with other plant AQPs, the surface of the *AtTIP2;1* is clearly more negative on the vacuolar surface. Acidic residues such as ASP48, ASP52, and ASP210 are most commonly exposed to the negative vacuolar surface of *AtTIP2;1*. Interestingly, our MD simulations reveal that ammonium ions are enriched locally at these acidic residues on the vascular side of the aquaporin.

MD simulation containing NH_4^+ ions shows the spontaneous movement of NH_4^+ ions into the channel from the vascular side of the AQPs. In doing so, NH_4^+ ions form a short-lived hydrogen bonding interaction between the extended filter region, comprising HSD63, HSD131, ARG200, and GLY194. It has been also suggested that the vacuoles side of the *AtTIP2;1* having low pH levels allow ammonium accumulation on the surface. And thus accumulation of NH_4^+ ions can contribute to the channelling of NH_3 molecule would benefit permeation efficiency [147].

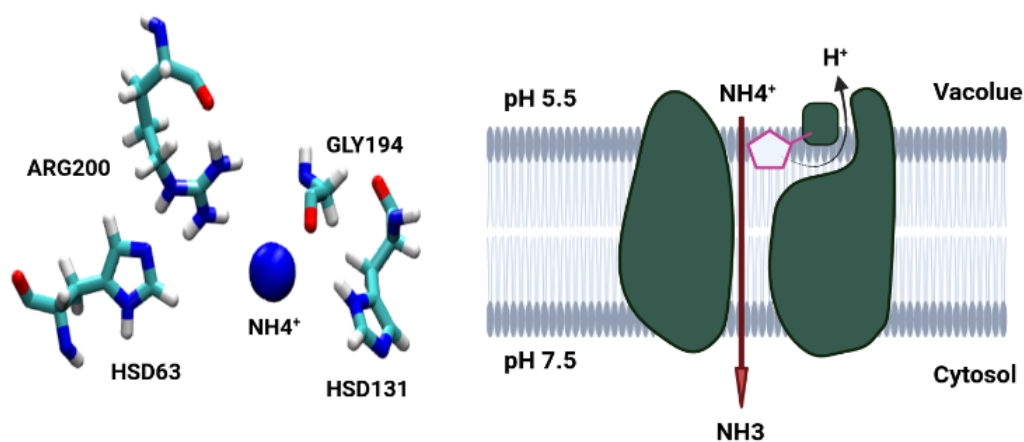


Figure 8.14.: The figure on the left shows the snapshot of electrostatic interaction between the selective filter region (HSD63, HSD131, ARG200, and GLY194) and the ammonium from our MD simulation, and the figure on the right shows the alternative path for the passage of NH_4^+ ions across the aquaporins.

But the question still remains the same, how does the accumulation of NH_4^+ ions will lead to the increase in the permeability of the NH_3 across the AQPs *AtTIP2;1*. It has been previously suggested by Kirscht et. al. regarding the possible mechanism for the NH_4^+ ions permeation via deprotonation [147]. How do the NH_4^+ ions deprotonate to form NH_3 , which can be passed through the membrane? The position of HSD131 residue in the selective

filter region can help in the deprotonation of NH_4^+ ions. Interestingly, our MD simulations suggest the large movement of HSD131, and often the hydrogen bond between HSD131 and ARG200 was broken creating a continuous side pore all the way to the surface of the vacuole, see fig. 8.13. The simulations provide evidence for the side pore's continuous solvation, from HSD131 to the vacuolar exit, so that it can provide a water wire for transferring protons to the vacuolar environment and thus NH_4^+ can be passed as NH_3 . The proton can be shuttled along the water wire by the Grotthuss mechanism. In order to see this proton transfer, QM/MM simulations have to be conducted, which can be a separate project.

8.5. Comparison with Experiment

A comparison was made between MD simulation results and experimental results for the H_2O and NH_3 permeation across both phases of lipid (Table 8.10). The results were quite astounding and in good agreement with the experimental observation.

	H_2O Passage (in Simulation)	Experiment (<i>pf</i>) ($10^{-13}\text{cm}^3\text{s}^{-1}$)	NH_3 Passage (in Simulation)	Experiment (<i>pf</i>) ($10^{-13}\text{cm}^3\text{s}^{-1}$)
Liquid at 310 K	3946	2.41	60	43.31
Gel at 270 K	645	0.31	7	4.96
Gel at 310 K	1711	0.81	17	12.99
Liquid/Gel 310 K	2.30	2.97	3.47	3.33

Table 8.10.: Comparison table of H_2O and NH_3 passage from experiments and the MD simulation in both phases of lipid. The passage in the gel phase is corrected to 310 K using the Arrhenius equation at activation energy of 4 kcal/mol.

The experiment suggested a decrease in the permeation of water as well as the ammonia permeation events across the AQPs in the gel phase. The ratio of water permeation events in the liquid phase at 310 K to the number of water permeation events in the gel phase corrected system at 310 K is 2.30; whereas the experimentally found ratio is 2.97 of the permeation, which is in good agreement with a slight deviation. The experimentally found osmotic probability (*pf*) of water in liquid phase lipid is $2.41 \times 10^{-13}\text{cm}^3\text{s}^{-1}$, which is also in good agreement with *pf* found from the simulation, $2.0 \times 10^{-13}\text{cm}^3\text{s}^{-1}$.

Further comparison of NH_3 permeation events from the experiment and those of MD simulations show a strikingly good agreement. The ratio of NH_3 permeation events in the liquid phase at 310 K to the number of NH_3 permeation events in the gel phase corrected system at 310 K is 3.47; whereas the experimentally found ratio is 3.33. The experiment suggests that for every 10 NH_3 molecule that passes in the liquid phase only one ammonia passes through the gel phase lipid. The results of our MD simulation are rightly supported by the results of experiments.

9. Summary

In this work, state-of-art computer simulations have been used to study the structure and dynamics of integral membrane proteins, and the results from molecular dynamics simulation were compared with experiments, whenever possible.

In chapter 6, all-atom simulations of an antiparallel TisB tetramer assembly in the POPC bilayer were performed. In our simulation, we have seen that the polar interface of the tetramer assembly can mediate the passage of water and proton transfer across the membrane. We further investigated the condition under which water molecules can pass through and the energy costs involved. Through our investigation, it is clear that the polar interface of the TisB is not enough for the passage of water, though it can stabilise the water molecules inside the TisB assembly. We found two factors that control the passage of water across the TisB assembly. First, the salt bridges between ASP-LYS across the subdimer lead to bridging the two subdimer closer, and thus it restricts the movement of water inside the pore of the assembly. Second, the even more important factor which controls the passage of water is the hydrogen bond interaction between Q19-Q19. The Q19-Q19 hydrogen bond across the subdimer blocks the passage of water, whereas the hydrogen bond within the subdimer leaves more room for the passage of water. This work is still in progress and it actually looks good with the current setting.

By combining extensive classical and QM/MM simulations, we also tried to probe the mechanics and energetics of the proton transfer in acidic and basic media. We have employed two reaction coordinates for the proton transfer: (1) centre of excess charge (mCEC) transformed to scalar quantity ζ , (2) z-component of ξ is considered. The maximum energy barrier for the proton transfer is ~ 16.0 kcal/mol and ~ 15.0 kcal/mol using ζ reaction coordinates in an acidic and basic medium, respectively. Whereas reaction coordinate z-component of ξ reveals barrier heights of 9.0 kcal/mol in acidic medium and 13.0 kcal/mol in basic medium. The free energy profile in the acidic and basic media did not reach convergence. The reason can be attributed to the deformation of the water wire, migration of an extra generated proton to an unfavourable position, and interference of MM water in the QM region.

In chapter 7, several sets of classical MD simulations of the proposed tetrameric and hexameric models of the E5-PDGFR β complex in the DGPC bilayer were performed. The complex formed between E5 and PDGFR β is formed via specific interactions between Gln17-Thr513 and Asp33-Lys499 in both 4-helix and 6-helix complexes. During the simulation run, both of these complexes remained stable with little change in the helical orientation, which indicates the hydrophobic contact suggested in the proposed model is maintained. In both of the complexes, the highest fluctuation was seen at the membrane

interface region, whereas the residues which are buried in the membrane show lower fluctuations. The MD simulations show short distances in membrane buried regions, for both of the complexes including Leu506–Leu24, Leu517–Leu10, Leu520–Leu10, and Leu517–Val13 which are in good agreement with the experimental distance measurements for both of the complexes. Helix-helix interactions or conformational stabilisation are likely mediated by these hydrophobic amino acids. Based on distance measurements as well as the helix and tilt angle measurements from our simulation, we concluded that the proposed 4-helix and 6-helix model is in good agreement with the experimental measurements. Further analysis of the PISA wheel of the 4-helix and 6-helix model perfectly matches the experimental measurements. Our MD simulation measurements fit well with both of the proposed models and the activation mechanism of receptors can be explained by these two models.

In chapter 8, several sets of $3\mu\text{s}$ long classical MD simulations of aquaporin *AtTIP2;1* embedded in the OSPC bilayer were performed in liquid and gel phase lipid. The experiment suggests the lower permeability of H_2O and NH_3 across gel phase lipid through aquaporin *AtTip2;1* as compared with liquid phase lipid. This work was mainly aimed at reproducing the experimental observations as well as providing the atomistic detail of the pore of the channel in order to provide a better understanding of what is happening inside the channel. Our simulation shows that all four channels allow passive permeation of water in either direction of the channel and that the average time taken by a H_2O molecule to pass through the gel phase lipid is almost two times higher than the liquid phase lipid. The average time taken by NH_3 in the gel phase is ~ 8 times higher than in liquid phase lipid. Furthermore, our simulation has also confirmed the decrease in the permeability of water by more than half compared to the gel phase lipid. A gel phase lipid system containing 1000 NH_4^+ exhibits considerable hindrance of the permeation of NH_4^+ . For every ten molecules of NH_3 passing through a liquid phase lipid, only one NH_3 is permitted to permeate a gel phase lipid, which perfectly matches the findings of the experiments. The presence of NH_3 has no severe influence on the water permeation in the gel phase and liquid. Compared to pure water systems, simulations of systems containing 1000 NH_4^+ ions show a decrease in the water permeability in liquid phase lipids but remain roughly the same in the gel phase. This is further supported by conducted experiments at Johannes Kepler University, Linz, by Thomas Barta and Dr. Andreas Horner.

According to our simulation trajectory, there are several factors that may influence the aquaporin channel openings and closings and which control the passage of water or ammonia across the aquaporins. Firstly, the movement of the HSD131 at the selective filter region on the vascular surface controls the passage of water across the main pore. The formation of hydrogen bond between HSD131–ARG200 at SF region, complete blockage of water in the side pore but main pore still narrow. Second, the hydrogen bonding interaction at the NPA motif between MET196–ASN83 could also potentially block the aquaporin channel. The breaking of this hydrogen bond can narrow the space in the main pore due to large side chains of MET196, which can potentially block the main pore. Third, the formation and deformation of a hydrogen bond between HSD81–GLU24 on the cytosolic side of the aquaporin can hinder the passage of water/ammonia in the channel. If this

hydrogen bond is dislodged, the side chain of the HSD81 can go into the channel and thus narrow the main channel. All of these three interactions are vital for opening and closing the channel and a single interaction can not completely block or open the channel. But the hydrogen bonding interaction between MET196–ASN83 at the NPE region is more important as it is the centre of the channel and breaks the chain of water in the main pore.

Bibliography

- [1] Mary Luckey. *Membrane structural biology: with biochemical and biophysical foundations*. Cambridge University Press, 2014.
- [2] Erik Strandberg and J Antoinette Killian. “Snorkeling of lysine side chains in transmembrane helices: how easy can it get?” In: *Febs Letters* 544.1-3 (2003), pp. 69–73.
- [3] Mary Luckey. “Introduction to the structural biology of membrane proteins”. In: (2016).
- [4] Denis Maag, Thilo Mast, Marcus Elstner, Qiang Cui, and Tomáš Kubař. “O to bR transition in bacteriorhodopsin occurs through a proton hole mechanism”. In: *Proceedings of the National Academy of Sciences* 118.39 (2021), e2024803118.
- [5] Monika Bayrhuber, Thomas Meins, Michael Habeck, Stefan Becker, Karin Giller, Saskia Villinger, Clemens Vonrhein, Christian Griesinger, Markus Zweckstetter, and Kornelius Zeth. “Structure of the human voltage-dependent anion channel”. In: *Proceedings of the National Academy of Sciences* 105.40 (2008), pp. 15370–15375.
- [6] Manfred S Weiss and Georg E Schulz. “Structure of porin refined at 1.8 Å resolution”. In: *Journal of Molecular Biology* 227.2 (1992), pp. 493–509.
- [7] Suman Bhatia, Phool Singh, and Prabha Sharma. “Hodgkin–Huxley model based on ionic transport in axoplasmic fluid”. In: *Journal of Integrative Neuroscience* 16.4 (2017), pp. 401–417.
- [8] Arthur C Guyton, John Edward Hall, et al. *Textbook of medical physiology*. Vol. 548. Saunders Philadelphia, 1986.
- [9] Dylan Chivian, Timothy Robertson, Richard Bonneau, and David Baker. “Ab initio methods”. In: *Methods of biochemical analysis* 44 (2003), pp. 547–558.
- [10] Anders S Christensen, Tomas Kubar, Qiang Cui, and Marcus Elstner. “Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications”. In: *Chemical Reviews* 116.9 (2016), pp. 5301–5337.
- [11] Adele D Laurent and Denis Jacquemin. “TD-DFT benchmarks: a review”. In: *International Journal of Quantum Chemistry* 113.17 (2013), pp. 2019–2039.
- [12] Dennis C Rapaport and Dennis C Rapaport Rapaport. *The art of molecular dynamics simulation*. Cambridge university press, 2004.
- [13] Stefan Grimme. “Semiempirical GGA-type density functional constructed with a long-range dispersion correction”. In: *Journal of computational chemistry* 27.15 (2006), pp. 1787–1799.

- [14] Peter Elliott and Kieron Burke. “Non-empirical derivation of the parameter in the B88 exchange functional”. In: *Canadian Journal of Chemistry* 87.10 (2009), pp. 1485–1491.
- [15] Peter MW Gill, Benny G Johnson, John A Pople, and Michael J Frisch. “The performance of the Becke–Lee–Yang–Parr (B–LYP) density functional theory with various basis sets”. In: *Chemical Physics Letters* 197.4-5 (1992), pp. 499–505.
- [16] Matthias Ernzerhof and Gustavo E Scuseria. “Assessment of the Perdew–Burke–Ernzerhof exchange–correlation functional”. In: *The Journal of chemical physics* 110.11 (1999), pp. 5029–5036.
- [17] Dirk Porezag, Th Frauenheim, Th Köhler, Gotthard Seifert, and R Kaschner. “Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon”. In: *Physical Review B* 51.19 (1995), p. 12947.
- [18] Gregory M Preston, Tiziana Piazza Carroll, William B Guggino, and Peter Agre. “Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein”. In: *Science* 256.5055 (1992), pp. 385–387.
- [19] Marcus Elstner, Dirk Porezag, G Jungnickel, J Elsner, M Haugk, Th Frauenheim, Sandor Suhai, and Gotthard Seifert. “Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties”. In: *Physical Review B* 58.11 (1998), p. 7260.
- [20] Yang Yang, Haibo Yu, Darrin York, Qiang Cui, and Marcus Elstner. “Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction”. In: *The Journal of Physical Chemistry A* 111.42 (2007), pp. 10861–10873.
- [21] Michael Gaus, Qiang Cui, and Marcus Elstner. “DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB)”. In: *Journal of chemical theory and computation* 7.4 (2011), pp. 931–948.
- [22] Marcus Elstner. “The SCC-DFTB method and its application to biological systems”. In: *Theoretical Chemistry Accounts* 116.1 (2006), pp. 316–325.
- [23] Van Quan Vuong, Jissy Akkarapattiakal Kuriappan, Maximilian Kubillus, Julian J Kranz, Thilo Mast, Thomas A Niehaus, Stephan Irle, and Marcus Elstner. “Parametrization and benchmark of long-range corrected DFTB2 for organic molecules”. In: *Journal of Chemical Theory and Computation* 14.1 (2018), pp. 115–125.
- [24] Michael Gaus, Xiya Lu, Marcus Elstner, and Qiang Cui. “Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications”. In: *Journal of chemical theory and computation* 10.4 (2014), pp. 1518–1537.
- [25] Andrew R Leach and Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.
- [26] Anthony K Rappe and Carla J Casewit. *Molecular mechanics across chemistry*. University Science Books, 1997.

-
- [27] NL Allinger. "Calculation of molecular structure and energy by force-field methods". In: *Advances in physical organic chemistry*. Vol. 13. Elsevier, 1976, pp. 1–82.
- [28] J Phillip Bowen and Norman L Allinger. "Molecular mechanics: the art and science of parameterization". In: *Reviews in computational chemistry* (1991), pp. 81–97.
- [29] Frank H Westheimer and Joseph E Mayer. "The theory of the racemization of optically active derivatives of diphenyl". In: *The Journal of Chemical Physics* 14.12 (1946), pp. 733–738.
- [30] Terrell L Hill. "On steric effects". In: *The Journal of Chemical Physics* 14.7 (1946), pp. 465–465.
- [31] Terrell L Hill. "Steric effects. I. Van der Waals potential energy curves". In: *The Journal of Chemical Physics* 16.4 (1948), pp. 399–404.
- [32] NL Allinger and JT Sprague. "Calculation of Structures of Hydrocarbons Containing Delocalized Electronic Systems by Molecular Mechanics Method". In: *Journal of the American Chemical Society* 95.12 (1973), pp. 3893–3907.
- [33] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. "LINCS: a linear constraint solver for molecular simulations". In: *Journal of computational chemistry* 18.12 (1997), pp. 1463–1472.
- [34] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of computational physics* 23.3 (1977), pp. 327–341.
- [35] Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.
- [36] Loup Verlet. "Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules". In: *Physical review* 159.1 (1967), p. 98.
- [37] Roger Williams Hockney, SP Goel, and JW Eastwood. "Quiet high-resolution computer models of a plasma". In: *Journal of Computational Physics* 14.2 (1974), pp. 148–158.
- [38] Michael P Allen et al. "Introduction to molecular dynamics simulation". In: *Computational soft matter: from synthetic polymers to proteins* 23.1 (2004), pp. 1–28.
- [39] Daan Frenkel, Berend Smit, Jan Tobochnik, Susan R McKay, and Wolfgang Christian. "Understanding molecular simulation". In: *Computers in Physics* 11.4 (1997), pp. 351–354.
- [40] Lars V Bock, Michal H Kolář, and Helmut Grubmüller. "Molecular simulations of the ribosome and associated translation factors". In: *Current opinion in structural biology* 49 (2018), pp. 27–35.
- [41] Gongpu Zhao, Juan R Perilla, Ernest L Yufenyuy, Xin Meng, Bo Chen, Jiyong Ning, Jinwoo Ahn, Angela M Gronenborn, Klaus Schulten, Christopher Aiken, et al. "Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics". In: *Nature* 497.7451 (2013), pp. 643–646.

- [42] Violetta Schneider, Parvesh Wadhvani, Johannes Reichert, Jochen Bürck, Marcus Elstner, Anne S Ulrich, and Tomáš Kubař. “Tetrameric Charge-Zipper Assembly of the TisB Peptide in Membranes—Computer Simulation and Experiment”. In: *The Journal of Physical Chemistry B* 123.8 (2019), pp. 1770–1779.
- [43] Paul D Lyne, Adrian J Mulholland, and W Graham Richards. “Insights into chormate mutase catalysis from a combined QM/MM simulation of the enzyme reaction”. In: *Journal of the American Chemical Society* 117.45 (1995), pp. 11345–11350.
- [44] Chenyi Liao and Jian Zhou. “Replica-exchange molecular dynamics simulation of basic fibroblast growth factor adsorption on hydroxyapatite”. In: *The Journal of Physical Chemistry B* 118.22 (2014), pp. 5843–5852.
- [45] Changjun Chen, Yanzhao Huang, and Yi Xiao. “Enhanced sampling of molecular dynamics simulation of peptides and proteins by double coupling to thermal bath”. In: *Journal of Biomolecular Structure and Dynamics* 31.2 (2013), pp. 206–214.
- [46] Johannes Kästner. “Umbrella sampling”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.6 (2011), pp. 932–942.
- [47] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566.
- [48] Thomas Huber, Andrew E Torda, and Wilfred F Van Gunsteren. “Local elevation: a method for improving the searching properties of molecular dynamics simulation”. In: *Journal of computer-aided molecular design* 8.6 (1994), pp. 695–708.
- [49] Helmut Grubmüller. “Predicting slow structural transitions in macromolecular systems: Conformational flooding”. In: *Physical Review E* 52.3 (1995), p. 2893.
- [50] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. “Well-tempered metadynamics: a smoothly converging and tunable free-energy method”. In: *Physical review letters* 100.2 (2008), p. 020603.
- [51] Tobias Dörr, Marin Vulić, and Kim Lewis. “Ciprofloxacin Causes Persister Formation by Inducing the TisB toxin in *Escherichia coli*”. In: *PLOS Biology* 8.2 (2010), pp. 1–8.
- [52] E. Gerhart H. Wagner and Cecilia Unoson. “The toxin-antitoxin system tisB-istR1”. In: *RNA Biology* 9.12 (2012), pp. 1513–1519.
- [53] Cecilia Unoson and E Gerhart H Wagner. “A small SOS-induced toxin is targeted against the inner membrane in *Escherichia coli*”. In: *Molecular microbiology* 70.1 (2008), pp. 258–270.
- [54] Yoshihiro Yamaguchi, Jung-Ho Park, and Masayori Inouye. “Toxin-antitoxin systems in bacteria and archaea”. In: *Annual review of genetics* 45 (2011), pp. 61–79.
- [55] Philip A Gurnev, Ron Ortenberg, Tobias Dörr, Kim Lewis, and Sergey M Bezrukov. “Persister-promoting bacterial toxin TisB produces anion-selective pores in planar lipid bilayers”. In: *FEBS letters* 586.16 (2012), pp. 2529–2534.
- [56] Torsten H Walther and Anne S Ulrich. “Transmembrane helix assembly and the role of salt bridges”. In: *Current opinion in structural biology* 27 (2014), pp. 63–68.

-
- [57] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. “GROMACS: A message-passing parallel molecular dynamics implementation”. In: *Computer Physics Communications* 91.1 (1995), pp. 43–56. ISSN: 0010-4655. DOI: 10.1016/0010-4655(95)00042-E.
- [58] Erik Lindahl, Berk Hess, and David Van Der Spoel. “GROMACS 3.0: a package for molecular simulation and trajectory analysis”. In: *Molecular modeling annual 7.8* (2001), pp. 306–317.
- [59] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. “GROMACS: fast, flexible, and free”. In: *Journal of computational chemistry* 26.16 (2005), pp. 1701–1718.
- [60] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. “GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation”. In: *Journal of chemical theory and computation* 4.3 (2008), pp. 435–447.
- [61] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David Van Der Spoel, et al. “GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit”. In: *Bioinformatics* 29.7 (2013), pp. 845–854.
- [62] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. “PLUMED 2: New feathers for an old bird”. In: *Computer Physics Communications* 185.2 (2014), pp. 604–613.
- [63] Giovanni Bussi, Davide Donadio, and Michele Parrinello. “Canonical sampling through velocity rescaling”. In: *The Journal of chemical physics* 126.1 (2007), p. 014101.
- [64] Shūichi Nosé. “A molecular dynamics method for simulations in the canonical ensemble”. In: *Molecular physics* 52.2 (1984), pp. 255–268.
- [65] Shuichi Nosé. “A unified formulation of the constant temperature molecular dynamics methods”. In: *The Journal of chemical physics* 81.1 (1984), pp. 511–519.
- [66] Shuichi Nosé and ML Klein. “Constant pressure molecular dynamics for molecular systems”. In: *Molecular Physics* 50.5 (1983), pp. 1055–1076.
- [67] Demian Riccardi, Peter König, Xavier Prat-Resina, Haiibo Yu, Marcus Elstner, Thomas Frauenheim, and Qiang Cui. ““Proton holes” in long-range proton transfer reactions in solution and enzymes: A theoretical analysis”. In: *Journal of the American Chemical Society* 128.50 (2006), pp. 16302–16311.
- [68] Thomas Steinbrecher, Sebastian Prock, Johannes Reichert, Parvesh Wadhvani, Benjamin Zimpfer, Jochen Bürck, Marina Berditsch, Marcus Elstner, and Anne S Ulrich. “Peptide-lipid interactions of the stress-response peptide TisB that induces bacterial persistence”. In: *Biophysical journal* 103.7 (2012), pp. 1460–1469.
- [69] C Wayne Smith. “3. Adhesion molecules and receptors”. In: *Journal of Allergy and Clinical Immunology* 121.2 (2008), S375–S379.
- [70] Svein G Dahl, Ingebrigt Sylte, and Aina Westrheim Ravna. “Structures and models of transporter proteins”. In: *Journal of Pharmacology and Experimental Therapeutics* 309.3 (2004), pp. 853–860.

- [71] Edwin Li and Kalina Hristova. "Role of receptor tyrosine kinase transmembrane domains in cell signaling and human pathologies". In: *Biochemistry* 45.20 (2006), pp. 6241–6251.
- [72] Axel Ullrich and Joseph Schlessinger. "Signal transduction by receptors with tyrosine kinase activity". In: *Cell* 61.2 (1990), pp. 203–212.
- [73] Bengt Westermark, Lena Claesson-Welsh, and Carl-Henrik Heldin. "Structural and functional aspects of the receptors for platelet-derived growth factor". In: *Progress in growth factor research* 1.4 (1989), pp. 253–266.
- [74] Carl-Henrik Heldin and Bengt Westermark. "Mechanism of action and in vivo role of platelet-derived growth factor". In: *Physiological reviews* (1999).
- [75] Jean-Baptiste Demoulin and Ahmed Essaghir. "PDGF receptor signaling networks in normal and cancer cells". In: *Cytokine & growth factor reviews* 25.3 (2014), pp. 273–283.
- [76] Johanna Andrae, Radosa Gallini, and Christer Betsholtz. "Role of platelet-derived growth factors in physiology and medicine". In: *Genes & development* 22.10 (2008), pp. 1276–1312.
- [77] Dan R Robinson, Yi-Mi Wu, and Su-Fang Lin. "The protein tyrosine kinase family of the human genome". In: *Oncogene* 19.49 (2000), pp. 5548–5557.
- [78] Mark A Lemmon and Joseph Schlessinger. "Cell signaling by receptor tyrosine kinases". In: *Cell* 141.7 (2010), pp. 1117–1134.
- [79] Stevan R Hubbard, Jeffrey H Till, et al. "Protein tyrosine kinase structure and function". In: *Annual review of biochemistry* 69.1 (2000), pp. 373–398.
- [80] Stevan R Hubbard. "Structural analysis of receptor tyrosine kinases". In: *Progress in biophysics and molecular biology* 71.3-4 (1999), pp. 343–358.
- [81] Aude I Ségaliny, Marta Tellez-Gabriel, Marie-Françoise Heymann, and Dominique Heymann. "Receptor tyrosine kinases: Characterisation, mechanism of action and therapeutic interests for bone cancers". In: *Journal of bone oncology* 4.1 (2015), pp. 1–12.
- [82] RA Bradshaw, RJ Chalkley, J Biarc, and AL Burlingame. "Receptor tyrosine kinase signaling mechanisms: Devolving TrkA responses with phosphoproteomics". In: *Advances in biological regulation* 53.1 (2013), pp. 87–96.
- [83] Colin W Ward, Michael C Lawrence, Victor A Streltsov, Timothy E Adams, and Neil M McKern. "The insulin and EGF receptor structures: new insights into ligand-induced receptor activation". In: *Trends in biochemical sciences* 32.3 (2007), pp. 129–137.
- [84] Linda Fredriksson, Hong Li, and Ulf Eriksson. "The PDGF family: four gene products form five dimeric isoforms". In: *Cytokine & growth factor reviews* 15.4 (2004), pp. 197–204.

-
- [85] Veli-Matti Leppänen, Andrea E Prota, Michael Jeltsch, Andrey Anisimov, Nisse Kalkkinen, Tomas Strandin, Hilkka Lankinen, Adrian Goldman, Kurt Ballmer-Hofer, and Kari Alitalo. “Structural determinants of growth factor binding and specificity by VEGF receptor 2”. In: *Proceedings of the National Academy of Sciences* 107.6 (2010), pp. 2425–2430.
- [86] Ichiro N Maruyama. “Activation of transmembrane cell-surface receptors via a common mechanism? The “rotation model””. In: *Bioessays* 37.9 (2015), pp. 959–967.
- [87] Charlotte A Bell, John A Tynan, Kristen C Hart, April N Meyer, Scott C Robertson, and Daniel J Donoghue. “Rotational coupling of the transmembrane and kinase domains of the Neu receptor tyrosine kinase”. In: *Molecular biology of the cell* 11.10 (2000), pp. 3589–3599.
- [88] Takanori Moriki, Hiroko Maruyama, and Ichi N Maruyama. “Activation of pre-formed EGF receptor dimers by ligand-induced rotation of the transmembrane domain”. In: *Journal of molecular biology* 311.5 (2001), pp. 1011–1026.
- [89] Oded Livnah, Enrico A Stura, Steven A Middleton, Dana L Johnson, Linda K Jolliffe, and Ian A Wilson. “Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation”. In: *Science* 283.5404 (1999), pp. 987–990.
- [90] Guoqiang Jiang and Tony Hunter. “Receptor signaling: when dimerization is not enough”. In: *Current Biology* 9.15 (1999), R568–R571.
- [91] DM Thompson and GN Gill. “The EGF receptor: structure, regulation and potential role in malignancy.” In: *Cancer surveys* 4.4 (1985), pp. 767–788.
- [92] Peter Blume-Jensen and Tony Hunter. “Oncogenic kinase signalling”. In: *Nature* 411.6835 (2001), pp. 355–365.
- [93] Anne Burkhardt, Mark Willingham, Cyril Gay, Kuala-Teh Jeang, and Richard Schlegel. “The E5 oncoprotein of bovine papillomavirus is oriented asymmetrically in Golgi and plasma membranes”. In: *Virology* 170.1 (1989), pp. 334–339.
- [94] Daniel DiMaio and Dawn Mattoon. “Mechanisms of cell transformation by papillomavirus E5 proteins”. In: *Oncogene* 20.54 (2001), pp. 7866–7873.
- [95] Char-Chang Lai, Carl Henningson, and Daniel DiMaio. “Bovine papillomavirus E5 protein induces oligomerization and trans-phosphorylation of the platelet-derived growth factor β receptor”. In: *Proceedings of the National Academy of Sciences* 95.26 (1998), pp. 15241–15246.
- [96] Joanne Oates, Matthew Hicks, Timothy R Dafforn, Daniel DiMaio, and Ann M Dixon. “In vitro dimerization of the bovine papillomavirus E5 protein transmembrane domain”. In: *Biochemistry* 47.34 (2008), pp. 8985–8992.
- [97] Toral Surti, Ophir Klein, Kathryn Aschheim, Daniel DiMaio, and Steven O Smith. “Structural models of the bovine papillomavirus E5 protein”. In: *Proteins: Structure, Function, and Bioinformatics* 33.4 (1998), pp. 601–612.

- [98] Dirk Windisch, Silke Hoffmann, Sergii Afonin, Stefanie Vollmer, Soraya Benamira, Birgid Langer, Jochen Bürck, Claudia Muhle-Goll, and Anne S Ulrich. “Structural role of the conserved cysteines in the dimerization of the viral transmembrane oncoprotein E5”. In: *Biophysical journal* 99.6 (2010), pp. 1764–1772.
- [99] Dirk Windisch, Colin Ziegler, Jochen Bürck, and Anne S Ulrich. “Structural characterization of a C-terminally truncated E5 oncoprotein from papillomavirus in lipid bilayers”. In: *Biological chemistry* 395.12 (2014), pp. 1443–1452.
- [100] Gavin King, Joanne Oates, Dharmesh Patel, Hugo A van den Berg, and Ann M Dixon. “Towards a structural understanding of the smallest known oncoprotein: investigation of the bovine papillomavirus E5 protein using solution-state NMR”. In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1808.6 (2011), pp. 1493–1501.
- [101] Sebastian [VerfasserIn] Otteni. *Helix-Helix-Wechselwirkungen in Tyrosinkinaserzeptoren*. Karlsruhe, [2021].
- [102] Carl-Henrik Heldin, Arne Östman, and Lars Rönstrand. “Signal transduction via platelet-derived growth factor receptors”. In: *Biochimica et Biophysica Acta (BBA)-reviews on cancer* 1378.1 (1998), F79–F113.
- [103] Lisa Petti, Laura A Nilson, and Daniel DiMaio. “Activation of the platelet-derived growth factor receptor by the bovine papillomavirus E5 transforming protein.” In: *The EMBO journal* 10.4 (1991), pp. 845–855.
- [104] LAURA A Nilson and DANIEL DiMaio. “Platelet-derived growth factor receptor can mediate tumorigenic transformation by the bovine papillomavirus E5 protein”. In: *Molecular and Cellular Biology* 13.7 (1993), pp. 4137–4145.
- [105] David J Goldstein, Weiqun Li, Ling-Mei Wang, Mohammad A Heidarani, Stuart Aaronson, Richard Shinn, Richard Schlegel, and Jacalyn H Pierce. “The bovine papillomavirus type 1 E5 transforming protein specifically binds and activates the beta-type receptor for the platelet-derived growth factor but not other related tyrosine kinase-containing receptors to induce cellular transformation”. In: *Journal of Virology* 68.7 (1994), pp. 4432–4441.
- [106] Claudia Muhle-Goll, Silke Hoffmann, Sergii Afonin, Stephan L Grage, Anton A Polyansky, Dirk Windisch, Marcel Zeitler, Jochen Bürck, and Anne S Ulrich. “Hydrophobic matching controls the tilt and stability of the dimeric platelet-derived growth factor receptor (PDGFR) β transmembrane segment”. In: *Journal of Biological Chemistry* 287.31 (2012), pp. 26178–26186.
- [107] DA Drummond-Barbosa, Richard R Vaillancourt, Andrius Kazlauskas, and Daniel DiMaio. “Ligand-independent activation of the platelet-derived growth factor beta receptor: requirements for bovine papillomavirus E5-induced mitogenic signaling”. In: *Molecular and Cellular Biology* 15.5 (1995), pp. 2570–2581.
- [108] BRUCE D Cohen, DAVID J Goldstein, LAURA Rutledge, WILLIAM C Vass, DR Lowy, R Schlegel, and JT Schiller. “Transformation-specific interaction of the bovine papillomavirus E5 oncoprotein with the platelet-derived growth factor receptor transmembrane domain and the epidermal growth factor receptor cytoplasmic domain”. In: *Journal of virology* 67.9 (1993), pp. 5303–5311.

-
- [109] Lisa Petti and Daniel DiMaio. “Specific interaction between the bovine papillomavirus E5 transforming protein and the beta receptor for platelet-derived growth factor in stably transformed and acutely transfected cells”. In: *Journal of virology* 68.6 (1994), pp. 3582–3592.
- [110] Valerie M Nappi and Lisa M Petti. “Multiple transmembrane amino acid requirements suggest a highly specific interaction between the bovine papillomavirus E5 oncoprotein and the platelet-derived growth factor beta receptor”. In: *Journal of Virology* 76.16 (2002), pp. 7976–7986.
- [111] Ophir Klein, Deena Kegler-Ebo, Jennifer Su, Steven Smith, and Daniel DiMaio. “The bovine papillomavirus E5 protein requires a juxtamembrane negative charge for activation of the platelet-derived growth factor β receptor and transformation of C127 cells”. In: *Journal of virology* 73.4 (1999), pp. 3264–3272.
- [112] Lisa M Petti, Venkat Reddy, Steven O Smith, and Daniel DiMaio. “Identification of amino acids in the transmembrane and juxtamembrane domains of the platelet-derived growth factor receptor required for productive interaction with the bovine papillomavirus E5 protein”. In: *Journal of virology* 71.10 (1997), pp. 7318–7327.
- [113] Alexander G Karabadzha, Lisa M Petti, Francisco N Barrera, Anne PB Edwards, Andrés Moya-Rodríguez, Yury S Polikanov, J Alfredo Freitas, Douglas J Tobias, Donald M Engelman, and Daniel DiMaio. “Two transmembrane dimers of the bovine papillomavirus E5 oncoprotein clamp the PDGF β receptor in an active dimeric conformation”. In: *Proceedings of the National Academy of Sciences* 114.35 (2017), E7262–E7271.
- [114] Sunhwan Jo, Taehoon Kim, Vidyashankara G Iyer, and Wonpil Im. “CHARMM-GUI: a web-based graphical user interface for CHARMM”. In: *Journal of computational chemistry* 29.11 (2008), pp. 1859–1865.
- [115] Irina Kufareva and Ruben Abagyan. “Methods of protein structure comparison”. In: *Homology modeling*. Springer, 2011, pp. 231–257.
- [116] Karen Sargsyan, Cédric Grauffel, and Carmay Lim. “How molecular size impacts RMSD applications in molecular dynamics simulations”. In: *Journal of chemical theory and computation* 13.4 (2017), pp. 1518–1524.
- [117] Wolfgang Kabsch. “A solution for the best rotation to relate two sets of vectors”. In: *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 32.5 (1976), pp. 922–923.
- [118] Jianzhong Chen, Jinan Wang, and Weiliang Zhu. “Molecular mechanism and energy basis of conformational diversity of antibody SPE7 revealed by molecular dynamics simulation and principal component analysis”. In: *Scientific reports* 6.1 (2016), pp. 1–12.
- [119] Nematollah Gheibi, Mohamad Ghorbani, Hanifeh Shariatifar, and Alireza Farasat. “Effects of unsaturated fatty acids (Arachidonic/Oleic Acids) on stability and structural properties of Calprotectin using molecular docking and molecular dynamics simulation approach”. In: *PloS one* 15.3 (2020), e0230780.

- [120] J Antoinette Killian. “Hydrophobic mismatch between proteins and lipids in membranes”. In: *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* 1376.3 (1998), pp. 401–416.
- [121] Richard C Page, Conggang Li, Jian Hu, Fei Philip Gao, and Timothy A Cross. “Lipid bilayers: an essential environment for the understanding of membrane proteins”. In: *Magnetic Resonance in Chemistry* 45.S1 (2007), S2–S11.
- [122] Yee-Hung M Chan and Steven G Boxer. “Model membrane systems and their applications”. In: *Current opinion in chemical biology* 11.6 (2007), pp. 581–587.
- [123] Hongxia Zhao and Pekka Lappalainen. “A simple guide to biochemical approaches for analyzing protein–lipid interactions”. In: *Molecular biology of the cell* 23.15 (2012), pp. 2823–2830.
- [124] Stanley J Opella and Francesca M Marassi. “Structure determination of membrane proteins by NMR spectroscopy”. In: *Chemical reviews* 104.8 (2004), pp. 3587–3606.
- [125] Lindsay A Baker and Marc Baldus. “Characterization of membrane protein function by solid-state NMR spectroscopy”. In: *Current opinion in structural biology* 27 (2014), pp. 48–55.
- [126] Francesca M Marassi, A Ramamoorthy, and Stanley J Opella. “Complete resolution of the solid-state NMR spectrum of a uniformly ¹⁵N-labeled membrane protein in phospholipid bilayers”. In: *Proceedings of the National Academy of Sciences* 94.16 (1997), pp. 8551–8556.
- [127] Stephan L Grage, Xiaojun Xu, Markus Schmitt, Parvesh Wadhwani, and Anne S Ulrich. “¹⁹F-labeling of peptides revealing long-range NMR distances in fluid membranes”. In: *The Journal of Physical Chemistry Letters* 5.24 (2014), pp. 4256–4259.
- [128] Wenbin Luo and Mei Hong. “Determination of the oligomeric number and intermolecular distances of membrane protein assemblies by anisotropic ¹H-driven spin diffusion NMR spectroscopy”. In: *Journal of the American Chemical Society* 128.22 (2006), pp. 7242–7251.
- [129] Lynne S Batchelder, CE Sullivan, Lynn W Jelinski, and DA Torchia. “Characterization of leucine side-chain reorientation in collagen-fibrils by solid-state ²H NMR.” In: *Proceedings of the National Academy of Sciences* 79.2 (1982), pp. 386–389.
- [130] Weiwen Ying, Scott E Irvine, Richard A Beekman, David J Siminovitch, and Steven O Smith. “Deuterium NMR reveals helix packing interactions in phospholamban”. In: *Journal of the American Chemical Society* 122.45 (2000), pp. 11125–11128.
- [131] Wei Liu, Evan Crocker, Stefan N Constantinescu, and Steven O Smith. “Helix packing and orientation in the transmembrane dimer of gp55-P of the spleen focus forming virus”. In: *Biophysical journal* 89.2 (2005), pp. 1194–1202.
- [132] David J Siminovitch. “Solid-state NMR studies of proteins: the view from static ²H NMR experiments”. In: *Biochemistry and cell biology* 76.2-3 (1998), pp. 411–422.

-
- [133] Thomas J Piggot, Jane R Allison, Richard B Sessions, and Jonathan W Essex. “On the calculation of acyl chain order parameters from lipid simulations”. In: *Journal of chemical theory and computation* 13.11 (2017), pp. 5683–5696.
- [134] J Wang, J Denny, C Tian, S Kim, Y Mo, F Kovacs, Z Song, K Nishimura, Z Gan, R Fu, et al. *Imaging membrane protein helical wheels*. 2000.
- [135] CH Wu, A Ramamoorthy, LM Gierasch, and SJ Opella. “Simultaneous characterization of the amide 1H chemical shift, 1H-15N dipolar, and 15N chemical shift interaction tensors in a peptide bond by three-dimensional solid-state NMR spectroscopy”. In: *Journal of the American Chemical Society* 117.22 (1995), pp. 6148–6149.
- [136] Li [VerfasserIn] Tian. *Self-assembly of transmembrane peptides studied by solid-state NMR*. Karlsruhe, [2021?]
- [137] Peter Agre, Mélanie Bonhivers, and Mario J Borgnia. “The aquaporins, blueprints for cellular plumbing systems”. In: *Journal of Biological Chemistry* 273.24 (1998), pp. 14659–14662.
- [138] B Wu and E Beitz. “Aquaporins with selectivity for unconventional permeants”. In: *Cellular and molecular life sciences* 64.18 (2007), pp. 2413–2421.
- [139] Abul Kalam Azad, Topu Raihan, Jahed Ahmed, Al Hakim, Tanvir Hossain Emon, and Parveen Afroz Chowdhury. “Human aquaporins: functional diversity and potential roles in infectious and non-infectious diseases”. In: *Frontiers in Genetics* 12 (2021), p. 654865.
- [140] Jennifer M Carbrey and Peter Agre. “Discovery of the aquaporins and development of the field”. In: *Aquaporins* (2009), pp. 3–28.
- [141] Urszula Kosinska Eriksson, Gerhard Fischer, Rosmarie Friemann, Giray Enkavi, Emad Tajkhorshid, and Richard Neutze. “Subangstrom resolution X-ray structure details aquaporin-water interactions”. In: *science* 340.6138 (2013), pp. 1346–1349.
- [142] Jochen S Hub and Bert L De Groot. “Mechanism of selectivity in aquaporins and aquaglyceroporins”. In: *Proceedings of the National Academy of Sciences* 105.4 (2008), pp. 1198–1203.
- [143] Susanna Törnroth-Horsefield, Yi Wang, Kristina Hedfalk, Urban Johanson, Maria Karlsson, Emad Tajkhorshid, Richard Neutze, and Per Kjellbom. “Structural mechanism of plant aquaporin gating”. In: *Nature* 439.7077 (2006), pp. 688–694.
- [144] Haixin Sui, Bong-Gyoon Han, John K Lee, Peter Walian, and Bing K Jap. “Structural basis of water-specific transport through the AQP1 water channel”. In: *Nature* 414.6866 (2001), pp. 872–878.
- [145] Daxiong Fu, Andrew Libson, Larry JW Miercke, Cindy Weitzman, Peter Nollert, Jolanta Krucinski, and Robert M Stroud. “Structure of a glycerol-conducting channel and the basis for its selectivity”. In: *science* 290.5491 (2000), pp. 481–486.

- [146] Kazuyoshi Murata, Kaoru Mitsuoka, Teruhisa Hirai, Thomas Walz, Peter Agre, J Bernard Heymann, Andreas Engel, and Yoshinori Fujiyoshi. “Structural determinants of water permeation through aquaporin-1”. In: *Nature* 407.6804 (2000), pp. 599–605.
- [147] Andreas Kirscht, Shreyas S Kaptan, Gerd Patrick Bienert, François Chaumont, Poul Nissen, Bert L de Groot, Per Kjellbom, Pontus Gourdon, and Urban Johanson. “Crystal structure of an ammonia-permeable aquaporin”. In: *PLoS biology* 14.3 (2016), e1002411.
- [148] Schrödinger, LLC and Warren DeLano. *PyMOL 2.2.0*. <http://www.pymol.org/pymol>. July 24, 2018.
- [149] Jing Huang and Alexander D. MacKerell Jr. “CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data”. In: *Journal of Computational Chemistry* 34.25 (2013), pp. 2135–2145. DOI: 10.1002/jcc.23354.
- [150] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD – Visual Molecular Dynamics”. In: *Journal of Molecular Graphics* 14 (1996), pp. 33–38.
- [151] J. D. Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.
- [152] Fangqiang Zhu, Emad Tajkhorshid, and Klaus Schulten. “Theory and simulation of water permeation in aquaporin-1”. In: *Biophysical Journal* 86.1 (2004), pp. 50–57.
- [153] Bert L De Groot, Tomaso Frigato, Volkhard Helms, and Helmut Grubmüller. “The mechanism of proton exclusion in the aquaporin-1 water channel”. In: *Journal of molecular biology* 333.2 (2003), pp. 279–293.
- [154] Bert L de Groot and Helmut Grubmüller. “The dynamics and energetics of water permeation and proton exclusion in aquaporins”. In: *Current opinion in structural biology* 15.2 (2005), pp. 176–183.
- [155] Siladitya Padhi and U Deva Priyakumar. “Selectivity and transport in aquaporins from molecular simulation studies”. In: *Vitamins and Hormones* 112 (2020), pp. 47–70.
- [156] Hanning Chen, Boaz Ilan, Yujie Wu, Fangqiang Zhu, Klaus Schulten, and Gregory A Voth. “Charge delocalization in proton channels, I: the aquaporin channels and proton blockage”. In: *Biophysical journal* 92.1 (2007), pp. 46–60.
- [157] Nilmadhab Chakrabarti, Benoit Roux, and Régis Pomès. “Structural determinants of proton blockage in aquaporins”. In: *Journal of molecular biology* 343.2 (2004), pp. 493–510.
- [158] Viveca Lindahl, Pontus Gourdon, Magnus Andersson, and Berk Hess. “Permeability and ammonia selectivity in aquaporin TIP2; 1: linking structure to function”. In: *Scientific reports* 8.1 (2018), pp. 1–13.
- [159] Shreyas Kaptan, Mette Assentoft, Hans Peter Schneider, Robert A Fenton, Joachim W Deitmer, Nanna MacAulay, and Bert L de Groot. “H95 is a pH-dependent gate in aquaporin 4”. In: *Structure* 23.12 (2015), pp. 2309–2318.

-
- [160] Bert L de Groot and Helmut Grubmuller. “Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF”. In: *Science* 294.5550 (2001), pp. 2353–2357.

Part IV.
Appendix

10. Water and Proton Transport Across Tetrameric Charge Zipper Protein

10.1. Potential Gating mechanism at the center of the peptide

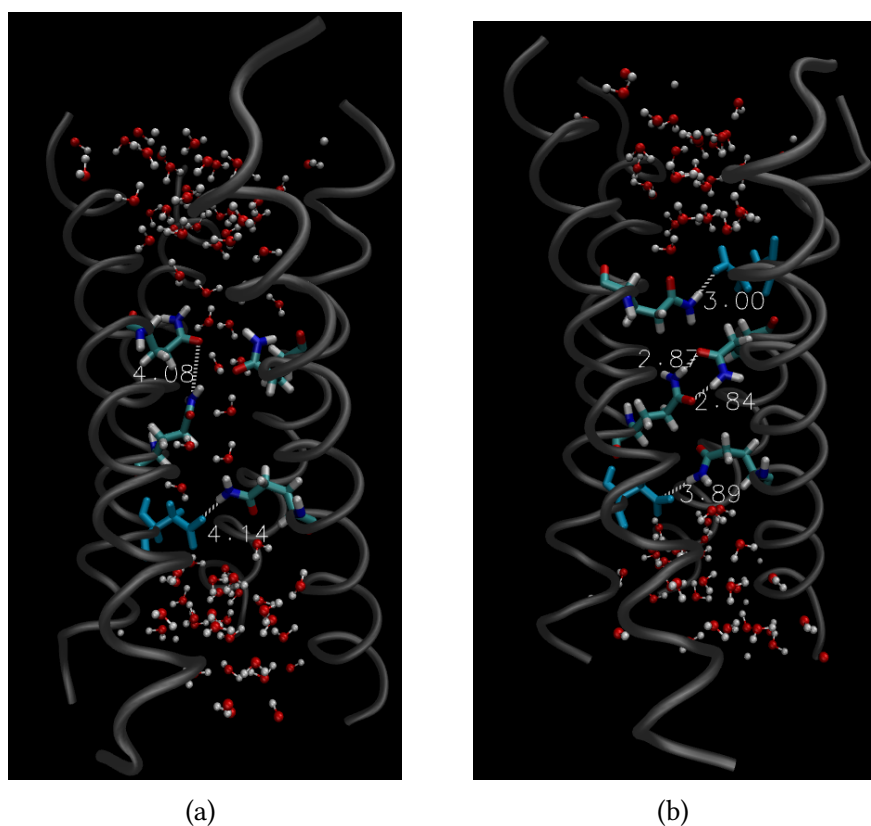
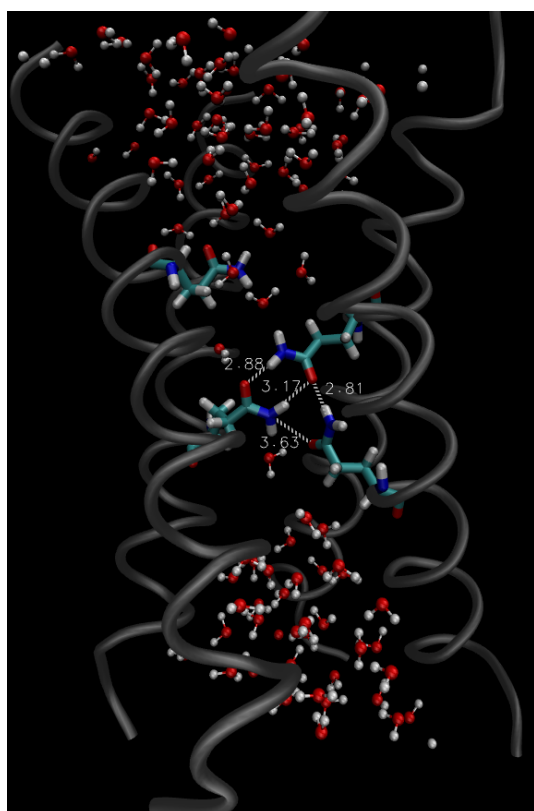
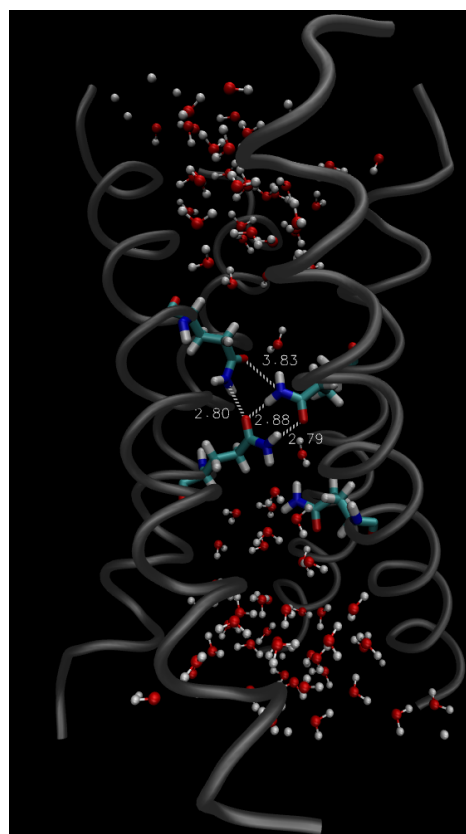


Figure 10.1.: The figure shows the centrally located four Q19 residues and D22 (cyan) in cluster 1. Figure (a) shows the hydrogen bonding interaction between the two Q19 residues within the subdimer. Whereas one of the residues of the other subdimer are dynamically interacting with D22 and other solvated with water molecules. Figure (b) shows the hydrogen bond between Q19–Q19 residues across the subdimer. The remaining two Q19 residues are interacting with D22 of the opposite helix.



(a)



(b)

Figure 10.2.: The figure shows the centrally located interactions between the three Q19 residues. The two Q19 residues are engaged in hydrogen bonding across the subdimer and the third Q19 residue interacts with these two Q19 residues from below (a) and above (b) in the vertical direction.

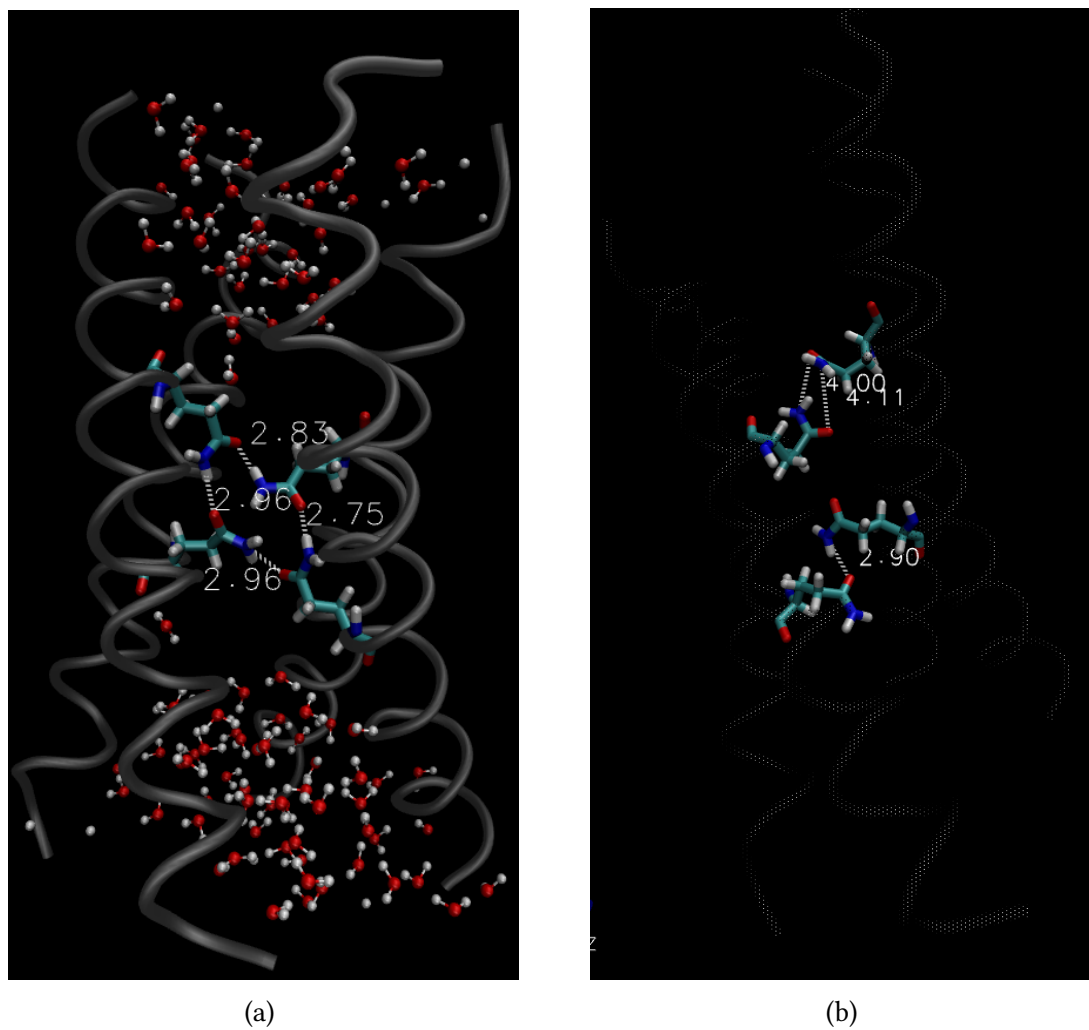


Figure 10.3.: Figure (a) shows the centrally located four Q19 residues which are engaged in extensive hydrogen bonding between the subdimer and across the helix, forming six hydrogen bonds. Figure (b) shows the two different sets of hydrogen bonding interactions between Q19 residues across the helix.

11. Structural Analysis of E5-PDGFR Transmembrane Proteins

11.1. Order parameter of N-H bond in tetrameric complexes

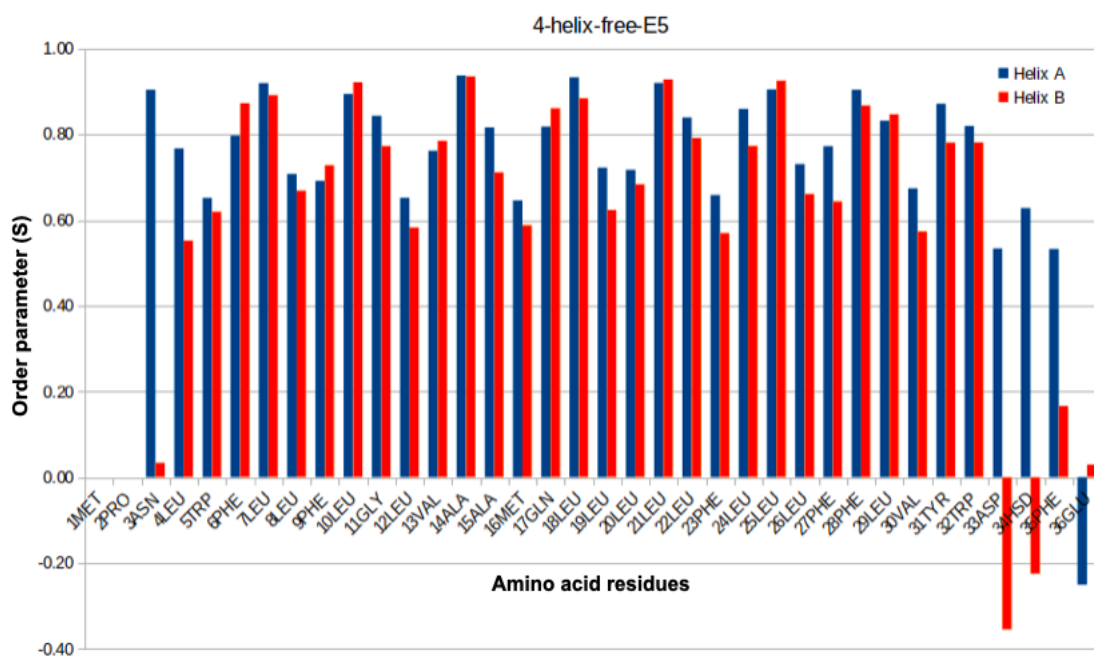


Figure 11.1.: N-H Order parameter along a pair of E5 helices of the 4-helix complex from the simulation free simulation when there is no restraint enforced during equilibration and production simulation. Each E5 has 36 residues.

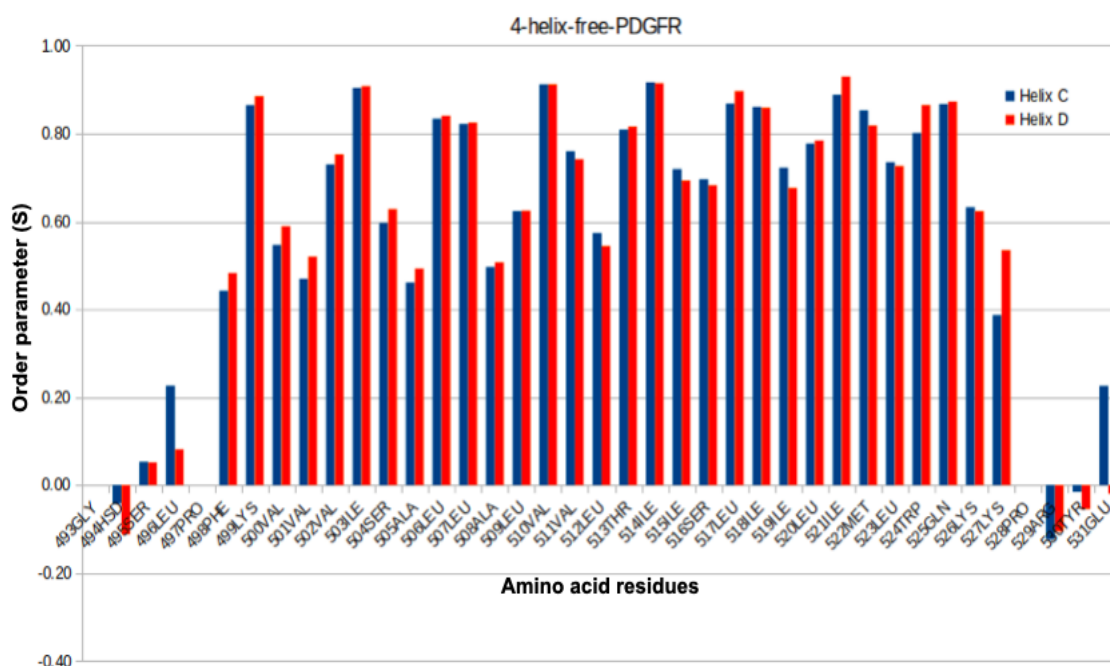


Figure 11.2.: N-H Order parameter along two flanking PDGFR β helices of the 4-helix complex from the simulation free simulation when there is no restraint enforced during equilibration and production simulation. Each PDGFR β has 39 residues.

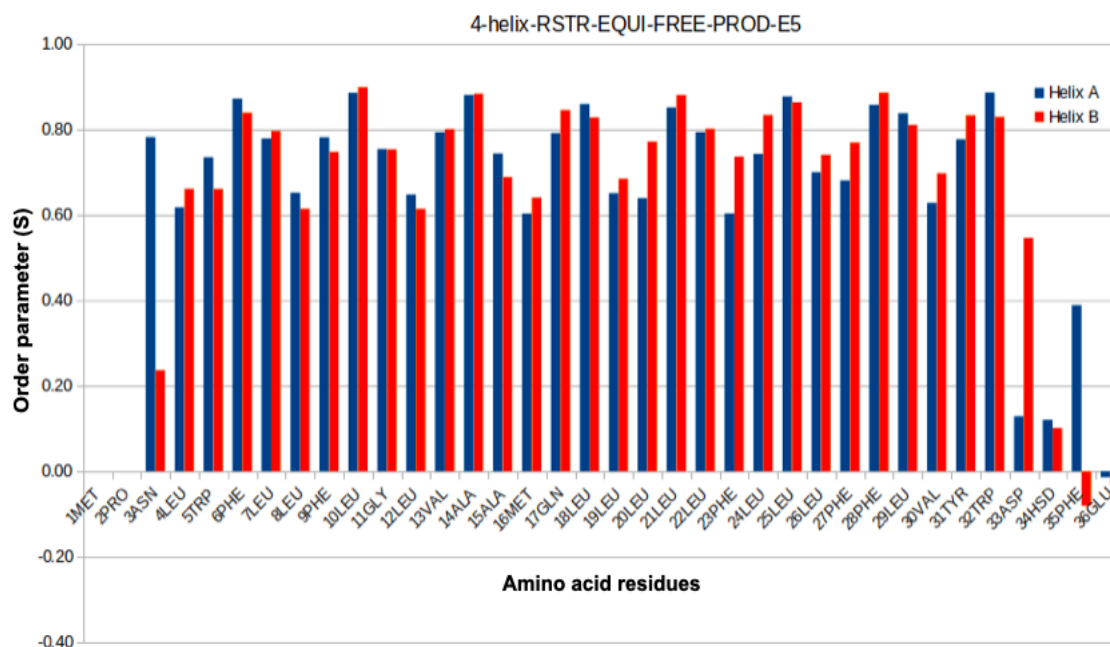


Figure 11.3.: N-H Order parameter along a pair of E5 helices of the 4-helix complex from the simulation when restrained is enforced during equilibration and released during production simulation. Each E5 has 36 residues.

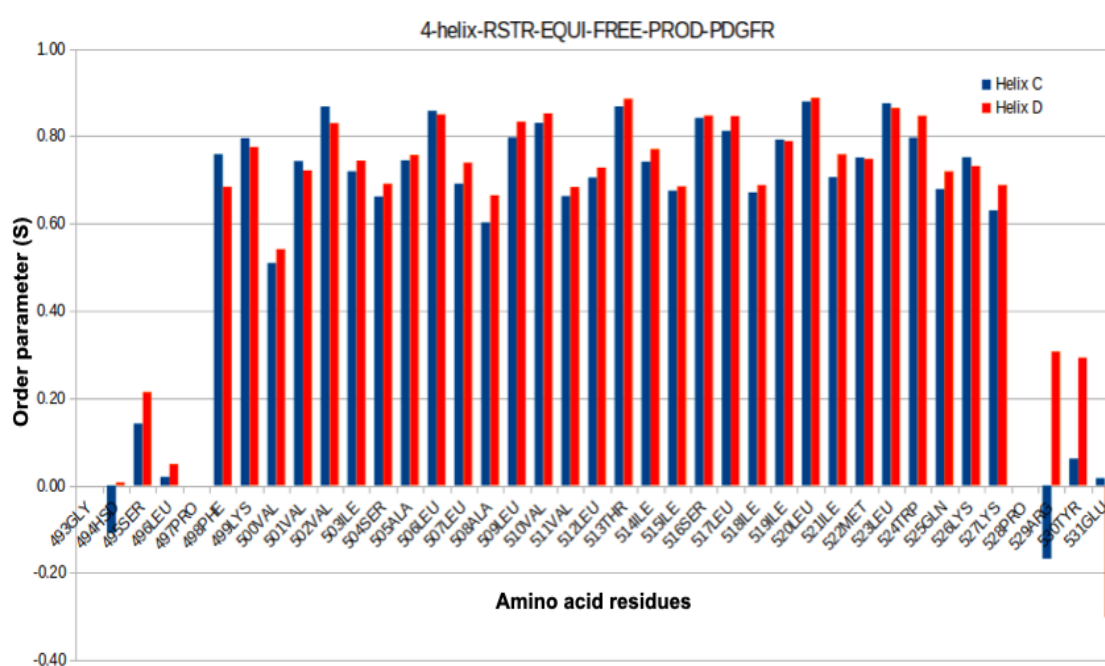


Figure 11.4.: N-H Order parameter along two flanking PDGFR β helices of the 4-helix complex from the simulation when restrained is enforced during equilibration and released during production simulation. Each PDGFR β has 39 residues.

11.2. Order parameter of N-H bond in hexameric complexes

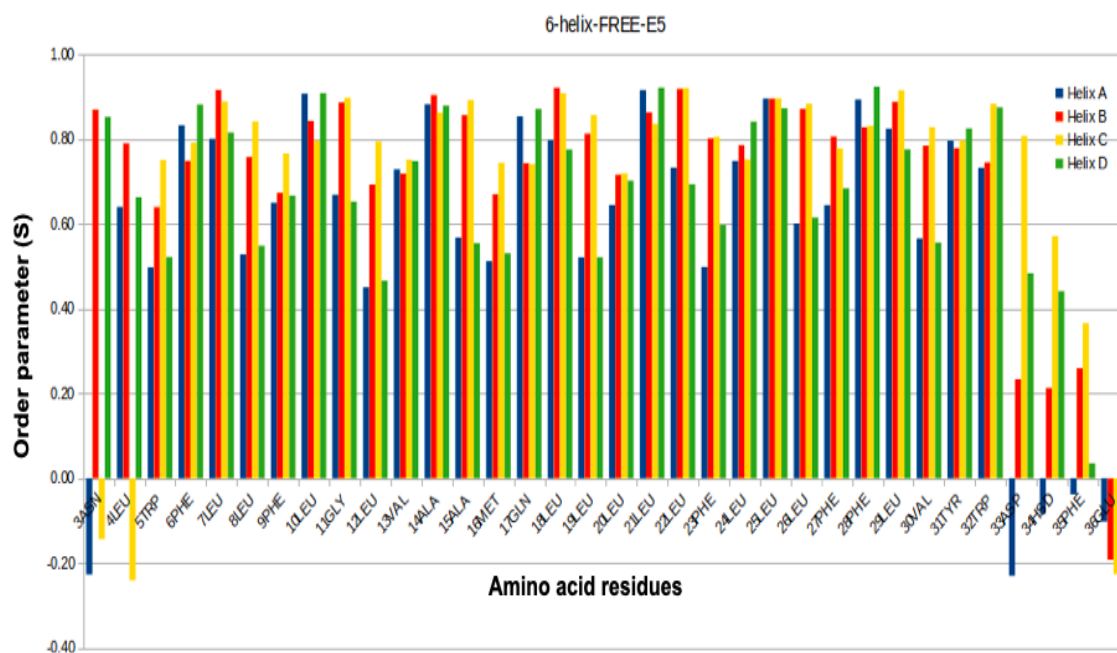


Figure 11.5.: N-H Order parameter along a pair of E5 helices(A, B, C, and D) of the 6-helix complex from the simulation free simulation when there is no restraint enforced during equilibration and production simulation. Each E5 has 36 residues.

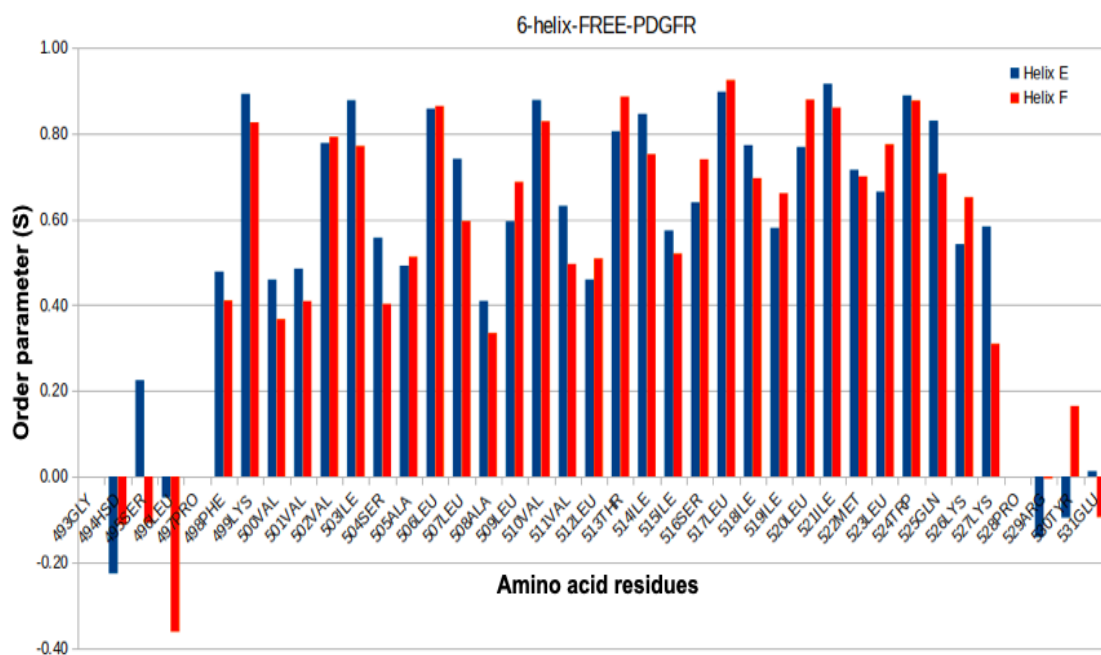


Figure 11.6.: N-H Order parameter along two flanking PDGFR β helices(E and F) of the 6-helix complex from the simulation free simulation when there is no restraint enforced during equilibration and production simulation. Each PDGFR β has 39 residues.



Figure 11.7.: N-H Order parameter along a pair of E5 helices(A, B, C, and D) of the 6-helix complex from the simulation when restrained is enforced during equilibration and released during production simulation. Each E5 has 36 residues.

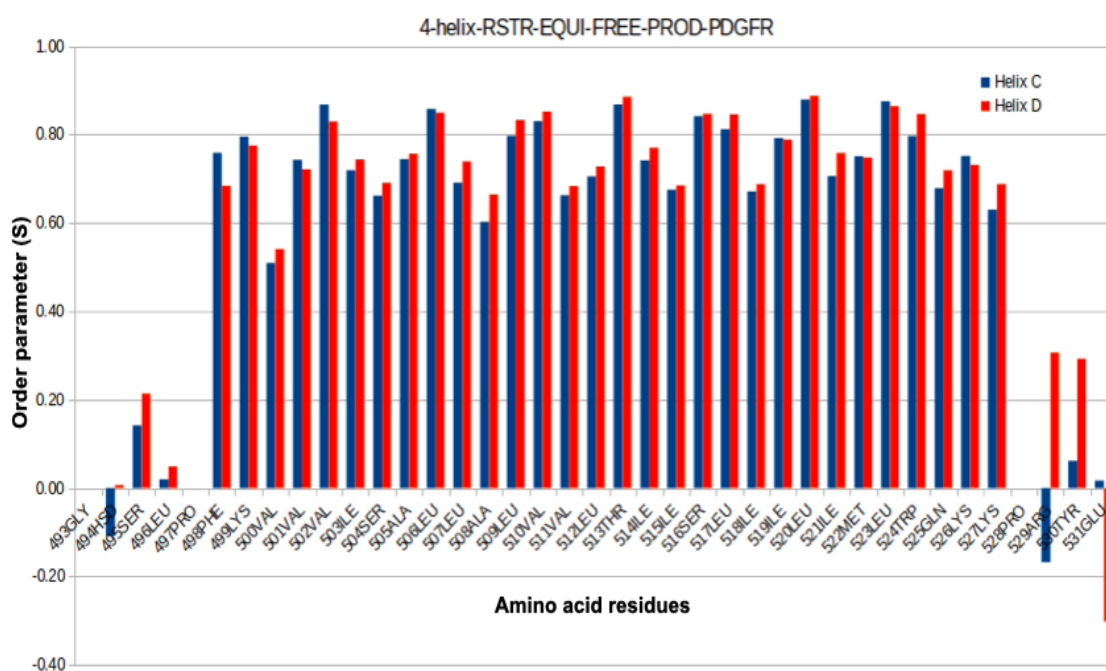


Figure 11.8.: N-H Order parameter along two flanking PDGFR β helices(E and F) of the 6-helix complex from the simulation when restrained is enforced during equilibration and released during production simulation. Each PDGFR β has 39 residues.

11.3. Distance measurements in tetrameric complexes

E5	PDGFR β	A - C	A - D	B - C	B - D	Experiment
Phe28 (F)	(C ¹ H ₃) Val502	9.0	17.0	15.8	10.3	7.4 ~ 9.8
Phe28 (F)	(C ² H ₃) Val502	10.4	18.7	17.6	11.9	
Phe28 (F)	(C ¹ H ₃) Leu506	15.2	9.8	10.1	14.6	8.3 ~ 11.3
Phe28 (F)	(C ² H ₃) Leu506	15.7	11.0	10.6	15.6	
Phe27 (F)	(C ¹ H ₃) Leu506	11.5	16.8	17.9	9.4	> 11.3
Phe27 (F)	(C ² H ₃) Leu506	12.0	17.8	18.3	10.6	
Leu24 (D ₃ C ¹)	(C ¹ H ₃) Leu506	12.8	6.0	7.9	11.3	8.0 ~ 10.0
Leu24 (D ₃ C ²)	(C ² H ₃) Leu506	12.0	7.0	8.0	11.2	
Leu24 (D ₃ C ²)	(C ¹ H ₃) Leu506	11.8	6.3	7.9	10.6	
Leu24 (D ₃ C ¹)	(C ² H ₃) Leu506	13.1	6.9	8.2	12.0	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu517	11.0	5.8	4.8	10.9	6.0 ~ 7.0
Leu10 (D ₃ C ²)	(C ² H ₃) Leu517	11.2	5.8	6.3	11.0	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu517	10.9	5.7	5.6	11.0	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu517	11.2	5.9	5.4	11.0	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu520	13.8	7.3	8.1	13.8	-
Leu10 (D ₃ C ²)	(C ² H ₃) Leu520	13.3	6.6	7.8	12.7	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu520	13.3	7.5	7.8	13.4	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu520	13.8	6.5	8.1	13.0	
Ala14 (D ₃ C)	(C ¹ H ₃) Leu517	5.5	8.8	9.2	4.9	
Ala14 (H ₃ C)	(C ² H ₃) Leu517	5.3	8.8	9.1	4.9	
Val13 (D ₃ C ¹)	(C ¹ H ₃) Leu517	12.3	7.3	5.9	11.1	6.0 ~ 7.0
Val13 (D ₃ C ²)	(C ² H ₃) Leu517	11.1	5.8	5.3	10.5	
Val13 (D ₃ C ²)	(C ¹ H ₃) Leu517	11.4	5.8	5.3	10.6	
Val13 (D ₃ C ¹)	(C ² H ₃) Leu517	12.0	7.4	6.2	11.1	
Leu10 (H ₃ C ¹)	(F) Trp524	13.6	9.6	9.1	1.33	7.4 ~ 8.3
Leu10 (H ₃ C ²)	(F) Trp524	13.5	9.4	9.3	1.39	
Gln17 (C)	(N) Thr513	12.4	5.6	12.5	5.6	
Asp33 (O)	(N) Lys499	12.6	18.6	16.0	13.4	

Table 11.1.: The distance (in Å) between the selected pairs of amino acids side chains from free production simulation(SP1). Tetramer is composed of two E5 helix (A, B) and two PDGFR β (C, D). The distances are compared with the experiment which is measured by the REDOR method.

E5	PDGFR β	A - C	A - D	B - C	B - D	Experiment
Phe28 (F)	(C ¹ H ₃) Val502	13.3	15.7	17.7	9.2	7.4 ~ 9.8
Phe28 (F)	(C ² H ₃) Val502	14.2	17.2	18.7	10.5	
Phe28 (F)	(C ¹ H ₃) Leu506	14.5	14.2	9.4	14.2	8.3 ~ 11.3
Phe28 (F)	(C ² H ₃) Leu506	16.2	13.6	11.1	13.6	
Phe27 (F)	(C ¹ H ₃) Leu506	10.2	18.0	16.8	15.2	> 11.3
Phe27 (F)	(C ² H ₃) Leu506	11.6	17.1	18.4	14.7	
Leu24 (D ₃ C ¹)	(C ¹ H ₃) Leu506	12.1	10.3	7.4	13.2	8.0 ~ 10.0
Leu24 (D ₃ C ²)	(C ² H ₃) Leu506	12.7	9.3	8.5	11.9	
Leu24 (D ₃ C ²)	(C ¹ H ₃) Leu506	11.6	9.7	7.6	12.5	
Leu24 (D ₃ C ¹)	(C ² H ₃) Leu506	13.3	9.9	8.6	12.5	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu517	9.4	5.0	5.9	8.6	6.0 ~ 7.0
Leu10 (D ₃ C ²)	(C ² H ₃) Leu517	9.6	6.1	4.8	9.2	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu517	9.3	4.7	3.9	8.6	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu517	9.7	6.4	9.7	9.6	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu520	14.1	5.0	5.9	8.6	-
Leu10 (D ₃ C ²)	(C ² H ₃) Leu520	15.3	5.6	9.7	11.2	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu520	15.3	6.4	9.7	11.9	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu520	14.1	5.6	9.6	9.8	
Ala14 (D ₃ C)	(C ¹ H ₃) Leu517	4.9	7.8	8.3	11.9	
Ala14 (H ₃ C)	(C ² H ₃) Leu517	4.5	8.2	8.3	9.8	
Val13 (D ₃ C ¹)	(C ¹ H ₃) Leu517	11.2	7.3	7.4	10.6	6.0 ~ 7.0
Val13 (D ₃ C ²)	(C ² H ₃) Leu517	9.9	6.3	6.6	9.8	
Val13 (D ₃ C ²)	(C ¹ H ₃) Leu517	10.6	5.7	6.0	10.1	
Val13 (D ₃ C ¹)	(C ² H ₃) Leu517	10.3	8.3	8.3	10.2	
Leu10 (H ₃ C ¹)	(F) Trp524	7.8	9.9	5.3	11.8	7.4 ~ 8.3
Leu10 (H ₃ C ²)	(F) Trp524	7.5	9.5	3.7	9.7	
Gln17 (C)	(N) Thr513	13.0	5.4	5.4	12.3	
Asp33 (O)	(N) Lys499	3.5	20.9	23.0	6.4	

Table 11.2.: The distance (in Å) between the selected pairs of amino acids side chains from simulation where protein are restrained in equilibration and released during production simulation(SP2). The tetramer is composed of two E5 helix (A, B) and two PDGFR β (C, D). The distances are compared with the experiment which is measured by the REDOR method.

11.4. Distance measurements in hexameric complexes

E5	PDGFR β	A - E	B - E	C - F	D - F	Experiment
Phe28 (F)	(C ¹ H ₃) Val502	9.6	16.8	21.2	17.7	7.4 ~ 9.8
Phe28 (F)	(C ² H ₃) Val502	8.6	15.4	19.5	16.2	
Phe28 (F)	(C ¹ H ₃) Leu506	9.8	14.7	18.2	15.9	8.3 ~ 11.3
Phe28 (F)	(C ² H ₃) Leu506	10.7	15.5	17.5	15.4	
Phe27 (F)	(C ¹ H ₃) Leu506	16.8	8.7	16.7	19.1	> 11.3
Phe27 (F)	(C ² H ₃) Leu506	17.5	9.5	16.3	18.5	
Leu24 (D ₃ C ¹)	(C ¹ H ₃) Leu506	11.0	5.4	12.4	15.0	8.0 ~ 10.0
Leu24 (D ₃ C ²)	(C ² H ₃) Leu506	10.9	6.3	10.8	14.2	
Leu24 (D ₃ C ²)	(C ¹ H ₃) Leu506	11.4	6.1	12.0	14.5	
Leu24 (D ₃ C ¹)	(C ² H ₃) Leu506	10.4	5.8	1.12	14.5	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu517	10.9	6.8	5.0	9.9	6.0 ~ 7.0
Leu10 (D ₃ C ²)	(C ² H ₃) Leu517	10.6	4.4	4.5	9.9	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu517	10.9	4.8	5.2	10.0	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu517	10.8	6.2	4.5	9.9	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu520	13.3	7.9	8.4	14.0	-
Leu10 (D ₃ C ²)	(C ² H ₃) Leu520	12.8	8.3	8.5	14.3	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu520	12.4	7.2	8.3	14.1	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu520	11.8	8.7	8.7	14.3	
Ala14 (D ₃ C)	(C ¹ H ₃) Leu517	5.6	9.3	8.9	4.9	
Ala14 (H ₃ C)	(C ² H ₃) Leu517	11.8	5.4	7.3	11.3	
Val13 (D ₃ C ¹)	(C ¹ H ₃) Leu517	11.4	5.1	7.2	10.2	6.0 ~ 7.0
Val13 (D ₃ C ²)	(C ² H ₃) Leu517	12.8	5.7	8.4	11.6	
Val13 (D ₃ C ²)	(C ¹ H ₃) Leu517	12.4	5.9	8.8	10.5	
Val13 (D ₃ C ¹)	(C ² H ₃) Leu517	11.8	5.4	7.3	11.3	
Leu10 (H ₃ C ¹)	(F) Trp524	13.4	9.5	10.8	11.9	7.4 ~ 8.3
Leu10 (H ₃ C ²)	(F) Trp524	12.2	9.6	10.3	12.0	
Gln17 (C)	(N) Thr513	11.6	4.4	8.3	13.4	
Asp33 (O)	(N) Lys499	12.2	13.5	26.7	6.2	

Table 11.3.: The distance (in Å) between the selected pairs of amino acids side chains from free production simulation(SP1). Hexamer is composed of two pair E5 helix (A, B, C, D) and two PDGFR β (E, F). There are no close contacts in the pairs E-C, E-D, F-A, and F-B in the hexamer due to the construction of the model, so these data are not shown here. The distances are compared with the experiment which is measured by the REDOR method.

E5	PDGFR β	A - E	B - E	C - F	D - F	Experiment
Phe28 (F)	(C ¹ H ₃) Val502	16.6	2.32	18.7	9.6	7.4 ~ 9.8
Phe28 (F)	(C ² H ₃) Val502	14.8	2.18	16.8	7.9	
Phe28 (F)	(C ¹ H ₃) Leu506	12.7	19.7	15.1	7.9	8.3 ~ 11.3
Phe28 (F)	(C ² H ₃) Leu506	12.4	19.4	16.1	9.2	
Phe27 (F)	(C ¹ H ₃) Leu506	19.4	16.3	9.4	16.5	> 11.3
Phe27 (F)	(C ² H ₃) Leu506	19.2	16.1	10.4	17.7	
Leu24 (D ₃ C ¹)	(C ¹ H ₃) Leu506	17.8	13.7	7.2	11.7	8.0 ~ 10.0
Leu24 (D ₃ C ²)	(C ² H ₃) Leu506	17.3	13.5	7.7	12.3	
Leu24 (D ₃ C ²)	(C ¹ H ₃) Leu506	17.4	13.3	7.6	12.4	
Leu24 (D ₃ C ¹)	(C ² H ₃) Leu506	17.6	13.8	7.3	11.6	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu517	17.0	9.2	6.9	13.0	6.0 ~ 7.0
Leu10 (D ₃ C ²)	(C ² H ₃) Leu517	17.4	7.7	9.4	13.1	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu517	16.9	8.0	8.9	13.0	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu517	17.5	8.9	7.4	13.0	
Leu10 (D ₃ C ¹)	(C ¹ H ₃) Leu520	21.1	10.5	4.8	17.3	-
Leu10 (D ₃ C ²)	(C ² H ₃) Leu520	21.4	10.4	6.3	16.8	
Leu10 (D ₃ C ²)	(C ¹ H ₃) Leu520	21.0	9.8	6.0	17.2	
Leu10 (D ₃ C ¹)	(C ² H ₃) Leu520	21.6	11.0	4.7	16.8	
Ala14 (D ₃ C)	(C ¹ H ₃) Leu517	11.7	10.7	8.4	8.3	
Ala14 (H ₃ C)	(C ² H ₃) Leu517	17.9	7.9	5.2	13.4	
Val13 (D ₃ C ¹)	(C ¹ H ₃) Leu517	17.0	7.5	4.9	13.6	6.0 ~ 7.0
Val13 (D ₃ C ²)	(C ² H ₃) Leu517	19.6	8.9	7.3	15.0	
Val13 (D ₃ C ²)	(C ¹ H ₃) Leu517	18.7	8.9	7.0	15.2	
Val13 (D ₃ C ¹)	(C ² H ₃) Leu517	17.9	7.9	5.2	13.4	
Leu10 (H ₃ C ¹)	(F) Trp524	16.2	9.3	8.2	17.1	7.4 ~ 8.3
Leu10 (H ₃ C ²)	(F) Trp524	16.0	9.3	7.4	17.1	
Gln17 (C)	(N) Thr513	18.7	10.2	4.3	14.1	
Asp33 (O)	(N) Lys499	3.8	23.4	18.9	5.3	

Table 11.4.: The distance (in Å) between the selected pairs of amino acids side chains from simulation where protein are restrained in equilibration and released during production simulation(SP2). There are no close contacts in the pairs E-C, E-D, F-A, and F-B in the hexamer due to the construction of the model, so these data are not shown here. The distances are compared with the experiment which is measured by the REDOR method.

12. Water and Ammonia Selectivity Through Aquaporin AtTIP2;1

12.1. Permeation of water (H₂O) and ammonia (NH₃)

Channel	Up	Down	Total
A	1544	1266	2810
B	1452	1447	2899
C	1627	1383	3010
D	1738	1381	3119
			11838
Average Time: 2.0 ns		pf: 16.4	

(a) Lipid phase lipid at 310K

Channel	Up	Down	Total
A	335	361	696
B	205	199	404
C	59	67	126
D	382	353	735
			1961
Average Time: 10.3 ns		pf: 3.3	

(b) Gel phase lipid at 270K

Channel	Up	Down	Total
A	877	945	1821
B	536	521	1057
C	154	175	330
D	1000	924	1923
			5132

(c) Gel phase corrected to 310K

Table 12.1.: The passage of water across the aquaporin AtTIP2;1 in liquid (a) and gel (b) phase lipids in pure water in all channels from 3 μ s simulation. cytosol \rightarrow vacuole is referred to as Up and vacuole \rightarrow cytosol is referred to as Down. Final table (c) shows the corrected number of water passages across all channels in the gel phase lipid at 310K.

12. Water and Ammonia Selectivity Through Aquaporin AtTIP2;1

Channel	Up	Down	Total
A	15	17	32
B	25	18	43
C	33	24	57
D	21	24	45
			177
Average Time: 3.5 ns		pf: -	

(a) Lipid phase lipid at 310K

Channel	Up	Down	Total
A	1	2	3
B	1	3	4
C	0	1	1
D	9	2	11
			19
Average Time: 28.6 ns		pf: -	

(b) Gel phase lipid at 270K

Channel	Up	Down	Total
A	3	5	8
B	3	8	10
C	0	3	3
D	24	5	29
			50

(c) Gel phase corrected to 310K

Table 12.2.: The passage of NH₃ across the aquaporin AtTIP2;1 each channels in liquid (a) and gel (b) phase lipids in systems where 1000 ammonia molecules are present in the system from 3 μ s simulation. cytosol \rightarrow vacuole is referred to as Up and vacuole \rightarrow cytosol is referred to as Down. Final table (c) shows the corrected number of water passages across all channels in the gel phase lipid at 310K.

Channel	Up	Down	Total
A	1191	1199	2390
B	1266	1191	2457
C	1272	1206	2478
D	1566	1509	3075
			10400
Average Time: 2.3 ns		pf: 13.9	

(a) Lipid phase lipid at 310K

Channel	Up	Down	Total
A	125	143	268
B	149	210	359
C	79	68	147
D	345	353	698
			1472
Average Time: 12.6 ns		pf: 2.4	

(b) Gel phase lipid at 270K

Channel	Up	Down	Total
A	327	374	701
B	390	550	939
C	207	178	385
D	903	924	1827
			3852

(c) Gel phase corrected to 310K

Table 12.3.: The passage of water across the aquaporin AtTIP2;1 each channel in liquid (a) and gel (b) phase lipids in systems where 1000 NH₃ molecules are present in the system from 3 μ s simulation. cytosol \rightarrow vacuole is referred to as Up and vacuole \rightarrow cytosol is referred to as Down. Final table (c) shows the corrected number of water passages across all channels in the gel phase lipid at 310K.

Channel	Up	Down	Total
A	403	371	774
B	1113	1050	721
C	1229	1232	820
D	343	219	562
			1987
Average Time: 3.8		pf: 11.6	

(a) Lipid phase lipid at 310K

Channel	Up	Down	Total
A	53	65	118
B	234	226	460
C	383	362	745
D	159	150	309
			1632
Average Time: 10.6		pf: 3.4	

(b) Gel phase lipid at 270K

Channel	Up	Down	Total
A	139	170	309
B	612	591	1204
C	1002	947	1950
D	416	393	809
			4271

(c) Gel phase corrected to 310K

Table 12.4.: The passage of water across the aquaporin AtTIP2;1 each channel in liquid **(a)** and gel **(b)** phase lipids in systems where 1000 NH₄⁺ molecules are present in the system from 3 μs simulation. cytosol → vacuole is referred to as Up and vacuole → cytosol is referred to as Down. Final table **(c)** shows the corrected number of water passages across all channels in the gel phase lipid at 310K.

12.2. Distance measurement between important residues across the channels

The data are derived from simulations performed at 310K for liquid phase lipids in each of the three systems (pure water, ammonia, and ammonium ion-containing system.) Similarly, the data for gel phase simulation is performed at 270K in each of the systems.

GLU146 (O)–THR150(OG1)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	2.75 ± 0.14	2.75 ± 0.14	2.83 ± 0.22	3.41 ± 0.39	2.86 ± 0.25	2.71 ± 0.11
B	2.78 ± 0.18	2.74 ± 0.11	2.75 ± 0.14	2.73 ± 0.11	2.9 ± 0.26	2.73 ± 0.11
C	2.78 ± 0.18	2.74 ± 0.12	2.8 ± 0.16	2.78 ± 0.16	2.79 ± 0.16	2.81 ± 0.14
D	2.76 ± 0.13	2.74 ± 0.12	2.75 ± 0.13	2.73 ± 0.11	2.76 ± 0.13	2.77 ± 0.15

Table 12.5.: Average distance (in Å, mean ± std. deviation) measured from 3 μs simulation between the atom O of GLU146 and atom OG1 of THR150.

GLU146 (O)–SER192(OG1)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.77 ± 0.39	3.82 ± 0.43	4.19 ± 0.73	3.82 ± 0.34	5.63 ± 1.66	5.58 ± 0.69
B	3.66 ± 0.2	3.61 ± 0.12	6.04 ± 2.72	3.61 ± 0.12	4.58 ± 1.04	3.62 ± 0.13
C	5.04 ± 1.41	5.54 ± 1.07	5.51 ± 1.4	4.89 ± 0.9	5.08 ± 1.15	4.65 ± 1.03
D	3.62 ± 0.13	3.96 ± 0.57	3.67 ± 0.24	3.7 ± 0.13	4.0 ± 0.66	7.54 ± 3.13

Table 12.6.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom O of GLU146 and atom OG1 of SER192.

GLU146(CD)–SER195(N)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.76 ± 0.31	3.78 ± 0.22	4.56 ± 0.91	5.19 ± 0.65	6.68 ± 2.0	3.48 ± 0.18
B	3.72 ± 0.6	3.49 ± 0.12	3.7 ± 0.25	3.49 ± 0.12	4.73 ± 1.26	3.54 ± 0.17
C	5.39 ± 1.25	8.02 ± 0.7	6.04 ± 1.62	7.33 ± 0.77	6.52 ± 0.76	7.62 ± 0.39
D	3.48 ± 0.13	3.93 ± 0.33	3.63 ± 0.29	3.77 ± 0.17	3.95 ± 0.43	3.8 ± 0.21

Table 12.7.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom CD of GLU146 and atom N of SER195.

GLU146 (O) -THR150(OG1)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	2.75 ± 0.14	2.75 ± 0.14	2.83 ± 0.22	3.41 ± 0.39	2.86 ± 0.25	2.71 ± 0.11
B	2.78 ± 0.18	2.74 ± 0.11	2.75 ± 0.14	2.73 ± 0.11	2.9 ± 0.26	2.73 ± 0.11
C	2.78 ± 0.18	2.74 ± 0.12	2.8 ± 0.16	2.78 ± 0.16	2.79 ± 0.16	2.81 ± 0.14
D	2.76 ± 0.13	2.74 ± 0.12	2.75 ± 0.13	2.73 ± 0.11	2.76 ± 0.13	2.77 ± 0.15

Table 12.8.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom O of GLU146 and atom OG1 of THR150.

GLU24(CD) – GLN108(NE2)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.61 ± 0.23	3.83 ± 0.19	3.55 ± 0.2	3.8 ± 0.19	3.63 ± 0.22	5.0 ± 1.3
B	3.73 ± 0.23	3.74 ± 0.17	3.75 ± 0.21	3.7 ± 0.17	3.61 ± 0.23	3.68 ± 0.18
C	3.6 ± 0.21	3.53 ± 0.17	3.6 ± 0.22	3.59 ± 0.18	3.55 ± 0.2	3.55 ± 0.18
D	3.73 ± 0.18	3.59 ± 0.2	3.63 ± 0.22	3.7 ± 0.17	3.71 ± 0.25	3.67 ± 0.2

Table 12.9.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom CD of GLU24 and atom NE2 of GLN180.

ASN197(OD1)–ALA199(N)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.02 ± 0.16	3.02 ± 0.17	3.01 ± 0.16	2.95 ± 0.13	3.04 ± 0.19	3.12 ± 0.18
B	3.0 ± 0.16	2.98 ± 0.14	3.03 ± 0.17	2.97 ± 0.14	2.97 ± 0.15	3.0 ± 0.15
C	3.04 ± 0.17	2.96 ± 0.14	3.1 ± 0.2	2.98 ± 0.14	3.02 ± 0.18	3.04 ± 0.16
D	3.02 ± 0.16	3.0 ± 0.14	3.03 ± 0.16	3.0 ± 0.15	3.03 ± 0.16	3.08 ± 0.16

Table 12.10.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom OD1 of ASN197 and atom N of ALA199.

ASN197(N) – VAL83(O)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.24 ± 0.34	3.57 ± 0.45	3.3 ± 0.36	3.23 ± 0.29	3.83 ± 1.12	3.2 ± 0.37
B	3.61 ± 0.49	3.73 ± 0.33	3.57 ± 0.5	3.71 ± 0.33	3.2 ± 0.33	3.83 ± 0.35
C	3.41 ± 0.45	3.2 ± 0.29	3.64 ± 0.5	3.37 ± 0.39	3.33 ± 0.4	3.15 ± 0.31
D	3.69 ± 0.44	3.86 ± 0.35	3.32 ± 0.43	3.84 ± 0.35	3.59 ± 0.47	3.21 ± 0.29

Table 12.11.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom N of ASN197 and atom O of VAL83.

12.3. Time Evolution of Distances

ASN83(O)–THR87 (OG1)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.0 ± 0.27	2.94 ± 0.2	3.12 ± 0.34	2.92 ± 0.2	3.21 ± 0.42	3.22 ± 0.35
B	3.06 ± 0.28	3.02 ± 0.22	3.04 ± 0.28	3.01 ± 0.22	3.03 ± 0.3	3.0 ± 0.22
C	3.11 ± 0.34	3.47 ± 0.28	3.2 ± 0.38	3.07 ± 0.34	3.31 ± 0.4	3.35 ± 0.27
D	3.02 ± 0.26	3.05 ± 0.28	3.06 ± 0.31	3.06 ± 0.26	3.02 ± 0.26	3.02 ± 0.25

Table 12.12.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom O of ASN83 and atom OG1 of THR87.

ALA214(O) – ASP205 (N)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.12 ± 0.22	3.1 ± 0.2	3.11 ± 0.21	3.13 ± 0.21	3.11 ± 0.23	3.15 ± 0.21
B	3.13 ± 0.22	3.07 ± 0.19	3.12 ± 0.22	3.11 ± 0.2	3.11 ± 0.21	3.1 ± 0.2
C	3.12 ± 0.24	3.05 ± 0.18	3.12 ± 0.26	3.05 ± 0.2	3.05 ± 0.43	3.08 ± 0.19
D	3.12 ± 0.22	3.09 ± 0.2	3.12 ± 0.22	3.11 ± 0.2	3.14 ± 0.26	3.07 ± 0.19

Table 12.13.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom O of ALA214 and atom N of ASP205.

ALA134(O)–PHE192(OH)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.1 ± 0.36	3.12 ± 0.31	3.03 ± 0.24	3.03 ± 0.29	4.98 ± 1.79	2.97 ± 0.23
B	3.02 ± 0.23	3.0 ± 0.2	4.24 ± 1.72	3.12 ± 0.62	3.31 ± 0.94	3.01 ± 0.2
C	6.51 ± 2.84	6.73 ± 0.47	8.86 ± 2.02	6.79 ± 0.44	6.89 ± 1.17	7.44 ± 1.35
D	3.01 ± 0.23	3.13 ± 0.3	3.21 ± 0.82	3.08 ± 0.25	3.01 ± 0.25	5.16 ± 1.83

Table 12.14.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom O of ALA134 and atom OH of PHE192.

ALA50 (O) - TYR42 (OH)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	3.37 ± 0.79	3.5 ± 0.64	3.34 ± 0.68	3.33 ± 0.64	3.56 ± 0.68	3.24 ± 0.64
B	3.46 ± 0.7	3.53 ± 0.68	3.86 ± 0.69	3.5 ± 0.73	3.58 ± 0.64	3.68 ± 0.58
C	3.66 ± 0.66	3.72 ± 0.58	3.6 ± 0.63	3.58 ± 0.61	3.44 ± 0.69	3.66 ± 0.76
D	3.69 ± 0.62	3.63 ± 0.48	3.28 ± 0.65	3.64 ± 0.46	3.36 ± 0.6	3.9 ± 0.64

Table 12.15.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom O of ALA50 and atom OH of TYR42.

ALA199 (O) – SER112 (OG)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	2.87 ± 0.35	3.07 ± 0.38	2.88 ± 0.34	3.06 ± 0.37	2.81 ± 0.32	2.93 ± 0.33
B	3.25 ± 0.56	3.69 ± 0.23	3.07 ± 0.49	3.73 ± 0.24	2.91 ± 0.38	3.82 ± 0.26
C	2.93 ± 0.43	2.69 ± 0.1	3.01 ± 0.44	2.78 ± 0.26	2.75 ± 0.22	2.68 ± 0.11
D	3.67 ± 0.35	3.8 ± 0.24	3.25 ± 0.5	3.78 ± 0.25	3.49 ± 0.52	3.5 ± 0.24

Table 12.16.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom O of ALA199 and atom OG of SER112.

GLY80(N)-SER78(OG)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	6.25 ± 2.31	3.23 ± 0.24	7.86 ± 0.82	3.26 ± 0.25	7.6 ± 1.27	8.08 ± 0.66
B	4.79 ± 2.18	3.38 ± 0.37	3.27 ± 0.29	3.25 ± 0.3	7.53 ± 1.4	3.3 ± 0.25
C	7.85 ± 0.59	7.88 ± 0.48	7.97 ± 0.42	7.88 ± 0.48	7.95 ± 0.45	7.92 ± 0.45
D	3.27 ± 0.29	7.11 ± 1.85	6.01 ± 2.42	3.22 ± 0.3	7.4 ± 1.51	6.34 ± 2.25

Table 12.17.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom N of GLY80 and atom OG of SER78.

GLY193(NH2)–HSD131(NE2)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	4.23 ± 0.8	4.11 ± 0.72	3.94 ± 0.89	3.73 ± 0.79	4.99 ± 1.9	4.34 ± 1.06
B	3.99 ± 0.76	3.9 ± 0.64	4.91 ± 1.33	3.92 ± 0.67	3.65 ± 0.9	3.8 ± 0.6
C	5.66 ± 1.86	4.45 ± 1.36	5.67 ± 1.57	6.91 ± 1.82	6.47 ± 2.2	10.5 ± 2.82
D	3.98 ± 0.78	4.39 ± 0.6	4.18 ± 0.84	4.24 ± 0.59	8.38 ± 1.4	5.49 ± 1.06

Table 12.18.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom NH2 of GLY193 and atom NE2 of HSD131.

LEU109(O) – THR113(OG1)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	2.92 ± 0.27	2.85 ± 0.16	2.92 ± 0.26	2.85 ± 0.17	2.96 ± 0.34	2.9 ± 0.24
B	2.89 ± 0.27	2.85 ± 0.21	2.9 ± 0.27	2.81 ± 0.18	2.9 ± 0.26	2.8 ± 0.16
C	2.9 ± 0.27	2.91 ± 0.23	2.91 ± 0.3	2.97 ± 0.26	2.93 ± 0.32	3.0 ± 0.37
D	2.85 ± 0.26	2.78 ± 0.15	2.87 ± 0.23	2.8 ± 0.16	2.84 ± 0.22	2.85 ± 0.2

Table 12.19.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom O of LEU109 and atom OG1 of THR113.

SER212(**OG1**)-ASP210(**NE1**)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	4.07 ± 1.01	3.64 ± 0.65	4.22 ± 1.14	3.68 ± 0.67	4.17 ± 1.33	3.95 ± 1.13
B	4.02 ± 1.01	4.2 ± 1.05	4.03 ± 1.01	4.23 ± 1.11	4.28 ± 1.25	3.92 ± 0.96
C	4.24 ± 1.33	3.57 ± 0.57	4.54 ± 1.55	3.38 ± 0.18	3.95 ± 1.43	3.43 ± 0.35
D	4.01 ± 0.99	4.04 ± 0.96	4.04 ± 1.02	4.18 ± 0.98	4.47 ± 1.49	3.75 ± 0.76

Table 12.20.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom OG1 of SER212 and atom NE1 of ASP210.

TYR217(**OH**) – SER192(**O**)

Channel	Liquid H ₂ O	Gel H ₂ O	Liquid NH ₃	Gel NH ₃	Liquid NH ₄ ⁺	Gel NH ₄ ⁺
A	2.78 ± 0.34	2.72 ± 0.14	2.78 ± 0.24	2.8 ± 0.19	5.22 ± 2.1	2.71 ± 0.12
B	2.73 ± 0.19	2.71 ± 0.13	2.89 ± 0.7	2.7 ± 0.11	2.85 ± 0.52	2.72 ± 0.14
C	7.02 ± 2.85	8.58 ± 1.1	8.2 ± 1.72	8.25 ± 1.03	7.31 ± 1.41	6.55 ± 0.51
D	2.73 ± 0.16	2.79 ± 0.3	2.74 ± 0.22	2.71 ± 0.11	3.4 ± 1.32	6.33 ± 2.84

Table 12.21.: Average distance (in Å, std. ± deviation) measured from 3 μs simulation between the atom 192 of TYR217 and atom O of SER192.

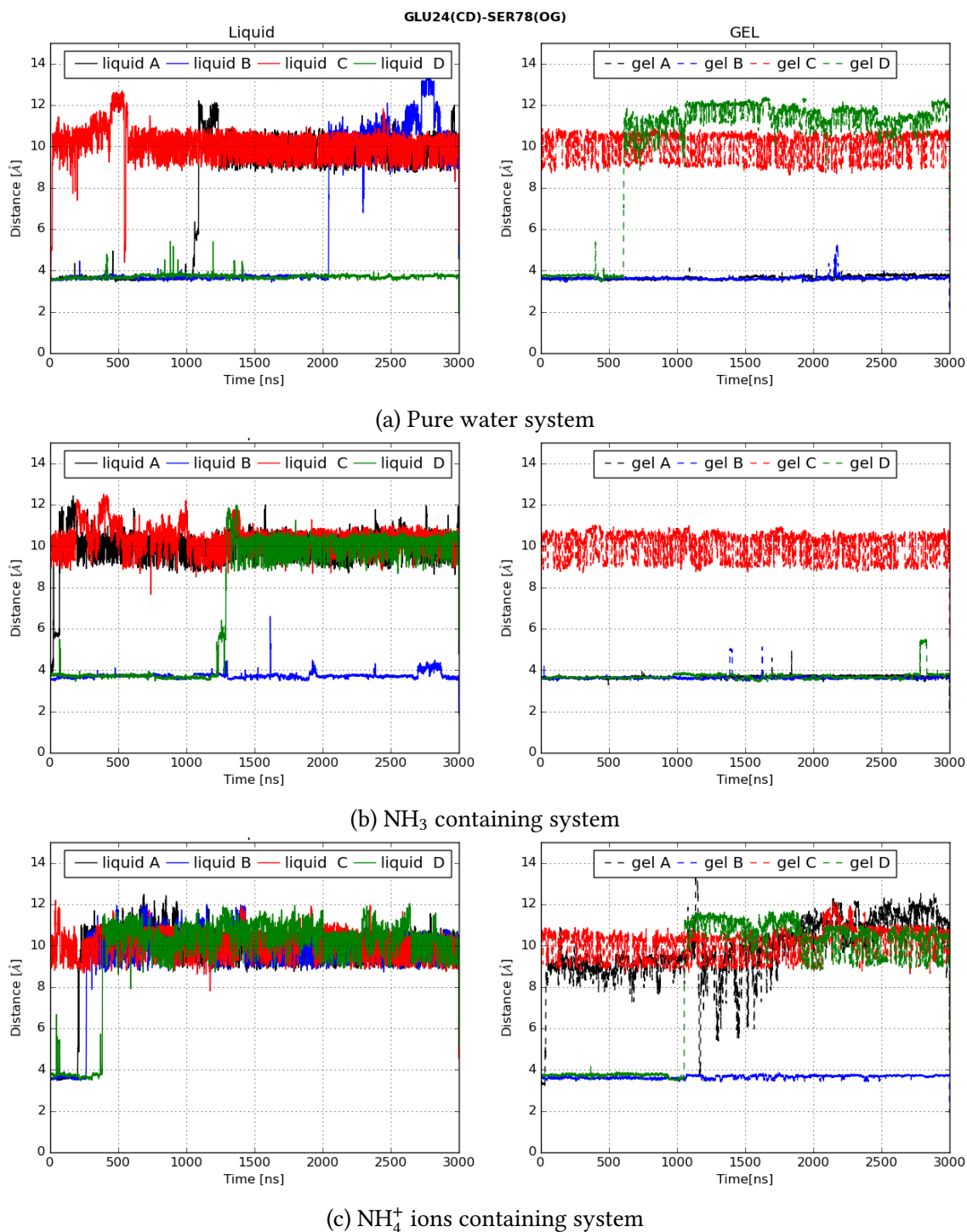


Figure 12.1.: The average distance (in Å) is shown as the evolution of time in liquid phase lipid (on the left side) and gel phase lipid (on the right side) in all the systems for 3 μ s. The distance between heavy atom O of GLU24 and atom O _{γ} of SER78 is measured.

12.4. Torsional angle analysis of the bulky residues

The data are derived from simulations performed at 310K for liquid phase lipids in each of the three systems (pure water, ammonia, and ammonium ion-containing system.) Similarly, the data for gel phase simulation is performed at 270K in each of the systems.

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	73.33 ± 47.08	60.93 ± 7.58	A	77.92 ± 55.19	61.06 ± 7.44
B	61.97 ± 8.14	62.48 ± 7.43	B	61.72 ± 11.1	62.49 ± 7.44
C	114.97 ± 91.28	65.35 ± 14.06	C	104.75 ± 86.31	63.13 ± 8.06
D	-103.24 ± 41.38	-95.29 ± 9.77	D	-154.29 ± 92.69	-94.86 ± 8.77

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	-18.75 ± 93.62	59.28 ± 7.37
B	61.51 ± 8.14	62.61 ± 7.48
C	95.26 ± 75.56	65.41 ± 7.94
D	-96.96 ± 18.99	-98.12 ± 24.48

(c) 1000 NH₄⁺ ions containing system

Table 12.22.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-C_δ in **ARG200**.

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	168.03 ± 10.59	169.27 ± 10.04	A	168.73 ± 11.84	173.8 ± 11.09
B	171.06 ± 11.11	173.85 ± 9.0	B	169.0 ± 11.15	174.11 ± 8.93
C	166.35 ± 12.81	161.82 ± 11.85	C	166.14 ± 13.38	163.43 ± 13.92
D	-148.22 ± 7.65	-148.41 ± 7.25	D	-145.89 ± 7.95	-149.05 ± 7.36

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	168.33 ± 25.54	165.35 ± 10.54
B	168.53 ± 11.42	174.19 ± 9.28
C	159.2 ± 14.9	166.49 ± 12.35
D	-146.98 ± 8.52	-146.17 ± 7.66

(c) 1000 NH₄⁺ ions containing system

Table 12.23.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-N_{δ2} of the residue **ASN83**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	157.44 ± 10.41	158.38 ± 10.36	A	161.6 ± 12.02	160.63 ± 9.78
B	160.27 ± 11.01	162.86 ± 9.3	B	157.2 ± 11.19	163.11 ± 9.31
C	163.27 ± 13.57	176.39 ± 11.43	C	164.5 ± 13.03	173.55 ± 10.98
D	-143.81 ± 39.09	-149.92 ± 8.37	D	-147.71 ± 28.85	-146.98 ± 8.01

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	181.21 ± 35.6	159.56 ± 10.41
B	160.99 ± 11.46	161.83 ± 9.18
C	170.65 ± 14.81	167.58 ± 12.28
D	-167.57 ± 15.88	-166.06 ± 14.7

(c) 1000 NH₄⁺ ions containing systemTable 12.24.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-N_{δ₂} of the residue **ASN197**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	34.85 ± 59.07	-78.11 ± 81.47	A	60.21 ± 38.48	-58.21 ± 9.03
B	37.95 ± 58.3	-59.15 ± 9.9	B	-38.15 ± 47.36	-60.62 ± 9.32
C	71.92 ± 14.12	63.97 ± 30.35	C	44.61 ± 56.57	71.39 ± 7.95
D	-20.37 ± 56.03	-10.91 ± 60.92	D	15.44 ± 65.45	-59.29 ± 9.72

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	34.85 ± 59.07	-65.8 ± 90.38
B	37.95 ± 58.3	-59.15 ± 9.9
C	71.92 ± 14.12	63.97 ± 30.35
D	-20.37 ± 56.03	-10.91 ± 60.92

(c) 1000 NH₄⁺ ions containing systemTable 12.25.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-C_δ of the residue **GLU24**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	-81.48 ± 29.17	-79.66 ± 9.37	A	-75.87 ± 14.56	-70.68 ± 10.48
B	-73.73 ± 10.99	-70.38 ± 9.06	B	-75.95 ± 22.14	-70.8 ± 9.05
C	-150.02 ± 58.52	-166.17 ± 33.45	C	-174.18 ± 28.8	-163.68 ± 29.21
D	-70.42 ± 9.73	-79.64 ± 11.12	D	-73.94 ± 11.19	-80.38 ± 9.05

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	-122.8 ± 51.41	-72.08 ± 9.2
B	-65.58 ± 15.42	-72.01 ± 9.57
C	-161.88 ± 33.78	-184.84 ± 11.46
D	-101.0 ± 50.54	-77.58 ± 13.99

(c) 1000 NH₄⁺ ions containing system

Table 12.26.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-C_δ of the residue **GLU146**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	80.58 ± 12.03	78.67 ± 11.1	A	80.54 ± 12.33	79.86 ± 10.5
B	79.82 ± 12.12	80.58 ± 10.76	B	80.46 ± 12.19	79.9 ± 10.7
C	81.47 ± 12.68	86.81 ± 16.13	C	81.72 ± 13.09	84.72 ± 10.52
D	79.9 ± 12.92	80.57 ± 11.16	D	81.14 ± 12.39	79.15 ± 11.33

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	81.9 ± 11.72	81.92 ± 17.37
B	80.95 ± 11.99	81.1 ± 10.84
C	81.26 ± 12.9	89.58 ± 19.02
D	78.13 ± 11.75	85.44 ± 12.61

(c) 1000 NH₄⁺ ions containing system

Table 12.27.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-N_{δ1} of the residue **HSD63**

Channel	Liquid H ₂ O	Liquid NH ₃
A	80.58 ± 12.03	78.67 ± 11.1
B	79.82 ± 12.12	80.58 ± 10.76
C	81.47 ± 12.68	86.81 ± 16.13
D	79.9 ± 12.92	80.57 ± 11.16

(a) Pure water system

Channel	Liquid H ₂ O	Liquid NH ₃
A	80.54 ± 12.33	79.86 ± 10.5
B	80.46 ± 12.19	79.9 ± 10.7
C	81.72 ± 13.09	84.72 ± 10.52
D	81.14 ± 12.39	79.15 ± 11.33

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	81.9 ± 11.72	81.92 ± 17.37
B	80.95 ± 11.99	81.1 ± 10.84
C	81.26 ± 12.9	89.58 ± 19.02
D	78.13 ± 11.75	85.44 ± 12.61

(c) 1000 NH₄⁺ ions containing system

Table 12.28.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μ s simulation along the C _{α} -C _{β} -C _{γ} -N _{δ_1} of the residue **HSD81**

Channel	Liquid H ₂ O	Liquid NH ₃
A	-87.85 ± 49.65	-81.43 ± 39.47
B	-89.42 ± 49.8	-81.33 ± 33.4
C	-24.61 ± 81.5	-76.08 ± 50.94
D	-94.27 ± 57.66	-96.77 ± 57.42

(a) Pure water system

Channel	Liquid H ₂ O	Liquid NH ₃
A	-91.99 ± 50.28	-82.11 ± 24.95
B	-96.35 ± 51.62	-83.4 ± 39.44
C	43.49 ± 102.9	-239.2 ± 70.46
D	-93.82 ± 58.76	-95.24 ± 56.8

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	-21.65 ± 87.77	-26.5 ± 113.78
B	-71.49 ± 52.54	-67.9 ± 48.17
C	40.63 ± 95.8	8.66 ± 95.17
D	6.51 ± 131.61	-94.97 ± 68.14

(c) 1000 NH₄⁺ ions containing system

Table 12.29.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μ s simulation along the C _{α} -C _{β} -C _{γ} -N _{δ_1} of the residue **HSD131**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	61.9 ± 16.95	61.67 ± 14.93	A	65.24 ± 17.51	69.05 ± 15.41
B	61.18 ± 19.5	58.92 ± 18.63	B	55.47 ± 35.02	60.88 ± 18.9
C	78.92 ± 36.14	73.36 ± 15.51	C	77.49 ± 23.64	68.17 ± 13.38
D	64.39 ± 21.16	56.65 ± 14.5	D	61.98 ± 25.18	61.56 ± 16.49

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	73.98 ± 21.35	62.94 ± 14.52
B	61.44 ± 19.84	64.77 ± 17.15
C	-6.68 ± 68.65	70.98 ± 12.28
D	61.8 ± 24.5	61.79 ± 17.47

(c) 1000 NH₄⁺ ions containing system

Table 12.30.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-N_{δ1} of the residue **HSD214**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	-136.31 ± 71.26	-139.51 ± 67.39	A	-162.57 ± 81.21	-129.62 ± 94.38
B	-111.33 ± 62.72	-85.54 ± 21.18	B	-130.56 ± 70.65	-87.8 ± 30.98
C	-147.92 ± 81.51	-183.65 ± 37.92	C	-155.05 ± 68.08	-177.25 ± 72.18
D	-77.57 ± 35.35	-143.29 ± 65.53	D	-29.9 ± 100.38	-135.89 ± 64.72

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	-165.08 ± 74.95	-148.23 ± 85.96
B	-148.62 ± 68.0	-100.16 ± 45.92
C	-176.0 ± 53.82	-136.1 ± 87.18
D	-126.05 ± 63.31	-150.13 ± 56.04

(c) 1000 NH₄⁺ ions containing system

Table 12.31.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_β-C_γ-S_δ-C_ε of the residue **MET196**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	68.65 ± 13.88	66.8 ± 12.52	A	71.22 ± 13.11	68.69 ± 12.23
B	-97.32 ± 47.69	-109.32 ± 11.24	B	-113.44 ± 13.88	-108.22 ± 11.2
C	62.84 ± 14.11	70.57 ± 14.28	C	65.1 ± 15.1	66.89 ± 13.25
D	61.8 ± 13.84	55.85 ± 11.7	D	69.43 ± 13.75	57.85 ± 11.67

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	63.21 ± 14.04	76.29 ± 14.06
B	-113.2 ± 13.69	-113.07 ± 10.7
C	67.17 ± 14.64	61.97 ± 17.07
D	67.06 ± 12.34	66.41 ± 12.71

(c) 1000 NH₄⁺ ions containing systemTable 12.32.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-C_δ of the residue **TYR42**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	32.95 ± 85.26	-30.9 ± 13.1	A	55.67 ± 93.51	-28.78 ± 13.56
B	47.0 ± 91.19	55.07 ± 94.52	B	78.17 ± 91.82	76.62 ± 93.5
C	40.08 ± 92.95	132.6 ± 11.33	C	-9.93 ± 78.17	133.69 ± 11.06
D	88.19 ± 89.77	28.6 ± 91.8	D	72.88 ± 95.2	107.48 ± 82.64

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	-32.02 ± 58.07	-3.08 ± 67.17
B	42.94 ± 92.88	99.55 ± 84.58
C	89.76 ± 78.08	132.96 ± 11.46
D	32.5 ± 94.51	96.38 ± 82.41

(c) 1000 NH₄⁺ ions containing systemTable 12.33.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-C_δ of the residue **TYR104**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	-118.88 ± 91.66	-99.46 ± 98.0	A	-98.65 ± 81.89	-47.61 ± 36.21
B	-178.1 ± 7.87	-178.41 ± 6.51	B	-180.35 ± 8.21	-179.16 ± 6.48
C	-61.35 ± 56.64	-10.49 ± 19.37	C	-119.13 ± 83.16	-14.7 ± 20.27
D	-116.67 ± 89.26	-64.04 ± 23.23	D	-107.73 ± 78.33	-54.76 ± 33.64

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	-186.29 ± 88.56	-177.82 ± 70.72
B	-181.03 ± 7.93	-179.47 ± 6.47
C	-162.37 ± 86.05	-23.21 ± 31.14
D	-68.01 ± 60.25	-6.47 ± 25.6

(c) 1000 NH₄⁺ ions containing system

Table 12.34.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-C_δ of the residue **TYR158**

Channel	Liquid H ₂ O	Liquid NH ₃	Channel	Liquid H ₂ O	Liquid NH ₃
A	-65.6 ± 10.39	-66.15 ± 10.07	A	-60.89 ± 11.3	-59.56 ± 9.66
B	-64.52 ± 9.86	-65.74 ± 8.37	B	-64.29 ± 10.58	-66.64 ± 8.6
C	247.0 ± 84.07	234.86 ± 89.87	C	225.19 ± 91.5	176.9 ± 66.85
D	-65.35 ± 9.37	-66.99 ± 10.6	D	-65.28 ± 10.32	85.45 ± 66.66

(a) Pure water system

(b) 1000 NH₃ containing system

Channel	Liquid H ₂ O	Liquid NH ₃
A	-106.6 ± 80.7	-58.05 ± 9.73
B	-60.47 ± 12.12	-63.9 ± 9.46
C	193.41 ± 96.17	153.16 ± 19.92
D	-40.47 ± 59.15	-62.8 ± 11.74

(c) 1000 NH₄⁺ ions containing system

Table 12.35.: Average torsional angle (in degree, mean ± std. deviation) measured from 3 μs simulation along the C_α-C_β-C_γ-C_δ of the residue **TYR217**

Curriculum Vitae

Deepak Kumar

Address: Jehanabad, Bihar, India
E-mail: deepaknextprez@gmail.com

Education

08/2019 – 12/2022 Doctoral student at the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
09/2017 – 07/2019 Master in Theoretical chemistry and computational modeling, University of Valencia, Valencia, Spain
08/2015 – 07/2017 Master in Chemistry, Indian Institute of Technology, Mandi, India
07/2011 – 05/2014 B.Sc. (Hons.) Chemistry, University of Delhi, Delhi, India
04/2009 – 03/2011 Senior secondary school, St. Michael's High School, Patna, India
04/2007 – 03/2009 Secondary School, B. D. Public School, Patna, India

List of publications

1. Rajendran Saravanan, **Deepak Kumar** and Aniruddha Chakraborty, "Quantum physics of the systems with ultrashort potentials: Introduction to a new analytically solvable model". arXiv:1805.01895v4 [quant-ph] 26th December 2018.
2. **Deepak Kumar**, Kirill Zinovjev, Carlos A. Ramos and Iñaki Tuñón, "Computational Simulations of Enzyme Reactivity in Chorismate Mutase and Mutants: Analysis of Electrostatic Contributions to Catalysis". (review)
3. Li Tian, **Deepak Kumar**, Stephan L. Grage, Tomas Kubar, Sebastian Otteni, Parvesh Wadhvani, Anne S. Ulrich, "Self-association of the oncoprotein E5 and PDGFR β revealed by distance measurements". (In preparation)
4. **Deepak Kumar**, Parvesh Wadhvani, Anne S. Ulrich, Marcus Elstner, Tomas Kubar, "Long-range proton transfer through tetrameric charge zipper TisB assembly of peptide". (In preparation)
5. Thomas Barta, **Deepak Kumar**, Andreas Horner, Tomas Kubar "Water Permeability and Ammonia Selectivity Through Aquaporin AtTIP2;1 in Liquid and Gel PhaseLipid". (In preparation)

Thank you for constant motivation,



and the journey begins...